



Sur les inégalités de Sobolev logarithmiques en théorie de l'information et pour des systèmes de spins conservatifs en mécanique statistique

Djalil Chafai

► To cite this version:

Djalil Chafai. Sur les inégalités de Sobolev logarithmiques en théorie de l'information et pour des systèmes de spins conservatifs en mécanique statistique. Mathématiques [math]. Université Paul Sabatier - Toulouse III, 2002. Français. NNT: . tel-00001382

HAL Id: tel-00001382

<https://theses.hal.science/tel-00001382>

Submitted on 7 Jun 2002

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Numéro d'ordre: XXXX

THÈSE

présentée en vue de l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PAUL SABATIER

Discipline : MATHÉMATIQUES

Spécialité : PROBABILITÉS

par Djalil CHAFAÏ

Sur les inégalités de Sobolev logarithmiques en théorie de l'information et pour des systèmes de spins conservatifs en mécanique statistique

Soutenue le 17 mai 2002 devant le jury composé de
Madame et Messieurs les Professeurs :

Dominique BAKRY	Université Paul Sabatier	Examineur
Mireille CAPITAINE	CNRS & Université Paul Sabatier	Examinatrice
Giambattista GIACOMIN	Université de Paris VII	Rapporteur
Claudio LANDIM	CNRS & Université de Rouen	Rapporteur
Michel LEDOUX	Université Paul Sabatier	Directeur de thèse
Gilles ROYER	Université d'Orléans	Président



Laboratoire de Statistique et Probabilités
Unité Mixte de Recherche C.N.R.S 5583
Bâtiment 1R1, U.F.R. M.I.G.
Université Paul Sabatier - Toulouse III
118 route de Narbonne, F-31062, CEDEX 4, France.

Mise en page par L^AT_EX 2_ε le 18 mai 2002.

Au timide que j'ai été... et à tous ses semblables.

Avant propos

Ce document constitue une présentation des modestes travaux que j'ai mené pendant quelques années sous la direction de Michel LEDOUX dans le domaine des inégalités de SOBOLEV logarithmiques. Les conventions font qu'il s'agit d'une « thèse » présentée en vue d'obtenir le titre de « docteur » à l'issue d'une « soutenance » devant un « jury ». En mathématiques comme ailleurs, il s'agit d'une sorte de rite initiatique nécessaire à l'intégration du monde de la recherche, le jury mesurant dans sa grande sagesse, au vu de la thèse et après un exposé, si le candidat est acceptable ou non par la tribu. À l'heure où j'écris ces lignes, je ne sais toujours pas si la tribu voudra bien de moi, mais j'ose espérer que ça sera le cas !

Je tiens à remercier Giambattista GIACOMIN et Claudio LANDIM d'avoir aimablement accepté d'être mes rapporteurs. Merci également à Mireille CAPITAIN et à Gilles ROYER, dont la présence dans le jury me fait très plaisir, et enfin à Dominique BAKRY qui a toujours pris le temps de répondre à mes questionnements avec sa verve habituelle.

La thèse est décidément une drôle d'aventure. On y apprend beaucoup sur soi, sur les choses de l'esprit et sur la nature et les limites du travail des enseignants-chercheurs. Michel LEDOUX a su m'accompagner avec une certaine finesse dans ce parcours initiatique. Je tiens à lui exprimer ici ma grande estime. C'est son cours de probabilité de maîtrise qui m'a fait redécouvrir et aimer cette discipline à travers l'analyse. J'ai pu apprécier, tout au long de ces années, son sens de l'esthétique, sa disponibilité et sa culture mathématique. J'aime cet équilibre qui le caractérise, entre création et (ré)interprétation. Ces années m'ont permis d'apprendre un certain nombre de techniques et de concepts. J'ai eu beaucoup de plaisir à compléter certaines des cases vides de mon puzzle personnel, au gré des rencontres et des lectures. J'ai aussi mieux cerné l'étendue de mon ignorance.

Mon travail de thèse a été interrompu par le service national que j'ai effectué au centre de recherches en météorologie de Météo-France à Toulouse. Fidèle à mon caractère, et peut-être au détriment de mon travail de doctorant, je n'ai pas pu m'empêcher de m'investir intellectuellement dans des problèmes concrets mêlant à la fois informatique, mathématiques et physique. J'en profite pour saluer l'ingénieure en météorologie Florence RABIER avec qui j'ai eu le bonheur de travailler pendant cette période. Je me suis ainsi fait une idée de ce que pouvait être un travail de recherche appliquée, avec les contraintes qui lui sont propres et les gratifications qu'il peut apporter.

Je garde de l'enfance un certain émerveillement pour les belles choses et une ouverture d'esprit face à la nouveauté. Les hasards de la vie et un certain goût pour

l'abstraction dépouillée ont fait que je soutiens aujourd'hui une thèse en mathématiques. Cela dit, il y a beaucoup d'autres domaines que je trouve attirants. Ces années de thèse ont été pour moi l'occasion de constater à quel point les chercheurs et les universitaires d'aujourd'hui sont spécialisés. Certes, il y a bien longtemps que les grands savants ont disparu, mais j'ai dû accepter, à contre cœur, que même les petits se font rares à présent. Il faut dire que les critères de sélection du système éducatif ne favorisent pas vraiment leur apparition. Pourtant, et de façon paradoxale, les approches transverses et les polycultures me semblent de plus en plus nécessaires afin de limiter les effets cloisonnants de l'hyperspécialisation. A contrario, il me paraît important d'ajouter que la dispersion est un gros risque pour ceux qui gardent l'esprit ouvert et la curiosité sans œillères. L'humilité et la relativisation qui en découlent sont parfois difficiles à vivre. Rétrospectivement, je m'en veux un peu de m'être éparpillé – mes talents n'étant pas à la hauteur de mes appétits – mais j'ose espérer qu'entre « têtes bien faites » et « têtes bien pleines » existent des états intermédiaires plus riches.

J'ai également investi beaucoup (trop ?) de mon temps et de mon énergie dans la construction d'une œuvre collective impliquant un certain nombre de « thésards » en probabilités du laboratoire : les « logsobs ». J'ai eu le bonheur de partager avec eux des années d'amitié, de coopération, de découvertes... Je pense en particulier à Grégory, dont j'apprécie le sens de l'esthétique et le tempérament insouciant et optimiste, à Florent, avec qui je partage un certain enthousiasme et un goût pour l'intuition et l'exposition, et à Cécile, avec qui j'ai eu l'occasion de travailler agréablement sur un sujet qui nous a réunis. Je leur serai toujours reconnaissant à tous les trois de m'avoir remonté le moral, chacun à sa manière, quand j'en avais besoin.

Je pense aussi à la tribu stochastique toulousaine, riche en personnalités singulières, qui a gentiment accueilli le timide que j'étais, à ses tout débuts, dans le microcosme de la recherche mathématique. Je me souviendrais toujours avec sourire de ma première rencontre avec Laurent MICLO : « hha-t'es en DEA, whaaaaa, i'sont'ous'nuls en DEA, waaaaa'eeuu... héh-héh ! ». Non, les gens ne sont pas forcément méchants, il faut s'ouvrir aux autres et écouter ce qu'ils ne nous disent pas !

À tous donc, et aux autres, je garde une place dans mes pensées

∴

*À peine nés de l'innocence
Un regard nous atterra
Éblouis les yeux baissés
Nous fûmes terre avant la lettre
Mais à germer nous voici graine
Dissidente des tournesols*

شَافِعِي رُشْدِي

Travaux

Publications

- D. CHAFAÏ, M. LEDOUX, « Méthodes fonctionnelles pour des grandes déviations quasi-gaussiennes », *C. R. Acad. Sci., Paris, Ser. I, Math.*, **329** (1999), no. 6, p. 523–526.
- C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO, G. SCHEFFER, « Sur les inégalités de Sobolev logarithmiques »¹, *Panoramas et Synthèses*, vol. 10, Société Mathématique de France, Paris (2000).
- D. CHAFAÏ, « Gaussian maximum of entropy and reversed logarithmic sobolev inequality », à paraître in *Séminaire de Probabilité XXXVI* (2002).

Prépublications

- D. CHAFAÏ, « Glauber versus Kawasaki for spectral gap and logarithmic Sobolev inequalities of some unbounded conservative spin systems », soumis en janvier 2002.

Travaux annexes non inclus

- F. RABIER, N. FOURRIÉ, D. CHAFAÏ, P. PRUNET, « Channel selection methods for infrared atmospheric sounding interferometer radiances », *Quarterly Journal of the Royal Meteorological Society*, Vol. 128 No. 581 - April 2002 Part A.

¹Seul le dernier chapitre écrit en collaboration avec Cécile ANÉ est inclus dans cette thèse; le premier chapitre – dont je suis également l’auteur – étant essentiellement didactique.

Table des matières

Avant propos	i
Liste des travaux	iii
Table des matières	v
0 Introduction	1
0.1 Inégalités fonctionnelles	3
0.1.1 Inégalités locales et critères de courbure	4
0.1.2 Inégalités inverses	7
0.2 Méthodes fonctionnelles pour PGD quasi-gaussiens	8
Perspectives	9
0.3 Entropie de Shannon et information de Fisher	10
Perspectives	12
0.4 Glauber vs Kawasaki pour certains systèmes de spins	13
Perspectives	17
Bibliographie	17
1 Méthodes fonctionnelles pour PGD quasi-gaussiens	19
1.1 Introduction	20
1.2 PGD pour certaines familles de mesures de Boltzmann	21
1.3 Semi-groupes diffusifs et généralisations possibles	23
Bibliographie	24
2 Inégalités entropiques en théorie de l'information	27
2.1 Introduction	27
2.2 L'entropie en théorie de l'information	29
2.2.1 Entropie d'une variable aléatoire discrète finie	29
2.2.2 L'entropie et le problème du codage	31
2.2.3 Entropie d'une variable aléatoire continue	33
2.2.4 Quelques propriétés immédiates de l'entropie	35
2.2.5 Information mutuelle et capacité d'un canal bruité	37

2.3	Version euclidienne de l'inégalité de Sobolev logarithmique	39
2.3.1	Entropie exponentielle de Shannon, information de Fisher . . .	40
2.4	Autour des inégalités de Shannon et de Blachman-Stam	41
2.4.1	Première méthode	41
2.4.2	Deuxième méthode	44
2.4.3	Troisième méthode	45
2.5	L'inégalité de Young et ses conséquences	46
2.6	Principes d'incertitude	48
2.6.1	Principe d'incertitude de Cramér-Rao	48
2.6.2	Principe d'incertitude de Weyl-Heisenberg	52
2.6.3	Principe d'incertitude de Beckner-Hirschman	53
2.7	Notes	56
	Bibliographie	58
3	Gaussian maximum of entropy and reversed LSI	63
3.1	Shannon's entropy power and Gross's inequality	63
3.2	Reversed Gross's logarithmic Sobolev inequality	65
3.3	Sketches of proofs	66
3.4	Remarks	68
	Acknowledgements	68
	Bibliography	68
4	Glauber vs Kawasaki for some unbounded spin systems	71
	Introduction	71
4.1	Proof of Proposition 4.0.2	78
4.2	Preliminaries to the proof of Theorem 4.0.3	79
4.3	Derivation of the Poincaré inequality	81
4.4	Derivation of the logarithmic Sobolev inequality	87
	Acknowledgements	90
	Bibliography	90
	Bibliographie générale	93

Chapitre 0

Introduction

Ce chapitre zéro a pour but de présenter les quatres suivants, dont un bref descriptif est donné ci-après. Ils ont tous déjà fait l'objet de publications, à l'exception du dernier qui constitue le travail accompli le plus récent. Chacun d'entre eux possède sa propre bibliographie, en plus de la bibliographie générale figurant en fin de document.

1. [CL99] « Méthodes fonctionnelles pour PGD quasi-gaussiens ». En collaboration avec M. LEDOUX, Note aux C. R. Acad. Sci. Paris 329, n°6, 523-526 (1999). Série I. Mathématiques.
2. [ABC⁺00] « Quelques inégalités entropiques en théorie de l'information ». Il s'agit d'un travail commun avec Cécile ANÉ. Il constitue le chapitre n°10 du livre intitulé « Sur les inégalités de Sobolev logarithmiques », Panoramas et Synthèses, vol. 10, Société Mathématique de France, Paris, 2000, écrit en collaboration avec S. BLACHÈRE, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO et G. SCHEFFER.
3. [Cha02a] « Gaussian maximum of entropy and reversed log-Sobolev inequality ». À paraître dans *Séminaire de Probabilité, XXXVI*, Lecture Notes in Math., Springer, Berlin, 2002.
4. [Cha02b] « Glauber versus Kawasaki for spectral gap and logarithmic Sobolev inequalities of some unbounded conservative spin systems ». Prépublication, 2002, en cours de soumission.

Le premier chapitre présente une application des inégalités fonctionnelles à l'établissement de principes de grandes déviations. Certaines inégalités fonctionnelles gaussiennes possèdent des généralisations simples dans des cas quasi-gaussiens. Elles permettent alors d'établir des principes et bornes de grandes déviations (PGD) quasi-gaussiens via des inégalités de concentration et de translation gaussiennes pour certaines familles de mesures de BOLTZMANN et lois de certains semi-groupes diffusifs en temps petit. Au-delà des résultats, l'accent est mis ici sur la méthode et la symétrie des arguments utilisés pour les bornes inférieures et supérieures.

Le second chapitre a pour objectif de présenter certains liens existant entre les mathématiques construites autour des inégalités de SOBOLEV logarithmiques abordées dans l'ouvrage [ABC⁺00] d'une part, et certaines inégalités faisant intervenir l'entropie de SHANNON ou l'information de FISHER d'autre part, qui jouent un très grand rôle dans ce que l'on appelle « théorie de l'information ». Cette théorie possède des liens naturels avec l'informatique et la théorie du signal. Cependant, nous nous intéresserons essentiellement ici 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 l'information de FISHER et l'entropie de SHANNON, 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 (inégalité de CRAMÉR-RAO), et en physique (principes d'incertitude de BECKNER-HIRSHMAN et de WEYL-HEISENBERG).

Le troisième chapitre présente un lien entre une forme inverse de l'inégalité de SOBOLEV logarithmique et le maximum gaussien de l'entropie exponentielle de SHANNON. Il y a donc un parallèle complet avec le lien bien connu entre l'inégalité de SOBOLEV logarithmique et ses traductions en théorie de l'information. Nous donnons également une preuve élémentaire de la forme inverse gaussienne de l'inégalité de SOBOLEV logarithmique via une inégalité sur l'espace à deux points, imitant en cela la preuve de GROSS pour la forme non inverse.

Enfin, le dernier chapitre consiste en une étude des inégalités de Poincaré et de Sobolev logarithmique pour des mesures de probabilités liées à des modèles conservatifs à spins continus en mécanique statistique. Inspiré des résultats récents de C. LANDIM, G. PANIZO et H.-T. YAU [LPY00] sur le trou spectral et l'inégalité de Sobolev logarithmique pour des systèmes conservatifs de spins non bornés, nous étudions des bornes *uniformes* dans ces inégalités pour la dynamique de GLAUBER associée à des hamiltoniens de la forme $\sum_{i=1}^n V(x_i) + V(M - x_1 - \dots - x_n)$. Plus précisément, nous examinons le cas où V est strictement convexe et, en suivant [LPY00], le cas où V est une perturbation bornée d'un potentiel quadratique. Par un argument de chemins élémentaire pour la marche aléatoire simple, les bornes uniformes pour la dynamique de GLAUBER entraînent, de façon limpide, la décroissance classique en L^{-2} pour la dynamique de KAWASAKI sur des cubes d'arête L d'un réseau de dimension d . Cependant, les arguments de nos preuves suivent de près et utilisent abondamment les estimées et la démarche de [LPY00].

Dans la suite de ce chapitre zéro, nous tentons de donner une présentation unifiée des principaux résultats obtenus dans chacun des quatre travaux. Nous commençons par une synthèse sur des inégalités fonctionnelles, qui constitue la section 0.1. Nous poursuivons ensuite par trois sections consacrées aux résultats obtenus. Chacune de ces sections se termine par des questions restées sans réponses, et qui constituent des perspectives de recherche future.

0.1 Inégalités fonctionnelles

Une mesure de probabilités μ sur $\mathbb{R}^n$ vérifie une inégalité de Poincaré lorsqu'il existe une constante positive $\mathcal{P}$ telle que pour toute fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$(0.1) \quad \mathbf{Var}_\mu(f) \leq \mathcal{P} \mathbf{E}_\mu(|\nabla f|^2).$$

Dans le même ordre d'idées, la mesure de probabilité μ vérifie une inégalité de Sobolev logarithmique lorsqu'il existe une constante positive $\mathcal{L}$ telle que pour toute fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$(0.2) \quad \mathbf{Ent}_\mu(f^2) \leq \mathcal{L} \mathbf{E}_\mu(|\nabla f|^2),$$

où $\mathbf{Ent}_\mu(f^2) := \mathbf{E}_\mu(f^2 \log f^2) - \mathbf{E}_\mu(f^2) \log \mathbf{E}_\mu(f^2)$. L'exemple emblématique est fourni par la loi gaussienne $\mathcal{N}(0, \text{Id}_n)$, qui vérifie les deux inégalités (0.1) et (0.2) avec les constantes optimales 1 et 2 respectivement. On vérifie facilement que la loi exponentielle satisfait (0.1) mais pas (0.2). L'inégalité de Sobolev logarithmique (0.2) est plus forte que celle de Poincaré (0.1) puisque l'on a

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

qui entraîne (0.1) avec $\mathcal{P} = \mathcal{L}/2$. Cette comparaison des constantes est optimale, comme le montre l'exemple de la gaussienne, et diffère de celle que pourrait laisser croire l'encadrement suivant valable pour toute fonction f lisse et positive,

$$\mathbf{E}_\mu(f) \mathbf{Ent}_\mu(f) \leq \mathbf{Var}_\mu(f) \leq \mathbf{Ent}_\mu(f^2),$$

qui se déduit de l'encadrement élémentaire $(\sqrt{x} - 1)^2 + x - 1 \leq x \log(x) \leq x^2 - x$ pour tout $x \geq 0$.

La mesure de probabilité μ vérifie une inégalité de BOBKOV lorsqu'il existe une constante positive $\mathcal{I}$ telle que pour toute fonction lisse $f : \mathbb{R}^n \rightarrow [0, 1]$,

$$(0.3) \quad \mathbf{I}(\mathbf{E}_\mu(f)) \leq \mathbf{E}_\mu\left(\sqrt{\mathbf{I}(f)^2 + \mathcal{I} |\nabla f|^2}\right),$$

où $\mathbf{I} : [0, 1] \rightarrow \mathbb{R}_+$ est la fonction isopérimétrique gaussienne donnée par

$$\mathbf{I}(u) = \Phi' \circ \Phi^{-1}(u),$$

où Φ est la fonction répartition de $\mathcal{N}(0, \text{Id}_n)$. À son tour, l'inégalité (0.3) est plus forte que (0.2). En effet, comme l'a judicieusement observé BECKNER, on retrouve (0.2) avec $\mathcal{L} = 2\mathcal{I}$ en posant $f = \varepsilon g^2$ dans (0.3) puis en utilisant le fait que

$$\mathbf{I}(u) \sim_0 u \sqrt{-2 \log u}.$$

Ici encore, l'inégalité (0.3) est vérifiée par la loi gaussienne standard $\mathcal{N}(0, \text{Id}_n)$ avec la constante optimale 1, cf. [Bob97].

En approchant les indicatrices de boréliens par des fonctions lisses, l'inégalité de BOBKOV (0.3) apparaît comme une forme fonctionnelle de l'inégalité isopérimétrique gaussienne,

$$\mathbf{I}(\mu(A)) \leq \mu_s(\partial A),$$

qui exprime le fait que parmi tous les boréliens A de mesure donnée, les demi-espaces sont ceux qui ont le plus petit bord au sens de la mesure de surface μ_s . D'un autre côté, les inégalités de POINCARÉ (0.1) et de SOBOLEV logarithmique (0.2) fournissent, lorsqu'elles sont appliquées à des fonctions de la forme $f = \exp(\lambda F)$ où F est lipschitzienne, des majorations exponentielles de $\mu(|F - \mathbf{E}_\mu(F)| \geq r)$. La majoration obtenue par (0.2), connue sous le nom de « concentration gaussienne », est de la forme $\exp(-r^2/\mathcal{L})$. Elle peut apparaître à son tour comme une forme faible de l'inégalité isopérimétrique gaussienne déduite de (0.3). Nous renvoyons par exemple à [Led99] pour ces aspects.

À ce stade, il est tout à fait légitime de s'interroger sur la validité des inégalités (0.1), (0.2) et (0.3) pour des mesures de probabilités μ non gaussiennes. On pense tout naturellement aux densités log-concaves, dont les lois gaussiennes sont des cas particuliers. Cette question trouve une réponse satisfaisante dans le critère de courbure de BAKRY et ÉMERY introduit dans la suite.

0.1.1 Inégalités locales et critères de courbure

Intéressons-nous à des mesures positives μ sur $\mathbb{R}^n$ qui s'écrivent sous la forme

$$\mu(dx) = \exp(-H(x)) dx,$$

où $H \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$. Lorsque cela est possible, on normalise μ en une mesure de probabilités en faisant intervenir un facteur multiplicatif $(Z_H)^{-1}$. La mesure μ apparaît comme la mesure invariante symétrique du processus de KOLMOGOROV $(X_t)_{t \geq 0}$ satisfaisant à l'É.D.S.

$$dX_t = \sqrt{2} dB_t - (\nabla H)(X_t) dt.$$

Cette équation décrit par exemple la vitesse X_t à l'instant t d'une particule qui subit à la fois une force d'agitation brownienne et une force de rappel associée au champ de potentiel H . Au processus de diffusion $(X_t)_{t \geq 0}$, on associe un semi-groupe de MARKOV $(\mathbf{P}_t)_{t \geq 0}$, défini pour tout $t \geq 0$, tout $x \in \mathbb{R}^n$ et toute fonction bornée $f : \mathbb{R}^n \rightarrow \mathbb{R}$ par

$$\mathbf{P}_t(f)(x) = \mathbf{E}(f(X_t) | X_0 = x).$$

Ce semi-groupe de contraction de $\mathbf{L}^2(\mathbb{R}^n, \mu, \mathbb{R})$ a pour générateur infinitésimal l'opérateur différentiel $\mathbf{L}$ défini par

$$\mathbf{L} := \Delta - \nabla H \cdot \nabla.$$

On a donc $\partial_t \mathbf{P}_t = \mathbf{L} \mathbf{P}_t = \mathbf{P}_t \mathbf{L}$. Il est alors bien connu que l'inégalité de Poincaré (0.1) de constante $\mathcal{P}$ pour μ est équivalente à l'existence d'un trou supérieur à $\mathcal{P}^{-1}$ dans le

spectre de $\mathbf{L}$, tandis que l'inégalité de SOBOLEV logarithmique (0.2) de constante $\mathcal{L}$ pour μ est équivalente (théorème de GROSS) à l'hypercontractivité du semi-groupe $(\mathbf{P}_t)_{t \geq 0}$:

$$\|\mathbf{P}_t\|_{2 \rightarrow q} \leq 1 \text{ pour } q \leq 1 + e^{4t/\mathcal{L}}.$$

Ce résultat s'obtient en remarquant que la dérivée en p de la norme $\|\cdot\|_p$ fait apparaître l'entropie. On déduit des inégalités (0.1) et (0.2) des convergences exponentielles vers l'équilibre, au sens de la variance pour l'inégalité de POINCARÉ (0.1)

$$\mathbf{Var}_\mu(\mathbf{P}_t(f)) \leq e^{-2t/\mathcal{P}} \mathbf{Var}_\mu(f),$$

et au sens de l'entropie pour l'inégalité de SOBOLEV logarithmique (0.2)

$$\mathbf{Ent}_\mu(\mathbf{P}_t(f)) \leq e^{-4t/\mathcal{L}} \mathbf{Ent}_\mu(f).$$

Par exemple, pour la variance, on écrit en utilisant l'inégalité de POINCARÉ (0.1),

$$\begin{aligned} \partial_t \mathbf{Var}_\mu(\mathbf{P}_t(f)) &= 2 \mathbf{E}_\mu(\mathbf{P}_t(f) \mathbf{L} \mathbf{P}_t(f)) \\ &= -2 \mathbf{E}_\mu(|\nabla \mathbf{P}_t(f)|^2) \leq -2 \mathcal{P}^{-1} \mathbf{Var}_\mu(\mathbf{P}_t(f)). \end{aligned}$$

On peut aisément faire de même pour l'inégalité de BOBKOV (0.3) et la fonctionnelle $\mathbf{I}(\mathbf{E}_\mu(\cdot)) - \mathbf{E}_\mu(\mathbf{I}(\cdot))$ en utilisant le fait que $\mathbf{I} \mathbf{I}'' = -1$.

Le critère de courbure de BAKRY et ÉMERY affirme que lorsqu'il existe une constante positive α telle que $\text{Hess}(H)(x) \geq \alpha \text{Id}_n$ pour tout $x \in \mathbb{R}^n$ au sens des formes quadratiques (i.e. H est uniformément strictement convexe, ou alternativement μ est uniformément log-concave), alors μ satisfait aux inégalités (0.1), (0.2) et (0.3) suivantes :

$$(0.4) \quad \mathbf{Var}_\mu(f) \leq \alpha^{-1} \mathbf{E}_\mu(|\nabla f|^2),$$

$$(0.5) \quad \mathbf{Ent}_\mu(f^2) \leq 2\alpha^{-1} \mathbf{E}_\mu(|\nabla f|^2),$$

et

$$(0.6) \quad \mathbf{I}(\mathbf{E}_\mu(f)) \leq \mathbf{E}_\mu\left(\sqrt{\mathbf{I}(f)^2 + \alpha^{-1} |\nabla f|^2}\right).$$

On retrouve la gaussienne standard $\mathcal{N}(0, \text{Id}_n)$ pour $\alpha = 1$. Par exemple, pour (0.4), on écrit en utilisant l'inégalité $|\nabla \mathbf{P}_t(f)|^2 \leq e^{-2\alpha t} \mathbf{P}_t(|\nabla f|^2)$, qui est une conséquence de la convexité uniforme de H ,

$$\begin{aligned} \mathbf{Var}_\mu(f) &= - \int_0^{+\infty} \partial_t \mathbf{E}_\mu((\mathbf{P}_t(f))^2) dt \\ &= -2 \int_0^{+\infty} \mathbf{E}_\mu(\mathbf{P}_t(f) \mathbf{L} \mathbf{P}_t(f)) dt \\ &= 2 \int_0^{+\infty} \mathbf{E}_\mu(|\nabla \mathbf{P}_t(f)|^2) dt \leq \alpha^{-1} \mathbf{E}_\mu(|\nabla f|^2). \end{aligned}$$

De façon plus générale et sous les mêmes hypothèses sur H , on obtient ces inégalités pour les lois du semi-groupe [Led00]. Plus précisément, pour tout temps $t > 0$ et toute fonction f raisonnable,

$$(0.7) \quad \mathbf{P}_t(f^2) - \mathbf{P}_t(f)^2 \leq \alpha(t) \mathbf{P}_t(|\nabla f|^2),$$

$$(0.8) \quad \mathbf{P}_t(f^2 \log f^2) - \mathbf{P}_t(f^2) \log \mathbf{P}_t(f^2) \leq 2\alpha(t) \mathbf{P}_t(|\nabla f|^2),$$

et

$$(0.9) \quad \mathbf{I}(\mathbf{P}_t(f)) \leq \mathbf{P}_t\left(\sqrt{\mathbf{I}(f)^2 + \alpha(t) |\nabla f|^2}\right).$$

où $\alpha(t) := \alpha^{-1}(1 - e^{-2\alpha t})$. Les inégalités locales (0.7), (0.8) et (0.9) fournissent les inégalités (0.1), (0.2) et (0.3) pour la mesure invariante μ par ergodicité en faisant tendre t vers l'infini. Il est important de comprendre que la réciproque n'a pas lieu en général, et que les inégalités pour la mesure invariante μ n'entraînent pas d'inégalités locales en toute généralité. D'autre part, on montre que les inégalités locales ont lieu pour tout t avec la constante $\alpha(t)$ si et seulement si H est uniformément convexe de constante $\alpha > 0$. Plus précisément, l'occurrence de l'une d'entre elle pour tout $t > 0$ est équivalente à la propriété de commutation suivante,

$$|\nabla \mathbf{P}_t(f)| \leq e^{-\alpha t} \mathbf{P}_t(|\nabla f|),$$

elle même équivalente à $\forall x, \text{Hess}(H)(x) \geq \alpha \text{Id}_n$. A contrario, les inégalités pour la mesure invariante μ peuvent avoir lieu pour des H non convexes et ne sont donc pas équivalentes à la convexité de H . L'inégalité de Poincaré (0.1) pour la mesure invariante μ fait figure d'exception puisque l'on dispose d'un critère sous forme d'équivalence : la mesure de probabilité μ vérifie (0.1) avec une constante $\mathcal{P} > 0$ si et seulement si, pour toute fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$(0.10) \quad \mathcal{P} \mathbf{E}_\mu((\mathbf{L}f)^2) \geq \mathbf{E}_\mu(|\nabla f|^2).$$

Lorsque $\text{Hess}(H) \geq \alpha \text{Id}_n$ avec $\alpha > 0$, cette inégalité est bien entendu satisfaite avec $\mathcal{P} = \alpha^{-1}$ puisque l'on a toujours

$$\mathbf{E}_\mu((\mathbf{L}f)^2) = \mathbf{E}_\mu(\|\text{Hess}(f)\|_2^2) + \mathbf{E}_\mu(\nabla f \cdot \text{Hess}(H) \nabla f).$$

Nous avons volontairement choisi de ne pas parler de critère de courbure-dimension en terme d'opérateurs $\mathbf{\Gamma}$ et $\mathbf{\Gamma}_2$, qui soutend l'analyse de la géométrie des semi-groupes de MARKOV car nous nous restreignons ici à des diffusions particulières sur $\mathbb{R}^n$. Nous renvoyons pour cela à la synthèse faite dans [Led00] et [ABC⁺00, chapitre 5].

0.1.2 Inégalités inverses

Soit μ une mesure de probabilités sur $\mathbb{R}^n$ telle que

$$\mu(dx) = (Z_H)^{-1} \exp(-H(x)) dx$$

pour une certaine fonction $H \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$. Si le semi-groupe de diffusion associé est ergodique et s'il existe une constante $\beta > 0$ telle que $\|\text{Hess}(H)\| \leq \beta$ au sens des applications linéaires, alors on obtient (cf. [Led00]) les inégalités suivantes, qui constituent des formes inverses de (0.1), (0.2) et (0.3) :

$$(0.11) \quad \frac{1}{\beta} |\mathbf{E}_\mu(\nabla f)|^2 \leq \mathbf{Var}_\mu(f),$$

$$(0.12) \quad \frac{1}{2\beta} |\mathbf{E}_\mu(\nabla f)|^2 \leq \mathbf{E}_\mu(f) \mathbf{Ent}_\mu(f),$$

et

$$(0.13) \quad \sqrt{\mathbf{E}_\mu(\mathbf{I}(f))^2 + \frac{1}{\beta} |\mathbf{E}_\mu(\nabla f)|^2} \leq \mathbf{I}(\mathbf{E}_\mu(f)).$$

En particulier, elles ont toutes lieu pour la loi gaussienne $\mathcal{N}(0, \text{Id}_n)$ avec $\beta = 1$. Il est facile de voir que (0.12) entraîne (0.11) mais la réciproque n'est pas claire. Cependant, les inégalités (0.13) et (0.12) sont équivalentes, aux constantes près, à l'inégalité de BOBKOV suivante (cf. [BCEF01] et [Bob99]) :

$$(0.14) \quad |\mathbf{E}_\mu(\nabla f)| \leq \sqrt{\beta} \mathbf{I}(\mathbf{E}_\mu(f)).$$

Pour la mesure gaussienne, cette inégalité exprime le fait que parmi tous les boréliens de mesure donnée, les demi-espaces sont ceux qui maximisent la moyenne de leur barycentre. En imitant les preuves des inégalités non inverses, nous pouvons établir des versions locales de (0.11), (0.13) et (0.12). Pour tout $t > 0$ et toute fonction raisonnable f , on a en notant $\beta(t) := \beta^{-1}(1 - e^{-2\beta t})$:

$$(0.15) \quad \beta(t) |\mathbf{P}_t(\nabla f)|^2 \leq \mathbf{P}_t(f^2) - \mathbf{P}_t(f)^2,$$

$$(0.16) \quad \frac{\beta(t)}{2} |\mathbf{P}_t(\nabla f)|^2 \leq \mathbf{P}_t(f) (\mathbf{P}_t(f \log f) - \mathbf{P}_t(f) \log \mathbf{P}_t(f)),$$

et

$$(0.17) \quad \sqrt{\mathbf{P}_t(\mathbf{I}(f))^2 + \beta(t) |\mathbf{P}_t(\nabla f)|^2} \leq \mathbf{I}(\mathbf{P}_t(f)).$$

Ces inégalités locales redonnent les inégalités (0.11), (0.13) et (0.12) pour la mesure invariante, par ergodicité en faisant tendre le temps t vers l'infini. La preuve des inégalités (0.13) et (0.17) fait appel aux relations fondamentales $\mathbf{I}\mathbf{I}'' = -1$ et

$$\nabla \mathbf{L} = \mathbf{L} \nabla - \text{Hess}(H) \nabla,$$

déjà utilisées dans la preuve de (0.3) et (0.9). Par exemple, pour établir (0.17), il suffit de montrer que

$$J(s) = \mathbf{P}_s(\mathbf{I}(\mathbf{P}_{t-s}(f)))^2 + \beta(s) |\mathbf{P}_s(\nabla \mathbf{P}_{t-s}(f))|^2$$

décroit de $s = 0$ à $s = t$. Un petit calcul utilisant $\nabla \mathbf{L} = \mathbf{L} \nabla - \text{Hess}(H) \nabla$ montre que :

$$\begin{aligned} J'(s) = & 2\mathbf{P}_s(\mathbf{I}(\mathbf{P}_{t-s}(f))) (\mathbf{L}\mathbf{P}_s(\mathbf{I}(\mathbf{P}_{t-s}(f))) - \mathbf{P}_s(\mathbf{I}'(\mathbf{P}_{t-s}(f))\mathbf{L}\mathbf{P}_{t-s}(f))) \\ & + 2e^{-2\beta t} |\mathbf{P}_s(\nabla \mathbf{P}_{t-s}(f))|^2 \\ & + 2\beta(s) \langle \mathbf{P}_s(\nabla \mathbf{P}_{t-s}(f)), \mathbf{P}_s(\text{Hess}(H) \nabla \mathbf{P}_{t-s}(f)) \rangle. \end{aligned}$$

En posant $F = \mathbf{P}_{t-s}(f)$ et en utilisant $\mathbf{I}\mathbf{I}'' = -1$, il vient :

$$\mathbf{L}\mathbf{I}(F) - \mathbf{I}(F)\mathbf{L}F = \mathbf{I}''(F)|\nabla F|^2 = -\frac{|\nabla F|^2}{\mathbf{I}(F)},$$

et donc :

$$\frac{1}{2}J'(s) = -\mathbf{P}_s(\mathbf{I}(F)) \mathbf{P}_s\left(\frac{|\nabla F|^2}{\mathbf{I}(F)}\right) + e^{-2\beta t} |\mathbf{P}_s(\nabla F)|^2 + \beta(s) \langle \mathbf{P}_s(\nabla F), \mathbf{P}_s(\text{Hess}(H) \nabla F) \rangle.$$

Or par hypothèse sur $\text{Hess}(H)$ et par l'inégalité de CAUCHY-SCHWARZ :

$$\langle \mathbf{P}_s(\nabla F), \mathbf{P}_s(\text{Hess}(H) \nabla F) \rangle \leq \beta \mathbf{P}_s(|\nabla F|^2)$$

et d'autre part, toujours par l'inégalité de CAUCHY-SCHWARZ :

$$|\mathbf{P}_s(\nabla F)|^2 \leq \mathbf{P}_s(|\nabla F|^2) \leq \mathbf{P}_s(\mathbf{I}(F)) \mathbf{P}_s\left(\frac{|\nabla F|^2}{\mathbf{I}(F)}\right)$$

d'où le résultat car $\beta(t) = (1 - e^{-2\beta t})/\beta$.

0.2 Méthodes fonctionnelles pour PGD quasi-gaussiens

L'inégalité de BOBKOV (0.3) de constante $\mathcal{I} = \alpha^{-1}$ entraîne, en l'appliquant aux indicatrices d'ensembles, une propriété isopérimétrique et de concentration qui affirme que pour tout borélien C et tout réel $r > 0$ on a

$$\mu(C + B_2(0, r)) \geq \Phi(\Phi^{-1}(\mu(C)) + \sqrt{\alpha} r),$$

où $B_2(0, r)$ est la boule euclidienne fermée de centre l'origine et de rayon $r > 0$. D'un autre côté, l'inégalité (0.14) entraîne que pour tout borélien C et tout vecteur h on a

$$\Phi(\Phi^{-1}(\mu(C)) + \sqrt{\beta} \|h\|_2) \geq \mu(C + h).$$

Ces deux propriétés permettent alors d'établir le théorème suivant :

Théorème 0.2.1. *Soit une famille de mesures de probabilités $(\mu_\varepsilon, \varepsilon > 0)$ sur $\mathbb{R}^n$ associées aux hamiltoniens $(H_\varepsilon, \varepsilon > 0)$ dans $\mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$, convergeant étroitement vers δ_0 quand ε tend vers 0. Supposons que pour tout $\varepsilon > 0$, il existe des réels α_ε et β_ε tels que $0 < \alpha_\varepsilon \text{Id}_n \leq \text{Hess}(H_\varepsilon) \leq \beta_\varepsilon \text{Id}_n$ uniformément sur $\mathbb{R}^n$ au sens des formes quadratiques avec $\lim_{\varepsilon \rightarrow 0^+} \alpha_\varepsilon = +\infty$. Alors pour tout borélien C on a*

$$-\inf_{\text{int}(C)} \mathbf{J} \leq \liminf_{\varepsilon \rightarrow 0} \beta_\varepsilon^{-1} \log \mu_\varepsilon(C) \leq \limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \log \mu_\varepsilon(C) \leq -\inf_{\text{adh}(C)} \mathbf{J}$$

où $\mathbf{J} := \frac{1}{2} \|\cdot\|_2^2$.

Le cas typique est celui donné par la contraction d'une loi de probabilité d'hamiltonien H vérifiant $0 < \alpha \text{Id}_n \leq \text{Hess}(H) \leq \beta \text{Id}_n$ uniformément sur $\mathbb{R}^n$ au sens des formes quadratiques. On a alors $H_\varepsilon(\cdot) = H(\varepsilon^{-1} \cdot)$ avec $\alpha_\varepsilon = \varepsilon^{-2} \alpha$ et $\beta_\varepsilon = \varepsilon^{-2} \beta$. On retrouve ainsi le PGD gaussien classique en prenant $H(\cdot) = \frac{1}{2} \|\cdot\|_2^2$ et $\alpha = \beta = 1$. On peut en déduire sans difficultés un PGD sans topologie. Les versions locales des inégalités fonctionnelles utilisées permettent, selon le même schéma, d'établir le résultat suivant.

Théorème 0.2.2. *Soit $(\mathbf{P}_t)_{t \geq 0}$ un semi-groupe de diffusion sur $\mathbb{R}^n$ de générateur infinitésimal $\mathbf{L} := \Delta - \nabla H \cdot \nabla$ tel que $\text{Hess}(H)$ soit uniformément borné en tant qu'opérateur linéaire sur $\mathbb{R}^n$. Alors, pour tout borélien C et tout x dans $\mathbb{R}^n$ on a*

$$-\inf_{\text{int}(C)} \mathbf{J}_x \leq \liminf_{t \rightarrow 0} 2t \log \mathbf{P}_t(C)(x) \leq \limsup_{t \rightarrow 0} 2t \log \mathbf{P}_t(C)(x) \leq -\inf_{\text{adh}(C)} \mathbf{J}_x$$

où $\mathbf{J}_x := \frac{1}{2} \|x - \cdot\|_2^2$.

Ainsi, on obtient, de façon très simple, un résultat de même nature que celui de [Var67]. On aurait pu faire de même en utilisant l'inégalité de SOBOLEV logarithmique (0.2) en lieu et place de l'inégalité isopérimétrique de BOBKOV (0.3), puisque (0.2) pour μ entraîne de la concentration gaussienne pour μ .

Perspectives

En prolongeant les inégalités de BOBKOV (0.3) et (0.13) à la loi de tout le processus, il devrait être possible d'énoncer des versions fonctionnelles des résultats obtenus.

0.3 Entropie de Shannon et information de Fisher

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

$$\mathbf{N}(X) := (2\pi e)^{-1} e^{\frac{2}{n}\mathbf{H}(X)},$$

où $\mathbf{H}(X) := -\mathbf{Ent}_{dx}(f)$ est l'entropie de SHANNON de X . L'entropie exponentielle de la loi gaussienne centrée de matrice de covariance K vaut donc

$$\mathbf{N}(\mathcal{N}(0, K)) = |K|^{1/n}.$$

De même que pour l'entropie de SHANNON, ces lois gaussiennes réalisent le maximum de $\mathbf{N}$ à covariance fixée, de sorte que si X a pour covariance K , alors

$$\mathbf{N}(X) \leq |K|^{1/n}.$$

À covariance fixée, la loi gaussienne est donc celle qui « contient le plus d'incertitude », au sens de l'entropie. Ainsi, de la même manière que $|K|^{1/n}$ représente le « rayon moyen » de la matrice K , $\mathbf{N}(X)$ représente en quelque sorte le « rayon d'incertitude » du vecteur aléatoire X .

Une deuxième quantité importante 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 par

$$\mathbf{J}(X) := \int |\nabla \log f|^2 f \, dx.$$

Elle apparaît comme la trace de la matrice d'information de FISHER $\mathbb{J}(X)$, bien connue en statistique et définie par

$$\mathbb{J}(X) := \int (\nabla \log f)(\nabla \log f)^\top f \, dx.$$

L'inégalité de SOBOLEV logarithmique (0.2) pour la loi gaussienne $\mathcal{N}(0, t\text{Id}_n)$ exprime que pour toute fonction f dérivable de $\mathbb{R}^n$ dans $\mathbb{R}$,

$$\mathbf{Ent}_{\mathcal{N}(0, t\text{Id}_n)}(f^2) \leq 2t \mathbf{E}_{\mathcal{N}(0, t\text{Id}_n)}(|\nabla f|^2).$$

Par un simple changement de fonction, cette inégalité est équivalente à l'inégalité suivante sur la mesure de LEBESGUE :

$$\int g^2 \log g^2 \, dx \leq \frac{n}{2} \log \left[\frac{2}{e\pi n} \int |\nabla g|^2 \, dx \right],$$

valable pour toute fonction g à décroissance rapide sur $\mathbb{R}^n$ telle que g^2 soit une densité pour la mesure de LEBESGUE. En posant $g = \sqrt{f}$, cette inégalité s'écrit alors à l'aide des quantités introduites

$$\mathbf{N}(X)\mathbf{J}(X) \geq n.$$

En appliquant cette inégalité au vecteur aléatoire $\mathbb{J}(X)^{1/2}X$, on obtient la forme suivante,

$$\mathbf{N}(X)|\mathbb{J}(X)|^{1/n} \geq 1,$$

qui redonne à son tour la précédente par l'inégalité arithmético-géométrique sur le spectre de $\mathbb{J}(X)$. Ainsi, l'inégalité de SOBOLEV logarithmique (0.12) pour les lois gaussiennes est équivalente à l'inégalité $\mathbf{N}(X)\mathbf{J}(X) \geq n$, connue depuis les années 1950 en théorie de l'information. Les objets $\mathbf{N}$ et $\mathbf{J}$ jouent un très grand rôle en théorie de l'information. Le chapitre 2 (page 27) constitue une synthèse sur leur multiples liens avec des objets et inégalités fonctionnelles en analyse. Nous y donnons en particulier une démonstration courte de la concavité de l'entropie exponentielle :

$$\partial_t^2 \mathbf{N}(X + \sqrt{t} Z) \leq 0,$$

où Z est un vecteur aléatoire gaussien standard indépendant du vecteur aléatoire X . En d'autres termes, l'entropie exponentielle $\mathbf{N}$ est concave par rapport à la variance d'une perturbation additive normale indépendante. Si $(\mathbf{P}_t)_{t \geq 0}$ désigne le semi-groupe de la chaleur associé au mouvement brownien, cette propriété peut s'écrire sous la forme $\partial_t (\mathbf{J}(\mathbf{P}_t(f))^{-1} \geq n^{-1}$, et s'établit alors simplement via l'inégalité

$$\|\text{Hess}(f)\|_2^2 \geq n^{-1} (\Delta f)^2.$$

Les spécialistes du « critère $\mathbf{I}_2$ » auront reconnu la partie dimensionnelle du critère de courbure-dimension en courbure nulle pour le laplacien usuel.

Nous montrons, dans le chapitre 3 (page 63) que le maximum gaussien de l'entropie de SHANNON est équivalent à la forme inverse (0.12) de l'inégalité de SOBOLEV logarithmique pour les lois gaussiennes. Plus précisément, on a l'énoncé suivant, qui établit un parallèle complet avec l'équivalence entre l'inégalité de SOBOLEV logarithmique gaussienne (0.2) et sa réécriture en terme des objets $\mathbf{N}$ et $\mathbf{J}$:

Théorème 0.3.1. *Si γ_n désigne la loi $\mathcal{N}(0, \text{Id}_n)$ et λ_n la mesure de LEBESGUE sur $\mathbb{R}^n$, alors les assertions suivantes sont équivalentes et vraies :*

(i) *Pour toute fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$,*

$$|\mathbf{E}_{\gamma_n}(\nabla f)|^2 \leq 2 \mathbf{Ent}_{\gamma_n}(f) \mathbf{E}_{\gamma_n}(f).$$

(ii) *Pour toute fonction lisse $g : \mathbb{R}^n \rightarrow \mathbb{R}^+$,*

$$-\mathbf{Ent}_{\lambda_n}(g) \leq \frac{n}{2} \log \left[\frac{2\pi e}{n} \text{Tr } K(g) \right].$$

(iii) *Pour tout vecteur aléatoire de $\mathbb{R}^n$ à densité lisse,*

$$n\mathbf{N}(X) \leq \text{Tr } K(X).$$

(iv) Pour tout vecteur aléatoire de $\mathbb{R}^n$ à densité lisse,

$$\mathbf{N}(X) \leq |K(X)|^{1/n}.$$

La forme inverse (0.12) de l'inégalité de SOBOLEV logarithmique gaussienne peut être obtenue par tensorisation à partir de l'inégalité à deux points suivante,

$$2p^2q^2(f(1) - f(0))^2 \leq \mathbf{Ent}_{\beta_p}(f) \mathbf{E}_{\beta_p}(f),$$

qui n'est rien d'autre qu'une version optimale de l'inégalité de PINSKER-CSISZÀR-KULLBACK pour la loi de BERNOULLI β_p . De manière générale, l'inégalité (0.12) se tensorise très facilement au moyen de l'inégalité

$$\mathbf{Ent}_{\mu}(f) \geq \sum_{i=1}^n \mathbf{Ent}_{\mu_i} \left(\int f d\mu_{\setminus i} \right),$$

valable pour toute mesure $\mu = \otimes_{i=1}^n \mu_i$, produit de n mesures positives μ_i et toute fonction positive f sur l'espace produit où $\mu_{\setminus i}$ désigne le produit des mesures μ_j pour $j \neq i$. De plus, l'égalité n'est réalisée que lorsque $f = (\int f d\mu) \otimes_{i=1}^n \int f d\mu_{\setminus i}$, μ -presque partout. Lorsque les mesures μ_i sont de probabilités, la fonction constante 1 est intégrable et l'inégalité précédente appliquée à la fonction $1 + \varepsilon f$ où f est bornée fournit :

$$\mathbf{Var}_{\mu}(f) \geq \sum_{i=1}^n \mathbf{Var}_{\mu_i} \left(\mathbf{E}_{\mu_{\setminus i}}(f) \right),$$

qui permet de tensoriser directement (0.11). Ces propriétés de tensorisation de la variance et de l'entropie apparaissent comme les contreparties inverses des inégalités de tensorisation présentées dans [ABC⁺00, chap. 1] :

$$(0.18) \quad \mathbf{Var}_{\mu}(f) \leq \sum_{i=1}^n \mathbf{E}_{\mu}(\mathbf{Var}_{\mu_i}(f)),$$

et

$$(0.19) \quad \mathbf{Ent}_{\mu}(f) \leq \sum_{i=1}^n \mathbf{E}_{\mu}(\mathbf{Ent}_{\mu_i}(f)).$$

Perspectives

L'invariance par translation et dilatation de la mesure de LEBESGUE joue un rôle important dans les propriétés des objets qui interviennent en théorie de l'information. Cependant, certaines restent valables pour d'autres mesures positives, et il serait intéressant de pouvoir en généraliser d'autres. L'inégalité de BLACHMAN-STAM (2.12) (cf. page 41) par exemple reste-t-elle valable pour des mesures plus générales ?

0.4 Glauber vs Kawasaki pour certains systèmes de spins

En mécanique statistique, les inégalités (0.1) et (0.2) établies sur $\mathbb{R}^n$ correspondent à des descriptions microscopiques, et la dépendance en n de leurs constantes $\mathcal{P}$ et $\mathcal{L}$ joue un rôle important pour le comportement des objets macroscopiques. Rappelons à ce propos que les inégalités (0.1) et (0.2) sont stables par perturbation et tensorisation :

- Si μ et ν sont deux mesures de probabilité sur $\mathbb{R}^n$ et $\mathbb{R}^m$ vérifiant (0.1) avec les constantes $\mathcal{P}_\mu$ et $\mathcal{P}_\nu$, alors $\mu \otimes \nu$ vérifie (0.1) sur $\mathbb{R}^{n+m}$ avec la constante $\max(\mathcal{P}_\mu, \mathcal{P}_\nu)$. Il en est de même pour l'inégalité (0.2).
- Si la mesure de probabilités μ vérifie (0.1) sur $\mathbb{R}^n$ avec la constante $\mathcal{P}$, et si $B \in \mathcal{C}_b(\mathbb{R}^n, \mathbb{R})$, alors la mesure de probabilités $(Z_B)^{-1} \exp(B(x)) d\mu(x)$ vérifie (0.1) avec la constante $\exp(2\text{osc}(B))\mathcal{P}$. Il en est de même pour l'inégalité (0.2).

Considérons par exemple la mesure de probabilité ν sur $\mathbb{R}$ définie par

$$d\nu(x) = Z^{-1} \exp(W(x) + F(x)) dx,$$

où W est uniformément convexe (i.e. $W''(x) \geq \alpha > 0$ pour tout x) et F est bornée. En vertu du critère de courbure et des propriétés de perturbation et de tensorisation, la mesure de probabilités $\nu^{\otimes n}$ sur $\mathbb{R}^n$ vérifie les inégalités (0.1) et (0.2) avec des constantes qui dépendent de α et $\|F\|_\infty$ mais pas de n . Il en va tout autrement si l'on considère des mesures qui ne sont pas produit, pour lesquelles il est alors légitime de s'interroger sur le comportement dimensionnel et perturbatif des constantes dans (0.1) et (0.2).

Soit $V \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$, $M \in \mathbb{R}$ et σ_M la mesure de probabilité sur $\mathbb{R}^n$ définie par

$$d\sigma_M(x_1, \dots, x_n) := (Z_{\sigma_M})^{-1} \exp\left(-\sum_{i=1}^n V(x_i) - V\left(M - \sum_{i=1}^n x_i\right)\right) dx_1 \cdots dx_n.$$

Cette mesure σ_M est liée à une mesure conditionnée. En effet, soit μ la mesure de probabilités sur $\mathbb{R}^{n+1}$ définie par

$$d\mu(x_1, \dots, x_{n+1}) := (Z_V)^{-1} \exp\left(\sum_{i=1}^n V(x_i)\right) dx_1 \cdots dx_{n+1},$$

et μ_M la mesure de probabilité conditionnée donnée par

$$\mu_M := \mu\left(\cdot \mid \sum_{i=1}^{n+1} x_i = M\right).$$

On a alors pour toute fonction $f \in \mathcal{C}_b(\mathbb{R}^{n+1}, \mathbb{R})$,

$$\mathbf{E}_{\mu_M}(f) = \int_{\mathbb{R}^n} f\left(x_1, \dots, x_n, M - \sum_{i=1}^n x_i\right) d\sigma_M(x_1, \dots, x_n).$$

Les mesures σ_M , μ et μ_M sont échangeables, et l'on voit bien que la coordonnée x_{n+1} ne joue par de rôle particulier. Ainsi, si σ_M vérifie les inégalités (0.1) et (0.2), alors μ_M vérifiera l'inégalité suivante, valable pour toute fonction lisse $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$,

$$\mathbf{Var}_{\mu_M}(f) \leq \frac{\mathcal{P}}{n+1} \sum_{i,j=1}^{n+1} \mathbf{E}_{\mu_M}(|\partial_i f - \partial_j f|^2),$$

et

$$\mathbf{Ent}_{\mu_M}(f^2) \leq \frac{\mathcal{L}}{n+1} \sum_{i,j=1}^{n+1} \mathbf{E}_{\mu_M}(|\partial_i f - \partial_j f|^2).$$

Intéressons nous à présent au cas $\mathbb{R}^{n+1} \simeq \mathbb{R}^\Lambda$, où $\Lambda = \{1, \dots, L\}^d$ est une boîte finie du réseau $\mathbb{Z}^d$. Un argument de chemins pour la marche aléatoire simple fournit la comparaison suivante valable pour tout $a \in \mathbb{R}^\Lambda$,

$$\frac{1}{|\Lambda|} \sum_{i,j \in \Lambda} (a_i - a_j)^2 \leq C_d L^2 \sum_{\substack{i,j \in \Lambda \\ |i-j|=1}} (a_i - a_j)^2,$$

où C_d ne dépend que de d . Puisque $|\Lambda| = n+1$, on déduit immédiatement des inégalités précédentes les deux suivantes,

$$\mathbf{Var}_{\mu_M}(f) \leq C_d L^2 \mathcal{P} \sum_{\substack{i,j \in \Lambda \\ |i-j|=1}} \mathbf{E}_{\mu_M}(|\partial_i f - \partial_j f|^2),$$

et

$$\mathbf{Ent}_{\mu_M}(f^2) \leq C_d L^2 \mathcal{L} \sum_{\substack{i,j \in \Lambda \\ |i-j|=1}} \mathbf{E}_{\mu_M}(|\partial_i f - \partial_j f|^2).$$

On parle d'inégalités sur la dynamique de KAWASAKI, par opposition à celle de GLAUBER qui fait intervenir le gradient euclidien $|\nabla f|^2$. Supposons à présent qu'il existe $\alpha > 0$ tel que $V''(u) \geq \alpha$ pour tout $u \in \mathbb{R}$. Le critère de courbure entraîne alors immédiatement que σ_M est uniformément log-concave de constante α , et donc σ_M satisfait aux inégalités de POINCARÉ et de SOBOLEV logarithmiques suivantes,

$$\mathbf{Var}_{\sigma_M}(f) \leq \alpha^{-1} \mathbf{E}_\mu(|\nabla f|^2),$$

et

$$\mathbf{Ent}_{\sigma_M}(f^2) \leq 2\alpha^{-1} \mathbf{E}_\mu(|\nabla f|^2).$$

Les constantes ne dépendent donc pas de n et M . Le procédé présenté précédemment fournit alors de façon limpide le comportement en L^2 des constantes pour la dynamique de KAWASAKI associée à μ_M sur la boîte $\{1, \dots, L\}^d$.

Il est naturel de chercher à étendre l'utilisation de ce procédé à des potentiels V non convexes. Il suffit pour cela d'établir des inégalités pour la dynamique de GLAUBER associée à σ_M . Comme l'a remarqué Ivan GENTIL, une utilisation judicieuse du critère de courbure sous sa forme d'équivalence (0.10) pour l'inégalité de POINCARÉ permet déjà d'obtenir le résultat perturbatif suivant :

Proposition 0.4.1. *Supposons que V soit de la forme $V(u) = u^2/2 + F(u)$ où $F : \mathbb{R} \rightarrow \mathbb{R}$. Alors, pour $\|F\|_\infty$ assez petit, il existe une constante $\mathcal{P} > 0$ dépendant uniquement de $\|F\|_\infty$ telle que pour tout n , M et toute fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}$,*

$$\mathbf{Var}_{\sigma_M}(f) \leq \mathcal{P} \mathbf{E}_{\sigma_M}(|\nabla f|^2).$$

Ceci reste valable si l'on remplace dans la définition de V la fonction quadratique $u \mapsto u^2/2$ par une fonction $u \mapsto \Phi(u)$ telle qu'il existe des constantes α et β vérifiant $0 < \alpha \leq \beta \leq 2\alpha$ et $\alpha \leq \Phi''(u) \leq \beta$ pour tout $u \in \mathbb{R}$. La constante $\mathcal{P}$ s'écrit alors

$$\frac{\exp(2 \operatorname{osc}(F))}{2\alpha \exp(-2 \operatorname{osc}(F)) - \beta}$$

pour $\operatorname{osc}(F) < \log \sqrt{2\alpha/\beta}$. Le traitement de l'inégalité de SOBOLEV logarithmique pose problème car l'on ne dispose pas de critère de courbure sous forme d'équivalence comme pour celle de POINCARÉ. La preuve de ce résultat perturbatif est calquée sur une méthode due à B. HELFFER dans le cadre de systèmes de spins avec conditions au bord, pour lesquels la mesure associée n'est pas échangeable. Cette méthode n'est sans doute pas adaptée à notre cas puisque toute réduction de F dans le terme d'interaction $V(M - x_1 - \dots - x_n)$ affecte le terme produit $V(x_1) + \dots + V(x_n)$ car la mesure possède une symétrie spéciale. On peut cependant espérer que cette symétrie supplémentaire entraîne un comportement non perturbatif, comme dans certains modèles « champ moyen ». Dans cette direction, C. LANDIM, G. PANIZO et H.-T. YAU ont récemment obtenu [LPY00] le comportement en L^2 des constantes pour la dynamique de KAWASAKI associée à μ_M . Leur preuve repose sur la décomposition markovienne de LU et YAU et sur un théorème central limite local. Le but du chapitre 4 page 71 est de montrer qu'en réalité, le facteur L^2 peut être obtenu de la même façon que pour le cas purement convexe via un argument de chemins. La preuve que nous proposons repose cependant de façon cruciale sur les estimées qui sont à la base de [LPY00]. Notre résultat principal est le suivant :

Théorème 0.4.2. *Supposons que V soit de la forme $V(u) = u^2/2 + F(u)$. Alors, si F est bornée et lipschitzienne, il existe une constante $\mathcal{P} > 0$ dépendant uniquement de $\|F\|_\infty$ et $\|F'\|_\infty$ telle que pour tout n , tout M et toute fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}$,*

$$\mathbf{Var}_{\sigma_M}(f) \leq \mathcal{P} \mathbf{E}_{\sigma_M}(|\nabla f|^2).$$

De plus, si F'' est également bornée, il existe une constante $\mathcal{L} > 0$ dépendant uniquement de $\|F\|_\infty$, $\|F'\|_\infty$ et $\|F''\|_\infty$ telle que pour tout n , tout M et toute fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\mathbf{Ent}_{\sigma_M}(f^2) \leq \mathcal{L} \mathbf{E}_{\sigma_M}(|\nabla f|^2).$$

La preuve reprend également le découpage en boîtes de taille contrôlée utilisé dans [LPY00], qui permet de tirer partie d'un théorème central limite local via une sorte de propagation du chaos donnant la convergence des marges de la mesure canonique vers celles de la mesure grand canonique produit. Les propriétés de tensorisation (0.18) et (0.19) sont remplacées par les suivantes :

$$\mathbf{Var}_{\sigma_M}(f) = \sum_{k=1}^n \mathbf{E}_{\sigma_M}(\mathbf{Var}_{\sigma^{(k-1)}}(f_k)),$$

et

$$\mathbf{Ent}_{\sigma_M}(f) = \sum_{k=1}^n \mathbf{E}_{\sigma_M}(\mathbf{Ent}_{\sigma^{(k-1)}}(f_k)),$$

où $\sigma^{(k)}$ désigne la mesure σ_M sachant $x_1, \dots, x_k$ et f_k l'espérance conditionnelle $\mathbf{E}_{\sigma_M}(f|x_1, \dots, x_k) = \mathbf{E}_{\sigma^{(k)}}(f)$. La structure particulière de σ_M fait que $\sigma^{(k)}$ n'est rien d'autre que $\sigma_{M-x_1-\dots-x_k}(dx_{k+1}, \dots, dx_n)$.

Remarque 0.4.3. Le facteur L^2 apparaît tout naturellement dans certains modèles d'interface. Soit par exemple $\Lambda := \{1, \dots, L\}^d \subset \mathbb{Z}^d$ et ν_Λ la mesure de probabilité sur $\mathbb{R}^\Lambda$ définie par

$$\nu_\Lambda(dx) = \exp(-H_\Lambda(x)) dx,$$

où

$$H_\Lambda(x) := \sum_{\substack{\{i,j\} \cap \Lambda \neq \emptyset \\ i \sim j}} V(x_i - x_j),$$

où $V \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$ et $x_i = \omega_i$ si $i \notin \Lambda$, ω étant fixé dans $\mathbb{Z}^d$. Un petit calcul donne pour tous $x \in \mathbb{R}^\Lambda$ et $i, j \in \Lambda$:

$$\partial_{ij}^2 H(x) = \begin{cases} \sum_{k \in \mathbb{Z}^d, k \sim i} V''(x_i - x_k) & \text{si } i = j \\ -V''(x_i - x_j) & \text{si } i \sim j \\ 0 & \text{sinon} \end{cases}$$

avec toujours $x_k = \omega_k$ si $k \notin \Lambda$. Ainsi, si $V'' \geq 0$ alors $\nabla^2 H \geq 0$ d'après le théorème de GERSHGORIN-HADAMARD. Il en découle que lorsque $0 < \alpha \leq V'' \leq \beta$ avec $\alpha \in]0, +\infty[$ et $\beta \in]0, +\infty]$, alors on a au sens des matrices symétriques

$$-2d\alpha \Delta_\Lambda \leq \nabla^2 H \leq -2d\beta \Delta_\Lambda,$$

où Δ_Λ est le laplacien discret sur Λ défini pour tout i, j dans Λ par

$$-2d(\Delta_\Lambda)_{ij} = 1 - \delta_{i \sim j}.$$

La diagonalisation de Δ_Λ s'obtient facilement à partir de celle du laplacien continu sur $[0, 1]^d$ avec condition au bord de DIRICHLET, et entraîne que

$$2 \sin^2 \left(\frac{\pi}{2(L+1)} \right) \leq -\Delta_\Lambda \leq 2 \cos^2 \left(\frac{\pi}{2(L+1)} \right)$$

au sens des matrices symétriques. Ainsi, le critère de courbure pour la mesure ν_Λ fournit, sous l'hypothèse $V'' \geq \alpha > 0$, les inégalités (0.1), (0.2) et (0.3) pour ν_Λ avec des constantes en $\alpha^{-1} L^2$. Ces inégalités font intervenir le gradient euclidien $\mathbf{E}_{\nu_\Lambda}(|\nabla f|^2)$, et il serait intéressant d'établir un lien clair avec les mesures conditionnées abordées plus haut. L'étude fine des mesures ν_L sous l'hypothèse $0 < \alpha \leq V'' \leq \beta$ est faite par exemple dans [DGI00] et [GOS01].

Perspectives

Très récemment, Pietro CAPUTO a montré que l'inégalité de POINCARÉ restait valable pour une perturbation bornée d'un potentiel strictement convexe [Cap02]. La méthode repose sur une exploitation des symétries et sur le caractère « spectral » de l'inégalité de POINCARÉ. Peut-on étendre ce résultat à l'inégalité de Sobolev logarithmique ? Les inégalités de concentration pour la mesure grand canonique permettent-elles de simplifier les preuves ? Est-il possible d'ajouter des interactions à la mesure produit μ sous-jacente, comme cela a été fait pour des modèles à spins discret par MARTINELLI et ses collaborateurs par exemple ? Dans cette direction, est-il possible de se ramener de la même manière à l'inégalité élémentaire sur la marche aléatoire pour les modèles de même type à spins discrets pour faire apparaître le facteur L^2 ?

Bibliographie

- [ABC⁺00] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO et G. SCHEFFER – *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses, vol. 10, Société Mathématique de France, Paris, 2000. 1, 2, 6, 12, 31, 36, 39, 40, 42, 43, 45, 64, 66, 72
- [BCEF01] F. BARTHE, D. CORDERO-ERAUSQUIN et M. FRADELIZI – « Shift inequalities of Gaussian type and norms of barycentres », *Studia Math.* **146** (2001), no. 3, p. 245–259. 7, 21, 68
- [Bob97] S. G. BOBKOV – « An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space », *Ann. Probab.* **25** (1997), no. 1, p. 206–214. 3, 20
- [Bob99] —, « The size of singular component and shift inequalities », *Ann. Probab.* **27** (1999), no. 1, p. 416–431. 7, 20, 21, 68
- [Cap02] P. CAPUTO – « Uniform Poincaré inequalities for unbounded conservative spin systems: The non-interacting case », Preprint archived as mp_arc 02-46, 2002. 17
- [Cha02a] D. CHAFAÏ – « Gaussian maximum of entropy and reversed logarithmic sobolev inequality », To appear in “Séminaire de Probabilité XXXVI, 2002”, october 2002. 1

- [Cha02b] — , « Glauber versus Kawasaki for spectral gap and logarithmic Sobolev inequalities of some unbounded conservative spin systems », preprint archived as mp_arc 02-30, january 2002. 1
- [CL99] D. CHAFAÏ et M. LEDOUX – « Méthodes fonctionnelles pour des grandes déviations quasi-gaussiennes. (On Gaussian-like large deviations by functional methods). », *C. R. Acad. Sci., Paris, Ser. I, Math.* **329** (1999), no. 6, p. 523–526 (Français). 1
- [DGI00] J.-D. DEUSCHEL, G. GIACOMIN et D. IOFFE – « Large deviations and concentration properties for ∇_φ interface models. », *Probab. Theor. Relat. Fields* **117** (2000), p. 49–111. 17
- [GOS01] G. GIACOMIN, S. OLLA et H. SPOHN – « Equilibrium fluctuations for ∇_φ interface model », Preprint, 2001. 17
- [Led99] M. LEDOUX – « Concentration of measure and logarithmic Sobolev inequalities », Séminaire de Probabilités, XXXIII, Lecture Notes in Math., Springer, Berlin, 1999, p. 120–216. 4, 20, 21
- [Led00] — , « The geometry of Markov diffusion generators », *Ann. Fac. Sci. Toulouse Math. (6)* **9** (2000), no. 2, p. 305–366. 6, 7, 66, 68
- [LPY00] C. LANDIM, G. PANIZO et H. T. YAU – « Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems », Preprint, to appear in Ann. Inst. H. Poincaré, 2000. 2, 15, 16, 71, 77, 78, 80, 84, 85, 86, 89, 90
- [Var67] S. R. S. VARADHAN – « Diffusion processes in a small time interval », *Comm. Pure Appl. Math.* **20** (1967), p. 659–685. 9, 24

Chapitre 1

Méthodes fonctionnelles pour PGD quasi-gaussiens

C. R. Acad. Sci. Paris 329, n°6, 523-526 (1999). Série I.

Djalil CHAFAÏ, <mailto:chafai@cict.fr>, <http://www-sv.cict.fr/lsp/Chafai/>
Michel LEDOUX, <mailto:ledoux@cict.fr>, <http://www-sv.cict.fr/lsp/Ledoux/>
Laboratoire de Statistique et Probabilités, UMR CNRS C5583,
Université Paul Sabatier,
Bât. 1R1, 118 route de Narbonne, 31062 Toulouse, cedex 4, France.

Résumé. Certaines inégalités fonctionnelles gaussiennes possèdent des généralisations simples dans des cas quasi-gaussiens. Elles permettent alors d'établir des principes et bornes de grandes déviations (PGD) quasi-gaussiens via des inégalités de concentration et de translation gaussiennes pour certaines familles de mesures de BOLTZMANN et lois de certains semi groupes diffusifs en temps petit. Au-delà des résultats, nous souhaitons mettre l'accent ici sur la méthode et la symétrie des arguments utilisés pour les bornes inférieures et supérieures.

On Gaussian-like large deviations by functional methods

Abstract. Some Gaussian functional inequalities have simple generalisations to some Gaussian-like cases. They allow us to establish Gaussian-like Large Deviations Principles and bounds via Gaussian concentration and shift inequalities for certain families of BOLTZMANN measures and laws of diffusion semi groups in short time. Beyond the results themselves, we would like to emphasize here the method and the symmetry of the arguments used for upper and lower bounds by means of the functional inequalities.

1.1 Introduction

Soit γ_n la mesure gaussienne standard sur $\mathbb{R}^n$ et soient la densité gaussienne unidimensionnelle φ et sa fonction de répartition Φ données respectivement par $\varphi(x) \stackrel{\text{déf.}}{=} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ et $\Phi(r) \stackrel{\text{déf.}}{=} \int_{-\infty}^r \varphi(x) dx$. La fonction isopérimétrique gaussienne est définie par $\mathbf{I} = \varphi \circ \Phi^{-1}$. S. BOBKOV [4] a établi une inégalité fonctionnelle qui stipule que pour toute fonction $f \in \mathcal{C}_c^2(\mathbb{R}^n, [0, 1])$ on a

$$(1.1) \quad \mathbf{I}(\mathbf{E}_{\gamma_n}(f)) \leq \mathbf{E}_{\gamma_n} \left(\sqrt{(\mathbf{I} \circ f)^2 + |\nabla f|^2} \right).$$

Une preuve par semi-groupes de cette inégalité est présentée dans [1] étendue, par comparaison, aux variétés riemanniennes. En l'appliquant aux indicatrices d'ensembles, l'inégalité (1.1) fournit une propriété isopérimétrique et de concentration qui affirme que pour tout borélien C et tout réel $r > 0$ on a

$$(1.2) \quad \gamma_n(C + B_2(0, r)) \geq \Phi(\Phi^{-1}(\gamma_n(C)) + r)$$

où $B_2(0, r)$ est la boule euclidienne fermée de centre l'origine et de rayon $r > 0$. Il est de plus montré dans [1] que l'inégalité fonctionnelle (1.1) possède une contrepartie gauche qui stipule que pour toute fonction $f \in \mathcal{C}_c^2(\mathbb{R}^n, [0, 1])$ on a

$$(1.3) \quad \sqrt{|\mathbf{E}_{\gamma_n}(\mathbf{I} \circ f)|^2 + |\mathbf{E}_{\gamma_n}(\nabla f)|^2} \leq \mathbf{I}(\mathbf{E}_{\gamma_n}(f)).$$

Ceci entraîne que pour toute fonction $f \in \mathcal{C}_c^2(\mathbb{R}^n, [0, 1])$ on a

$$(1.4) \quad |\mathbf{E}_{\gamma_n}(\nabla f)| \leq \mathbf{I}(\mathbf{E}_{\gamma_n}(f)).$$

Or d'après S. BOBKOV [5], cette inégalité est équivalente à une inégalité de translation gaussienne qui indique que pour tout borélien C et tout vecteur h on a

$$(1.5) \quad \Phi(\Phi^{-1}(\gamma_n(C)) + \|h\|_2) \geq \gamma_n(C + h).$$

Cette dernière est équivalente à

$$\Phi(\Phi^{-1}(\gamma_n(C)) - \|h\|_2) \leq \gamma_n(C + h).$$

par croissance de Φ et de son inverse. Les inégalités (1.2) et (1.5) permettent alors d'établir le très classique PGD gaussien de vitesse ε^2 et de fonction de taux $\frac{1}{2}\|\cdot\|_2^2$ pour la famille de gaussiennes $(\gamma_n(\cdot), \varepsilon > 0)$ où $\gamma_n(\cdot) = \gamma_n(\varepsilon^{-1}\cdot)$. L'inégalité de translation gaussienne (1.5) remplace alors avantageusement la formule de translation de CAMERON-MARTIN dans la preuve de la borne inférieure. Comme nous allons le voir par la suite, elle est plus souple que cette dernière et possède une généralisation simple dans certains cas quasi-gaussiens. Signalons enfin que ce PGD peut être généralisé en un PGD sans topologie [3, 7].

1.2 PGD pour certaines familles de mesures de Boltzmann

Notre but ici est de démontrer un PGD avec les méthodes présentées précédemment pour des mesures proches des gaussiennes. On se restreindra à $\mathbb{R}^n$ pour faciliter l'exposé mais les arguments présentés s'étendent à des cadres plus généraux (voir plus loin). Soit μ une mesure de BOLTZMANN sur $\mathbb{R}^n$ donnée par $d\mu(x) = \mathbf{Z}^{-1} \exp(-\mathbf{H}(x)) dx$ où l'hamiltonien $\mathbf{H}$ est dans $\mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$. Il est démontré dans [1] que si $\nabla^2 \mathbf{H} \geq \alpha > 0$ uniformément sur $\mathbb{R}^n$ au sens des formes quadratiques alors pour toute fonction $f \in \mathcal{C}_c^\infty(\mathbb{R}^n, [0, 1])$ on a

$$(1.6) \quad \mathbf{I}(\mathbf{E}_\mu(f)) \leq \mathbf{E}_\mu \left(\sqrt{(\mathbf{I} \circ f)^2 + \alpha^{-1} |\nabla f|^2} \right).$$

D'autre part, d'après [2], si $\mathbf{H}$ est convexe avec $\|\nabla^2 \mathbf{H}\| \leq \beta$ en tant qu'opérateur linéaire alors, pour toute fonction $f \in \mathcal{C}_c^\infty(\mathbb{R}^n, [0, 1])$ on a

$$(1.7) \quad \sqrt{|\mathbf{E}_\mu(\mathbf{I} \circ f)|^2 + \beta^{-1} |\mathbf{E}_\mu(\nabla f)|^2} \leq \mathbf{I}(\mathbf{E}_\mu(f)).$$

La gaussienne correspond à $\mathbf{H}(\cdot) = \frac{1}{2} \|\cdot\|_2^2$ avec $\alpha = \beta = 1$. Les démonstrations les plus courtes de (1.6) et (1.7) font intervenir le semi groupe de diffusion ergodique associé à μ . Comme pour (1.2) et (1.5), (1.6) entraîne que pour tout borélien C et tout réel $r > 0$ on a

$$(1.8) \quad \mu(C + B_2(0, r)) \geq \Phi(\Phi^{-1}(\mu(C)) + \sqrt{\alpha} r)$$

(cf. [7]) alors que (1.7) entraîne que pour tout borélien C et tout vecteur h on a

$$(1.9) \quad \Phi(\Phi^{-1}(\mu(C)) + \sqrt{\beta} \|h\|_2) \geq \mu(C + h)$$

(cf. [5]). Les inégalités (1.8) et (1.9) vont jouer dans l'établissement du PGD qui va suivre le même rôle que celui joué par (1.2) et (1.5) dans l'établissement du PGD gaussien évoqué précédemment.

Théorème 1.2.1. *Soit une famille de mesures de BOLTZMANN $(\mu_\varepsilon, \varepsilon > 0)$ sur $\mathbb{R}^n$ d'hamiltoniens $(\mathbf{H}_\varepsilon, \varepsilon > 0)$ dans $\mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ convergeant étroitement vers δ_0 quand ε tend vers 0. Supposons que pour tout $\varepsilon > 0$, il existe des réels α_ε et β_ε tels que $0 < \alpha_\varepsilon \mathbf{Id}_n \leq \nabla^2 \mathbf{H}_\varepsilon \leq \beta_\varepsilon \mathbf{Id}_n$ uniformément sur $\mathbb{R}^n$ au sens des formes quadratiques avec $\lim_{\varepsilon \rightarrow 0^+} \alpha_\varepsilon = +\infty$. Alors pour tout borélien C on a*

$$(1.10) \quad - \inf_{\text{int}(C)} \mathbf{J} \leq \liminf_{\varepsilon \rightarrow 0} \beta_\varepsilon^{-1} \log \mu_\varepsilon(C) \leq \limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \log \mu_\varepsilon(C) \leq - \inf_{\text{adh}(C)} \mathbf{J}$$

où $\mathbf{J} \stackrel{\text{déf.}}{=} \frac{1}{2} \|\cdot\|_2^2$.

Le cas typique est celui donné par la contraction d'une mesure de BOLTZMANN d'hamiltonien $\mathbf{H}$ vérifiant $0 < \alpha \mathbf{Id}_n \leq \nabla^2 \mathbf{H} \leq \beta \mathbf{Id}_n$ uniformément sur $\mathbb{R}^n$ au sens des formes quadratiques. On a alors $\mathbf{H}_\varepsilon(\cdot) = \mathbf{H}(\varepsilon^{-1}\cdot)$ avec $\alpha_\varepsilon = \varepsilon^{-2}\alpha$ et $\beta_\varepsilon = \varepsilon^{-2}\beta$. On retrouve ainsi le PGD gaussien classique en prenant $\mathbf{H}(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ et $\alpha = \beta = 1$.

Preuve. Les arguments de démonstration dépassent le cadre simple de $\mathbb{R}^n$ choisi par commodité. En vue des généralisations éventuelles, nous distinguons les boules B pour une norme quelconque sur $\mathbb{R}^n$ des boules euclidiennes B_2 .

Commençons par la borne inférieure. Par inclusion, il suffit de montrer que pour toute boule $B(x, \eta)$ on a pour $\varepsilon \leq \varepsilon_\eta$

$$\Phi\left(-\sqrt{\beta_\varepsilon}\|x\|_2\right) \leq \mu_\varepsilon(B(x, \eta)).$$

Or l'inégalité (1.9) entraîne

$$\mu_\varepsilon(B(0, \eta)) \leq \Phi\left(\Phi^{-1}(\mu_\varepsilon(B(x, \eta))) + \sqrt{\beta_\varepsilon}\|x\|_2\right).$$

Mais $\mu_\varepsilon(B(0, \eta))$ converge vers 1 quand $\varepsilon \rightarrow 0^+$ donc pour $\varepsilon \leq \varepsilon_\eta$, la quantité $\Phi^{-1}(\mu_\varepsilon(B(0, \eta)))$ est positive et on a alors le résultat par croissance de Φ et de son inverse.

Passons maintenant à la borne supérieure. Faisons la preuve pour C fermé. Il n'y a rien à démontrer quand $\inf_C \|\cdot\|_2 = 0$. Soit donc $0 < \eta < \inf_C \|\cdot\|_2$. Il suffit de montrer que pour $\varepsilon \leq \varepsilon_C$ on a

$$\mu_\varepsilon(C) \leq \Phi(-\sqrt{\alpha_\varepsilon}\eta).$$

Par définition de η , l'ensemble $C \cap B_2(0, \eta)$ est vide. C étant fermé et $B_2(0, \eta)$ étant compact, il existe un $\delta > 0$ tel que $A \cap (B_2(0, \eta) + B(0, \delta))$ soit vide. On a donc

$$\mu_\varepsilon(C) \leq 1 - \mu_\varepsilon(B_2(0, \eta) + B(0, \delta)).$$

Or $\mu_\varepsilon(B(0, \delta))$ converge vers 1 quand $\varepsilon \rightarrow 0^+$ donc en vertu de (1.8) on a pour $\varepsilon \leq \varepsilon_\delta$ l'inégalité

$$\mu_\varepsilon(C) \leq 1 - \Phi(\sqrt{\alpha_\varepsilon}\eta) = \Phi(-\sqrt{\alpha_\varepsilon}\eta).$$

□

Une lecture attentive de la démonstration permet de voir que pour tout borélien C relativement compact ne possédant pas 0 comme point adhérent, il existe un réel $\varepsilon_C > 0$ tel que pour tout $\varepsilon < \varepsilon_C$ on a

$$\Phi\left(-\sqrt{\beta_\varepsilon} \inf_{\text{int}(C)} \|\cdot\|_2\right) \leq \mu_\varepsilon(C) \leq \Phi\left(-\sqrt{\alpha_\varepsilon} \inf_{\text{adh}(C)} \|\cdot\|_2\right).$$

La démonstration de la borne supérieure du PGD utilise de façon cruciale l'inf-compacité de $\|\cdot\|_2$.

D'autre part, le théorème reste vrai si l'on suppose seulement que l'on a $\nabla^2 \mathbf{H}_\varepsilon = \nabla^2 \mathbf{G}_\varepsilon$ sur le complémentaire d'un compact K_ε convergent vers $\{0\}$ où $\nabla^2 \mathbf{G}_\varepsilon$, et non plus $\nabla^2 \mathbf{H}_\varepsilon$, vérifie l'hypothèse quasi-gaussienne.

Comme nous l'avons déjà mentionné pour le cas gaussien, on peut obtenir sans difficultés et de la même manière un PGD sans topologie qui généralise un peu le résultat (1.10). Pour ce faire, on définit en suivant [3] les fonctions $\mathbf{s}$ et $\mathbf{r}$ sur les boréliens par :

$$\mathbf{r}(A) = \sup \left\{ r \geq 0; \exists V / \liminf_{\varepsilon \rightarrow 0^+} \mu_\varepsilon(V) > 0, A \cap (V + B_2(0, r)) = \emptyset \right\}$$

et

$$\mathbf{s}(A) = \inf \left\{ s \geq 0; \exists V / \liminf_{\varepsilon \rightarrow 0^+} \mu_\varepsilon(V) > 0, B_2(0, r) \cap (A^c + V)^c = \emptyset \right\}.$$

Ce sont des fonctions décroissantes pour l'inclusion. On peut montrer que $\mathbf{r}^2 \geq 2\mathbf{J}$ avec égalité sur les fermés et que $\mathbf{s}^2 \leq 2\mathbf{J}$ avec égalité sur les ouverts. La preuve de (1.10) peut alors être adaptée pour montrer sous les mêmes hypothèses que pour tout borélien C de $\mathbb{R}^n$ on a

$$(1.11) \quad -\frac{1}{2}\mathbf{s}(C)^2 \leq \liminf_{\varepsilon \rightarrow 0} \beta_\varepsilon^{-1} \log \mu_\varepsilon(C) \leq \limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \log \mu_\varepsilon(C) \leq -\frac{1}{2}\mathbf{r}(C)^2.$$

1.3 Semi-groupes diffusifs et généralisations possibles

Nous allons voir que les méthodes fonctionnelles utilisées pour les familles de mesures de BOLTZMANN sont encore valables pour les lois de certains semi-groupes diffusifs. En effet, les inégalités (1.6) et (1.7) ont encore lieu sous une condition de courbure du générateur. Plus précisément, soit $(\mathbf{P}_t, t > 0)$ un semi-groupe diffusif sur $\mathbb{R}^n$ de générateur $\mathbf{L} = \Delta - \nabla \mathbf{H} \cdot \nabla$ où $\mathbf{H}$ est $\mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$. S'il existe un réel α tel que $\alpha \mathbf{Id}_n \leq \nabla^2 \mathbf{H}$ uniformément sur $\mathbb{R}^n$ au sens des formes quadratiques alors pour toute fonction $f \in \mathcal{C}_c^\infty(\mathbb{R}^n, [0, 1])$, tout $x \in \mathbb{R}^n$ et tout $t > 0$ on a

$$(1.12) \quad \mathbf{I}(\mathbf{P}_{t,x}(f)) \leq \mathbf{P}_{t,x} \left(\sqrt{(\mathbf{I} \circ f)^2 + \alpha(t) |\nabla f|^2} \right)$$

où $\alpha(t) \stackrel{\text{déf.}}{=} \alpha^{-1}(1 - e^{-2\alpha t})$ (cf. [1]). Quand $\alpha = 0$, on pose simplement $\alpha(t) = 2t$. D'autre part, si $\|\nabla^2 \mathbf{H}\| \leq \beta$ uniformément sur $\mathbb{R}^n$ en tant qu'opérateur linéaire alors pour toute fonction $f \in \mathcal{C}_c^\infty(\mathbb{R}^n, [0, 1])$, tout $x \in \mathbb{R}^n$ et tout $t > 0$ on a de la même façon

$$(1.13) \quad \sqrt{|\mathbf{P}_{t,x}(\mathbf{I} \circ f)|^2 + \beta(t) |\mathbf{P}_{t,x}(\nabla f)|^2} \leq \mathbf{I}(\mathbf{P}_{t,x}(f))$$

où $\beta(t) \stackrel{\text{déf.}}{=} \beta^{-1}(1 - e^{-2\beta t})$. Remarquons que lorsque $\alpha > 0$ et $\beta > 0$, les inégalités (1.12) et (1.13) entraînent respectivement (1.6) et (1.7) par ergodicité en faisant tendre le temps t vers l'infini.

Les inégalités (1.12) et (1.13) entraînent des inégalités gaussiennes du type (1.8) et (1.9) pour les lois du semi-groupe qui entraînent à leur tour comme nous l'avons vu un PGD gaussien pour les lois du semi-groupe en temps petit.

Théorème 1.3.1. *Soit $(\mathbf{P}_t, t > 0)$ un semi-groupe diffusif sur $\mathbb{R}^n$ de générateur $\mathbf{L} = \Delta - \nabla \mathbf{H} \cdot \nabla$ tel que $\nabla^2 \mathbf{H}$ soit uniformément borné en tant qu'opérateur linéaire sur $\mathbb{R}^n$. Alors, pour tout borélien C et tout x dans $\mathbb{R}^n$ on a*

$$(1.14) \quad - \inf_{\text{int}(C)} \mathbf{J}_x \leq \liminf_{t \rightarrow 0} 2t \log \mathbf{P}_{t,x}(C) \leq \limsup_{t \rightarrow 0} 2t \log \mathbf{P}_{t,x}(C) \leq - \inf_{\text{adh}(C)} \mathbf{J}_x$$

où $\mathbf{J}_x = \frac{1}{2} \|x - \cdot\|_2^2$.

Dans ce cadre restreint, et sous les hypothèses sur $\mathbf{H}$, le théorème précédent est évidemment bien connu [8]. Notre méthode fonctionnelle permet néanmoins d'en offrir une version sans topologie et de montrer des bornes précises à temps fixe : pour tout borélien C relativement compact ne possédant pas x comme point adhérent, il existe un réel $\varepsilon_C > 0$ tel que pour tout $\varepsilon < \varepsilon_C$ on a

$$\Phi\left(-\sqrt{\beta_t} \inf_{\text{int}(C)} \|x - \cdot\|_2\right) \leq \mathbf{P}_{t,x}(C) \leq \Phi\left(-\sqrt{\alpha_t} \inf_{\text{adh}(C)} \|x - \cdot\|_2\right).$$

En outre, les arguments développés ci-dessus s'étendent, comme annoncé, bien au-delà du cadre euclidien (toute autre métrique euclidienne convient également). Les inégalités (1.12) et (1.13) se prolongent à la loi de tout le processus [1], ce qui permet des versions fonctionnelles du premier théorème. En suivant certains travaux récents [6], on pourrait également considérer un semi-groupe diffusif associé à l'équation différentielle stochastique $dX_t = dB_t - \nabla \mathbf{H}(X_t) dt$ où $(B_t, t > 0)$ est un mouvement brownien à valeurs dans un espace de BANACH. Sous des conditions d'existence de la solution, les méthodes développées précédemment donnent lieu à un PGD pour la loi de X_t (sous des conditions de convexité vectorielle sur $\mathbf{H}$).

Signalons pour terminer que les inégalités (1.6) et (1.7) peuvent être remplacées par leur version de SOBOLEV logarithmique qui, sans la précision des précédentes, permettent des estimations de grandes déviations du même ordre.

Bibliographie

- [1] D. BAKRY et M. LEDOUX – « Lévy-Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator », *Invent. Math.* **123** (1996), no. 2, p. 259–281. 20, 21, 23, 24
- [2] F. BARTHE, D. CORDERO-ERAUSQUIN et M. FRADELIZI – « Shift inequalities of Gaussian type and norms of barycentres », *Studia Math.* **146** (2001), no. 3, p. 245–259. 7, 21, 68

- [3] G. BEN AROUS et M. LEDOUX – « Schilder's large deviation principle without topology », *Asymptotic problems in probability theory: Wiener functionals and asymptotics* (Sanda/Kyoto, 1990), Longman Sci. Tech., Harlow, 1993, p. 107–121. 20, 23
- [4] S. G. BOBKOV – « An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space », *Ann. Probab.* **25** (1997), no. 1, p. 206–214. 3, 20
- [5] — , « The size of singular component and shift inequalities », *Ann. Probab.* **27** (1999), no. 1, p. 416–431. 7, 20, 21, 68
- [6] S. FANG et T. S. ZHANG – « On the small time behavior of Ornstein-Uhlenbeck processes with unbounded linear drifts », *Probab. Theory Related Fields* **114** (1999), no. 4, p. 487–504. 24
- [7] M. LEDOUX – « Concentration of measure and logarithmic Sobolev inequalities », *Séminaire de Probabilités, XXXIII, Lecture Notes in Math.*, Springer, Berlin, 1999, p. 120–216. 4, 20, 21
- [8] S. R. S. VARADHAN – « Diffusion processes in a small time interval », *Comm. Pure Appl. Math.* **20** (1967), p. 659–685. 9, 24

Chapitre 2

Quelques inégalités entropiques en théorie de l'information

par Cécile ANÉ et Djalil CHAFAÏ

Chapitre n°10 du livre « Sur les inégalité de Sobolev logarithmiques », Panoramas et Synthèses, vol. n°10, Société Mathématique de France, Paris (2000).

« 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.” »

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].

2.1 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 ». Il en est de même pour l'information de FISHER de μ , donnée par

$$\mathbf{J}(\mu) = \mathbf{E}_{\mu}(|\nabla \log f|^2) = 4 \int \Gamma(\sqrt{f}) dx.$$

La « théorie de l'information », également appelée « théorie de la communication », 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 2.1.

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.

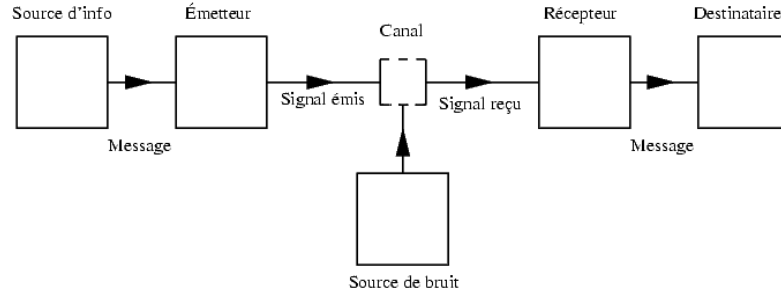


FIG. 2.1 – Diagramme schématisé d'un système de communication générique (tiré de [Sha48]).

2.2 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 2.1.

2.2.1 Entropie d'une variable aléatoire discrète finie

Considérons une variable aléatoire discrète X prenant les valeurs distinctes

$$x_1, \dots, x_n$$

avec probabilités $p_1, \dots, p_n$. 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 $(p_1, \dots, p_n)$ un nombre réel positif, noté $\mathbf{H}^{(n)}(p_1, \dots, p_n)$, croissant avec le « désordre » et la « variabilité ». 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 » sur son support. Ainsi, on s'attend à ce que la quantité $\mathbf{H}^{(n)}(X) = \mathbf{H}^{(n)}(p_1, \dots, p_n)$ 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 découleront comme nous allons le voir) :

1. $\mathbf{H}^{(n)}$ est continue en les variables p_i ;
2. $\mathbf{H}^{(n)}\left(\frac{1}{n}, \dots, \frac{1}{n}\right) < \mathbf{H}^{(n+1)}\left(\frac{1}{n+1}, \dots, \frac{1}{n+1}\right)$;
3. Pour tout $(b_1, \dots, b_k)$ tel que $b_1 + \dots + b_k = n$,

$$\mathbf{H}^{(n)}\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = \mathbf{H}^{(k)}\left(\frac{b_1}{n}, \dots, \frac{b_k}{n}\right) + \sum_{i=1}^k \frac{b_i}{n} \mathbf{H}^{(b_i)}\left(\frac{1}{b_i}, \dots, \frac{1}{b_i}\right).$$

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

$$(2.1) \quad \mathbf{H}_b(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log_b p_i = \sum_{i=1}^n p_i \log_b \frac{1}{p_i},$$

où $\log_b$ désigne le logarithme de base $b > 0$ et $0 \log_b 0 \stackrel{\text{déf.}}{=} 0$. 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

$$\left\{ (p_1, \dots, p_n) ; \forall i, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}.$$

Remarquons enfin que $-p_i \log_b p_i$ peut être vue comme l'« incertitude » 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 2.2.1 (Entropies $\mathbf{H}$ et $\mathbf{Ent}$). Nous ne reprenons pas ici la notation $\mathbf{Ent}$ des chapitres précédents pour l'entropie, malgré l'égalité formelle¹

$$\mathbf{H}_e(p_1, \dots, p_n) = -\mathbf{Ent}_\mu(p),$$

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 $x \log x$,

¹Cette opposition de signe a conduit le physicien BRILLOUIN à proposer le terme de « néguentropie » pour $\mathbf{H}$, par opposition à l'entropie $\mathbf{Ent}_{dx}$, qui apparaît en thermodynamique [Bri64].

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 $\mathbf{Ent}$ introduite au chapitre 1 du livre [ABC⁺00] seront donc perdues. C'est pour éviter toute confusion que nous avons préféré changer de notation.

2.2.2 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 « $\sqcup$ ». Par exemple, les lettres E et Z sont codées respectivement par « $\cdot \sqcup$ » et « $- - \cdot \sqcup$ ». 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 « $\cdot$ », « $-$ » et « $\sqcup$ ».

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, on peut prendre pour A l'alphabet latin additionné de l'espace « $\sqcup$ ». Un message est un élément de A^* . Pour le coder, on utilise un ensemble $E = \{e_1, \dots, e_b\}$. Par exemple, avec le code MORSE, on a $b = 3$ et $E = \{\cdot, -, \sqcup\}$. Un (b, n) -code consiste en un sous-ensemble $C = \{c_1, \dots, c_n\}$ 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 « $\cdot$ » ou « $-$ », suivis de l'espace « $\sqcup$ ».

On appelle « schéma de codage » 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 » par f en $f(x_{i_1}) \cdots f(x_{i_k})$, qui appartient à E^* . Dans la suite, nous supposons 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 » est codé « 01000101 »².

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

²En 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.

tout $i = 1, \dots, j$. 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 », 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 fautive en général.

KRAFT a montré en 1949 que si C est instantané, alors il satisfait à la condition suivante

$$\sum_{i=1}^n b^{-l_i} \leq 1,$$

où l_i est la longueur de c_i dans E^* . De plus, si des nombres entiers $(l_1, \dots, l_n)$ 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 :

$$|M|_C \stackrel{\text{déf.}}{=} \sum_{i=1}^n p_i l_i,$$

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 préférons la notation $|P|_C = |M|_C$.

Un théorème de « codage non bruité » (« Noiseless coding theorem » en anglais), dû à SHANNON [Sha48], affirme que si $\mathcal{C}(n, b)$ désigne l'ensemble des (n, b) -codes instantanés, on a

$$\mathbf{H}_b(p_1, \dots, p_n) \leq \inf_{C \in \mathcal{C}(n, b)} |P|_C < \mathbf{H}_b(p_1, \dots, p_n) + 1.$$

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 »³.

Le code ASCII est, à l'opposé des codes de HUFFMAN, un code à longueur fixe. Le codage de la lettre « E », 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 $\log_2 m = 8$ (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 $H_b(P) / \log_b m$ 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 par la prise en compte des fréquences d'apparition de chaînes de plusieurs symboles consécutifs⁴. Cependant, des algorithmes relativement simples dits « à dictionnaire » comme LZ(W)⁵ 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.

2.2.3 Entropie d'une variable aléatoire continue

La formule (2.1) 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

³D'autres codes aussi performants existent, comme par exemple les codes arithmétiques. Voir [CT91] ou [Sto88].

⁴On peut par exemple songer au message « $x_1x_2x_3 \cdots x_1x_2x_3$ ».

⁵Pour 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.

son entropie de base b par

$$(2.2) \quad \mathbf{H}_b(X) \stackrel{\text{déf.}}{=} - \int_{\mathbb{R}} f \log_b f dx = -\mathbf{E}(\log_b f(X)).$$

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 $\{i/b^p, 0 \leq i < b^p\}$ 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.

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

$$(2.3) \quad \mathbf{H}(X) \stackrel{\text{déf.}}{=} - \int_{\mathbb{R}^n} f \log f dx = -\mathbf{E}(\log f(X)) = -\mathbf{E}_{\mathcal{L}(X)}(\log f),$$

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.

2.2.4 Quelques propriétés immédiates de l'entropie

L'entropie $\mathbf{H}$ définie en (2.3) 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 ,

$$(2.4) \quad \mathbf{H}(\alpha X) = \mathbf{H}(X) + n \log \alpha.$$

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 2.2.2 (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*

$$(2.5) \quad \mathbf{H}(X) \leq \frac{1}{2} \log ((2\pi e)^n \det K),$$

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 :

$$\mathbf{Ent}(f | g) \stackrel{\text{déf.}}{=} \int \frac{f(x)}{g(x)} \log \frac{f(x)}{g(x)} g(x) dx = \int f(x) \log \frac{f(x)}{g(x)} dx.$$

La stricte concavité⁶ du logarithme permet de voir que $\mathbf{Ent}(f | g) \geq 0$, 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

$$0 \leq \mathbf{Ent}(f | \gamma_K) = -\mathbf{H}(X) + \mathbf{H}(\mathcal{N}(0, K)).$$

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 une constante près. Ainsi, on a

$$\int f \log \gamma_K dx = \int \gamma_K \log \gamma_K dx.$$

□

⁶Ou alternativement la stricte convexité de $x \log x$.

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 2.2.3 (Règle de la chaîne ou sous-additivité de l'entropie). *Soient n vecteurs aléatoires $X_1, \dots, X_n$ à densité par rapport à la mesure de LEBESGUE. Alors on a*

$$(2.6) \quad \mathbf{H}((X_1, \dots, X_n)) \leq \sum_{i=1}^n \mathbf{H}(X_i),$$

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 (proposition 1.4.1 page 9 du livre [ABC⁺00]) 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 :

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

où $\mu_{\setminus i}$ désigne la mesure produit des μ_j avec $j \neq i$.

Preuve du théorème 2.2.3. 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

$$\mathbf{H}(X) - \mathbf{H}(X|Y) = \mathbf{Ent}(f(x, y) | f_1(x)f_2(y)) \geq 0.$$

Ainsi, on voit que

$$\mathbf{H}(X|Y) \leq \mathbf{H}(X),$$

avec égalité si et seulement si X et Y sont indépendantes.

Il en découle que

$$\mathbf{H}((X_1, \dots, X_n)) = \sum_{i=1}^n \mathbf{H}(X_i | (X_{i-1}, \dots, X_1)) \leq \sum_{i=1}^n \mathbf{H}(X_i).$$

□

Une conséquence amusante de ce théorème est une version faible de l'inégalité de HADAMARD sur le déterminant d'une matrice⁷ qui s'obtient en appliquant la règle de la chaîne à des variables aléatoires gaussiennes bien choisies.

2.2.5 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

$$\mathbf{I}(X, Y) \stackrel{\text{déf.}}{=} \mathbf{H}((X, Y)) - \mathbf{H}(X|Y) - \mathbf{H}(Y|X) = \mathbf{I}(Y, X).$$

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].

Revenons à la figure 2.1 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 » X , et la réception par une variable aléatoire « réception » Y , dépendant de X . Les émissions et réceptions successives peuvent être alors vues comme des réalisations indépendantes du couple (X, Y) . Un canal idéal sans bruit correspondrait au cas où $Y = X$. Dans la pratique, on peut considérer que l'« émission » X subit, dans le canal, de nombreuses petites perturbations indépendantes dont l'effet cumulé est bien modélisé

⁷Si $A = (a_{ij})_{1 \leq i, j \leq n}$ est une matrice symétrique positive, alors $\det A \leq \prod_{i=1}^n a_{ii}$. On pourra consulter [CT91] ou [DCT91] pour d'autres applications matricielles de ce type.

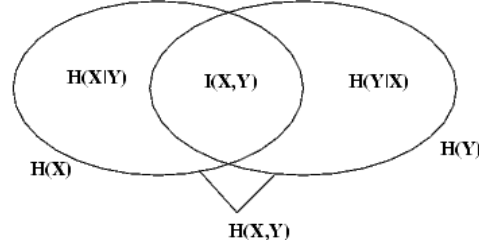


FIG. 2.2 – Imbrication des quantités liées à l'entropie et à l'information.

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 idéalisés⁸.

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 » X de variance P . Or le maximum gaussien de l'entropie de SHANNON à variance fixée (2.5) entraîne que :

$$\mathbf{I}(X, Y) \stackrel{\text{déf.}}{=} \mathbf{H}(Y) - \mathbf{H}(Z) \leq \frac{1}{2} \log 2\pi e(P + N) - \frac{1}{2} \log 2\pi eN = \frac{1}{2} \log \left(1 + \frac{P}{N} \right),$$

avec égalité si et seulement si X suit une loi gaussienne de variance P . Nous venons donc d'établir l'égalité suivante :

$$C(P) = \sup_{\text{Var}(X)=P} \mathbf{I}(X, X + Z) = \frac{1}{2} \log \left(1 + \frac{P}{N} \right).$$

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 2.1, 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

⁸Pour un canal faisant transiter des messages « discrets » constitués d'éléments $\{e_1, \dots, e_n\}$, 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].

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

$$\frac{1}{2}W \log \left(1 + \frac{P}{NW} \right),$$

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 (2.13) 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.

2.3 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 (théorème 1.5.2 page 12 du livre [ABC⁺00] par exemple), et une inégalité connue depuis les années 50 en théorie de l'information, et que nous nommerons « version euclidienne » 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 page 13 du livre [ABC⁺00], l'inégalité de SOBOLEV logarithmique pour la loi gaussienne exprime que pour toute fonction f dérivable de $\mathbb{R}^n$ dans $\mathbb{R}$,

$$(2.7) \quad \mathbf{Ent}_{\mathcal{N}(0, tI_n)}(f^2) \leq 2t \mathbf{E}_{\mathcal{N}(0, tI_n)}(|\nabla f|^2).$$

Comme nous l'avons vu à la section 4.6.1 page 70 du livre [ABC⁺00], cette inégalité est équivalente à l'inégalité suivante sur la mesure de LEBESGUE :

$$(2.8) \quad \int g^2 \log g^2 dx \leq \frac{n}{2} \log \left[\frac{2}{e\pi n} \int |\nabla g|^2 dx \right],$$

valable pour toute fonction g à décroissance rapide sur $\mathbb{R}^n$ appartenant à la sphère unité de $\mathbf{L}^2(dx)$. L'inégalité (2.8) est optimale et l'égalité est atteinte pour certaines fonctions « gaussiennes », exponentielles de formes quadratiques [Car91].

Dans la suite de ce chapitre, nous parlerons de l'inégalité de SOBOLEV logarithmique euclidienne (optimale) pour désigner (2.8).

Comme nous l'avons vu à la section 4.6.1 page 70 du livre [ABC⁺00], l'inégalité (2.8), qui est du type (EEL) (page 64 du livre [ABC⁺00]), 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 $n \rightarrow \infty$ de la constante optimale de l'inégalité de SOBOLEV dans $\mathbb{R}^n$. Notons également que dans (2.8), 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 $\alpha^{n/2}f(\alpha x)$.

2.3.1 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

$$(2.9) \quad \mathbf{N}(X) \stackrel{\text{déf.}}{=} \frac{1}{2\pi e} e^{\frac{2}{n}\mathbf{H}(X)}.$$

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 $\mathbf{N}(X) \leq (\det K)^{1/n}$. Ceci découle facilement de (2.5). À variance fixée, la loi gaussienne est donc celle qui « contient le plus d'incertitude », au sens de l'entropie. Ainsi, de la même manière que $(\det K)^{1/n}$ représente le « rayon moyen » de la matrice K , $\mathbf{N}(X)$ représente en quelque sorte le « rayon d'incertitude » 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'énergie⁹

$$(2.10) \quad \mathbf{J}(X) \stackrel{\text{déf.}}{=} 4 \int \left| \nabla \left(\sqrt{f} \right) \right|^2 dx.$$

La *version euclidienne de l'inégalité de SOBOLEV logarithmique* (2.8) pour $g = \sqrt{f}$ s'écrit alors à l'aide des quantités introduites

$$(2.11) \quad \mathbf{N}(X) \mathbf{J}(X) \geq n.$$

⁹On a aussi : $\mathbf{J}(X) = \mathbf{E}_{\mathcal{L}(X)}(|\nabla \log f|^2) = \int |\nabla \log f|^2 f dx = \int |\nabla f|^2 f^{-1} dx = \int \nabla f \cdot \nabla \log f dx$.

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 (2.11) est invariante par dilatation¹⁰.

2.4 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 de SOBOLEV logarithmique (2.11). 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 (2.11), 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].

Théorème 2.4.1 (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*

$$(2.12) \quad \lambda \mathbf{J}(X) + (1 - \lambda) \mathbf{J}(Y) \geq \mathbf{J}(\sqrt{\lambda}X + \sqrt{1 - \lambda}Y).$$

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 (2.11) de cette inégalité de BLACHMAN-STAM¹¹. 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.

2.4.1 Première méthode

Dans son article fondateur [Sha48], SHANNON prouve l'un des théorèmes de « codage bruité » grâce à l'inégalité suivante :

Théorème 2.4.2 (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*

¹⁰Plus 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 $\mathbf{J}(X) = \text{Tr}(\mathbb{J}(X))$ où $\mathbb{J}$ est définie plus loin en (2.21), et par $\mathbb{J}(AX) = A^{-1\top} \mathbb{J}(X) A^{-1}$.

¹¹ également connue sous la formulation équivalente suivante : pour tous vecteurs X et Y indépendants, $\mathbf{J}(X + Y)^{-1} \geq \mathbf{J}(X)^{-1} + \mathbf{J}(Y)^{-1}$.

de LEBESGUE, on a

$$(2.13) \quad \mathbf{N}(X + Y) \geq \mathbf{N}(X) + \mathbf{N}(Y).$$

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 (2.8) de cette inégalité de SHANNON (2.13). 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 2.4.3 (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*

$$(2.14) \quad \partial_t \left[\mathbf{H}(X + \sqrt{t}Z) \right] = \frac{1}{2} \mathbf{J}(X + \sqrt{t}Z).$$

Preuve. Soit $(\mathbf{P}_t)_{t \geq 0}$ 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 du livre [ABC⁺00]). La densité de $X + \sqrt{t}Z$ est alors $\mathbf{P}_t f$, où f est la densité de X , et $\mathbf{H}(X + \sqrt{t}Z) = \mathbf{H}(\mathbf{P}_t f)$. Ainsi la densité de $X + \sqrt{t}Z$ vérifie l'équation de la chaleur :

$$\partial_t \mathbf{P}_t f = \frac{1}{2} \Delta(\mathbf{P}_t f).$$

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 :

$$\begin{aligned} \partial_t \mathbf{H}(\mathbf{P}_t f) &= - \int \frac{1}{2} \Delta(\mathbf{P}_t f) (\log \mathbf{P}_t f + 1) dx \\ &= - \frac{1}{2} \int (\log \mathbf{P}_t f) \Delta(\mathbf{P}_t f) dx \\ &= \frac{1}{2} \mathbf{J}(\mathbf{P}_t f) = \frac{1}{2} \mathbf{J}(X + \sqrt{t}Z), \end{aligned}$$

ce qui achève la preuve de l'identité de DEBRUIJN. □

Remarque 2.4.4. Ce dernier calcul est exactement celui qui a été fait dans la preuve de la décroissance de l'entropie, théorème 2.6.7 page 35 du livre [ABC⁺00], et qui aboutissait à l'équation (2.17) de [ABC⁺00] :

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

Il est donc très tentant de calquer la preuve du théorème 2.6.7 du livre [ABC⁺00] dans notre contexte, ce que nous allons faire. En utilisant l'inégalité de SOBOLEV logarithmique (2.11), on majore $\mathbf{J}(\mathbf{P}_t f)$ par une expression en $\mathbf{H}(\mathbf{P}_t f)$, ce qui donne exactement

$$\partial_t \mathbf{N}(\mathbf{P}_t f) \geq 1.$$

En intégrant, on obtient

$$\mathbf{N}(X + \sqrt{t}Z) \geq \mathbf{N}(X) + t = \mathbf{N}(X) + \mathbf{N}(\sqrt{t}Z),$$

et on reconnaît alors l'inégalité de l'entropie exponentielle (2.13) avec une variable Y de loi normale $\mathcal{N}(0, tI_n)$.

Utilisons maintenant l'inégalité de SHANNON (2.13). Elle implique facilement que

$$\partial_{t=0} [\mathbf{N}(X + \sqrt{t}Z)] \geq \mathbf{N}(Z) = 1.$$

Mais par l'identité de DEBRUIJN,

$$\begin{aligned} \partial_{t=0} [\mathbf{N}(X + \sqrt{t}Z)] &= \frac{2}{n} \partial_{t=0} [\mathbf{H}(X + \sqrt{t}Z)] \mathbf{N}(X) \\ &= \frac{1}{n} \mathbf{N}(X) \mathbf{J}(X). \end{aligned}$$

On obtient alors l'inégalité de SOBOLEV logarithmique, version euclidienne (2.11).

Pour obtenir l'inégalité de SHANNON à partir de celle de BLACHMAN-STAM, nous passerons par les étapes suivantes :

Théorème 2.4.5 (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*

$$(2.15) \quad \mathbf{H}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda \mathbf{H}(X) + (1-\lambda) \mathbf{H}(Y).$$

Théorème 2.4.6. *Pour tous vecteurs aléatoires indépendants X et Y de $\mathbb{R}^n$ à densité par rapport à la mesure de LEBESGUE, on a*

$$(2.16) \quad \mathbf{H}(X + Y) \geq \mathbf{H}(\tilde{X} + \tilde{Y}),$$

où $\tilde{X}$ et $\tilde{Y}$ sont deux vecteurs aléatoires gaussiens indépendants, de covariances proportionnelles, tels que

$$\mathbf{H}(\tilde{X}) = \mathbf{H}(X) \quad \text{et} \quad \mathbf{H}(\tilde{Y}) = \mathbf{H}(Y).$$

On voit facilement que l'inégalité (2.16) 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 (2.15) et (2.16) sont équivalentes car pour deux lois gaussiennes de covariances proportionnelles, (2.15) 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 (2.15) de l'inégalité de BLACHMAN-STAM (2.12) [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 $\varphi(t) = \mathbf{H}(V_t) - \lambda\mathbf{H}(X_t) - (1-\lambda)\mathbf{H}(Y_t) \geq 0$. 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 $t \geq 0$,

$$V_t = \sqrt{t}V_1 + \sqrt{1-t}V_0 = \sqrt{t}(V_1 + \sqrt{\varepsilon_t}V_0),$$

avec $\varepsilon_t = (1-t)/t$. Ainsi, grâce à (2.4),

$$\varphi(t) = \mathbf{H}(V_1 + \sqrt{\varepsilon_t}V_0) - \lambda\mathbf{H}(X_1 + \sqrt{\varepsilon_t}X_0) - (1-\lambda)\mathbf{H}(Y_1 + \sqrt{\varepsilon_t}Y_0).$$

En dérivant, on a, par (2.14),

$$\partial_t \varphi(t) = \frac{1}{2}(\partial_t \varepsilon_t) \left(\mathbf{J}(V_1 + \sqrt{\varepsilon_t}V_0) - \lambda\mathbf{J}(X_1 + \sqrt{\varepsilon_t}X_0) - (1-\lambda)\mathbf{J}(Y_1 + \sqrt{\varepsilon_t}Y_0) \right).$$

Ceci permet, en appliquant l'inégalité de BLACHMAN-STAM (2.12), de voir la croissance de φ , et par là, la positivité de $\varphi(1)$, qui est exactement (2.15).

2.4.2 Deuxième méthode

Cette méthode est directe. Nous n'en présentons ici que les grandes lignes. En partant de l'inégalité de BLACHMAN-STAM (2.12), on peut démontrer l'inégalité de SOBOLEV logarithmique (2.8) pour un vecteur aléatoire X de carré intégrable en utilisant une « méthode de semi-groupe » (voir [Car91]). L'idée est de considérer l'entropie relative $\mathbf{Ent}(X | Z)$, où Z est un vecteur aléatoire normal indépendant de X , et d'écrire que

$$\mathbf{Ent}(X | Z) = - \int_0^\infty \partial_t \mathbf{Ent}(X_t | Z) dt,$$

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 (2.12). Lorsqu'on intègre ensuite en t , on obtient une majoration de $\mathbf{Ent}(X|Z)$ qui implique une inégalité paramétrée par t , qui, elle-même, fournit l'inégalité de SOBOLEV logarithmique (2.8) après une optimisation en t .

2.4.3 Troisième méthode

Cette dernière méthode passe par une nouvelle inégalité, la concavité de l'entropie exponentielle, liée au critère « courbure-dimension » ou encore « critère $\mathbf{I}_2$ » (défini au début du chapitre 5 du livre [ABC⁺00]).

Proposition 2.4.7 (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 ,*

$$(2.17) \quad \partial_t \left[\mathbf{J} \left(X + \sqrt{t}Z \right)^{-1} \right] \geq \frac{1}{n}.$$

Le nom de cette propriété vient de ce qu'elle peut se réécrire sous la forme

$$\partial_t^2 \mathbf{N} \left(X + \sqrt{t}Z \right) \leq 0.$$

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 (2.12) (voir [DCT91]), mais nous préférons ici utiliser le « critère $\mathbf{I}_2$ » pour le semi-groupe de la chaleur. Considérons son générateur $\mathbf{L} = (1/2)\Delta$ sur $\mathbb{R}^n$. Dans le chapitre 5 du livre [ABC⁺00], il a été vu à la remarque 5.3.2 page 78 du livre [ABC⁺00] que ce générateur vérifie $\mathbf{I}_2 \geq (\mathbf{L})^2/n$. Cette propriété, notée $\text{CD}(0, n)$, est le critère courbure-dimension de courbure nulle et de dimension n . C'est le terme dimensionnel qui est important ici. Posons $\mathbf{J}_t \stackrel{\text{def}}{=} \mathbf{J}(\mathbf{P}_t f) = \mathbf{J}(X + \sqrt{t}Z)$, 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 page 94 de [ABC⁺00]. Grâce à la propriété $\mathbf{I}_2 \geq (\mathbf{L})^2/n$, les calculs du paragraphe 5.6.3 de [ABC⁺00] permettent de montrer l'inégalité (5.24) de [ABC⁺00], qui se traduit dans notre contexte par

$$-\partial_t \mathbf{J}_t \geq \frac{1}{n} \mathbf{J}_t^2.$$

On en déduit alors immédiatement (2.17).

Nous nous proposons maintenant de retrouver l'inégalité de SOBOLEV logarithmique (2.8) à partir de cette inégalité (2.17). 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 (2.17) et à contrôler finement $\int \mathbf{P}_t f \log \mathbf{P}_t f dx$.

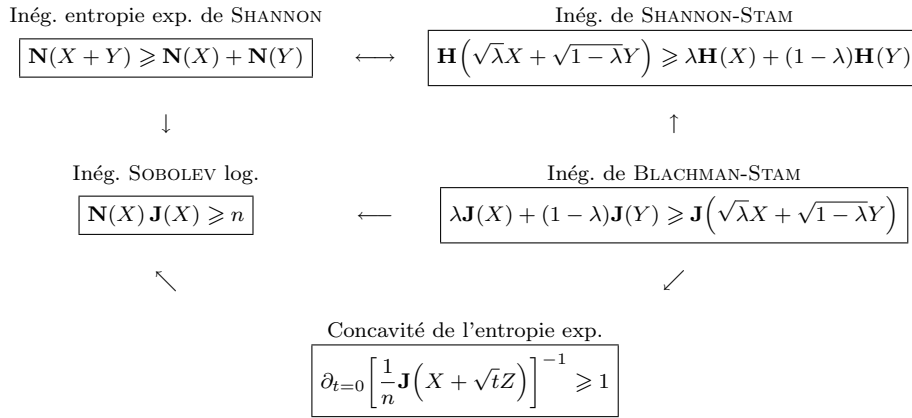
Il est aussi possible de faire comme suit. Soient X un vecteur aléatoire de $\mathbb{R}^n$ à densité, et Z un vecteur aléatoire normal standard indépendant de X . Notons alors $X_t = \sqrt{t}X + \sqrt{1-t}Z$, $\mathbf{N}_t \stackrel{\text{déf.}}{=} \mathbf{N}(X_t)$, et $\mathbf{J}_t \stackrel{\text{déf.}}{=} \mathbf{J}(X_t)$. Pour $t = 0$, il est évident que $\mathbf{N}_0 \mathbf{J}_0 = n$, et l'on veut montrer que $\mathbf{N}_1 \mathbf{J}_1 \geq n$. 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

$$\mathbf{N}_t \mathbf{J}_t = \mathbf{N}(X + \sqrt{\varepsilon}Z) \mathbf{J}(X + \sqrt{\varepsilon}Z).$$

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 (2.17). Ceci termine notre dernière preuve de l'inégalité de SOBOLEV logarithmique. En conclusion, nous pouvons dresser le tableau suivant :



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.

2.5 L'inégalité de Young et ses conséquences

L'inégalité de SHANNON-STAM (2.15) 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 (cf. [Bec75] et [BL76]).

Théorème 2.5.1 (Inégalité de Young). *Soient $1 \leq r, p, q \leq \infty$ des nombres 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 $\mathbf{L}^q(\mathbb{R}^n)$,*

$$\|f \star g\|_r \leq \left(\frac{c_p c_q}{c_r} \right)^{n/2} \|f\|_p \|g\|_q \quad \text{avec} \quad c_p = \frac{p^{1/p}}{|p'|^{1/p'}},$$

où p' est le conjugué de HÖLDER de p (i.e. $1/p + 1/p' = 1$) et $c_1 = c_\infty = 1$.

Réciproquement, si $0 \leq r, p, q \leq 1$, alors l'inégalité est inversée :

$$\|f \star g\|_r \geq \left(\frac{c_p c_q}{c_r} \right)^{n/2} \|f\|_p \|g\|_q.$$

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 (2.15) 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

$$\mathbf{H}^p(X) = \frac{1}{1-p} \log \mathbf{E}(f(X)^{p-1}) = \frac{p}{1-p} \log \|f\|_p.$$

L'entropie $\mathbf{H}^p$ est continue en p si l'on pose $\mathbf{H}^1 \stackrel{\text{déf.}}{=} \mathbf{H}$ et $\mathbf{H}^0(X) \stackrel{\text{déf.}}{=} \log |\{f > 0\}|$ où $|C|$ désigne la mesure de LEBESGUE de C .

Théorème 2.5.2. *Soient $0 < r \leq \infty$ 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*

$$(2.18) \quad \mathbf{H}^r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda \mathbf{H}^p(X) - (1-\lambda) \mathbf{H}^q(Y) \\ \geq \mathbf{H}^r(Z) - \lambda \mathbf{H}^p(Z) - (1-\lambda) \mathbf{H}^q(Z),$$

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 2.4.5. 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

$$\mathbf{H}^0(\lambda X + (1-\lambda)Y) - \lambda \mathbf{H}^0(X) - (1-\lambda) \mathbf{H}^0(Y) \geq 0.$$

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 :

$$(2.19) \quad |A + B| \geq \left| \frac{A}{\lambda} \right|^\lambda + \left| \frac{B}{1-\lambda} \right|^{1-\lambda},$$

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 :

$$|A + B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}.$$

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 (2.19), dans [BL00]. Ceci complète encore notre panorama. Le théorème 2.5.2 a de très nombreuses autres conséquences (voir pour cela [DCT91]).

2.6 Principes d'incertitude

Le principe d'incertitude le plus célèbre est sans doute celui de 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 inégalités 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]).

2.6.1 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 $x_1, \dots, x_n$ 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 $x_1, \dots, x_n$ sont des réalisations de variables aléatoires $X_1, \dots, X_n$ 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 $x_1, \dots, x_n$, appelée *estimateur*, qui soit « proche » de θ . Un choix très naturel est la moyenne arithmétique

$$Y(x_1, \dots, x_n) \stackrel{\text{déf.}}{=} \frac{x_1 + \dots + x_n}{n}.$$

Cet estimateur a l'avantage d'avoir exactement θ comme espérance, puisque

$$\mathbf{E}(Y) = \mathbf{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \mathbf{E}(X_1) = \theta.$$

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*¹². 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

$$\mathbf{L}_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i - \theta),$$

où f est la densité de la mesure gaussienne $\mathcal{N}(0, \sigma)$:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

¹² *Likelihood* en anglais, d'où la notation $\mathbf{L}_\theta$.

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 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

$$\mathbb{I}(\theta) = \mathbf{E}_\theta(\nabla_\theta \log \mathbf{L}_\theta \cdot \nabla_\theta \log \mathbf{L}_\theta^\top).$$

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 2.6.1 (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 $\nabla_\theta (\int \mathbf{L}_\theta \cdot Y d\mu) = \int (\nabla_\theta \mathbf{L}_\theta \cdot Y) d\mu$. On a alors*

$$\mathbf{Var}_\theta(Y) \geq \nabla_\theta \mathbf{E}_\theta(Y)^\top \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_\theta \mathbf{E}_\theta(Y).$$

Supposons que l'on désire estimer une fonction réelle F du paramètre θ . Le théorème 2.6.1 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 positive¹³ $\nabla_\theta F(\theta)^\top \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_\theta F(\theta)$, 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

$$\mathbf{Var}_\theta(Y) \geq \mathbb{I}(\theta)^{-1},$$

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 $Y = (x_1 + \dots + x_n)/n$ est de variance minimale.

¹³Les estimateurs sans biais pour lesquels la borne inférieure de l'inégalité est atteinte sont dits « efficaces ». Il n'en existe pas toujours. Dans l'exemple du sonar, $Y = (x_1 + \dots + x_n)/n$ est efficace.

Preuve du théorème 2.6.1. 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

$$\langle v, \nabla_{\theta} \mathbf{E}_{\theta}(Y) \rangle = \mathbf{E}_{\theta}(\langle v, \nabla_{\theta} \log \mathbf{L}_{\theta} \rangle (Y - \mathbf{E}_{\theta}(Y))),$$

et par l'inégalité de CAUCHY-SCHWARZ, on en déduit que

$$\mathbf{Var}_{\theta}(Y) \geq \frac{\langle v, \nabla_{\theta} \mathbf{E}_{\theta}(Y) \rangle^2}{\mathbf{E}_{\theta}(\langle v, \nabla_{\theta} \log \mathbf{L}_{\theta} \rangle^2)}.$$

En choisissant $v = \mathbb{I}(\theta)^{-1} \cdot \nabla_{\theta} \mathbf{E}_{\theta}(Y)$, on obtient exactement la conclusion du théorème :

$$\mathbf{Var}_{\theta}(Y) \geq \nabla_{\theta} \mathbf{E}_{\theta}(Y)^{\top} \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_{\theta} \mathbf{E}_{\theta}(Y).$$

□

Le théorème 2.6.1 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 2.6.1 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

$$(2.20) \quad K_{\theta}(Y) \geq \nabla_{\theta} F(\theta)^{\top} \cdot \mathbb{I}(\theta)^{-1} \cdot \nabla_{\theta} F(\theta),$$

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 (2.20) 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

$$(2.21) \quad \mathbb{J}(X) \stackrel{\text{déf.}}{=} \int \nabla f \cdot \nabla f^\top \frac{dx}{f} = \int \nabla \log f \cdot \nabla \log f^\top f dx.$$

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 (2.10) :

$$\text{Tr}(\mathbb{J}(X)) = \mathbf{J}(X).$$

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 (2.20), on obtient le résultat suivant (théorème 2.6.2). 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 2.6.2 (Principe d'incertitude de Cramér-Rao). *Si X est un vecteur aléatoire de covariance K_X , on a au sens des formes quadratiques*

$$(2.22) \quad K_X \geq \mathbb{J}(X)^{-1}.$$

En dimension 1, ce résultat est une conséquence directe de l'inégalité de SOBOLEV logarithmique (2.11), grâce à l'optimalité des lois gaussiennes à variance fixée (2.5), 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 (2.12) à partir du principe d'incertitude de CRAMÉR-RAO.

2.6.2 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 ». 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 » 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 2.6.3 (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*

$$(2.23) \quad 16\pi^2 K_Y - K_X^{-1} \geq 0 \quad \text{et de façon symétrique} \quad 16\pi^2 K_X - K_Y^{-1} \geq 0.$$

Une démonstration fait appel à une inégalité de PITT (voir [Bec95]), qui elle-même découle de l'inégalité de YOUNG (théorème 2.5.1).

Le célèbre principe d'incertitude de HEISENBERG s'écrit en dimension 1 comme suit :

$$2\pi\sigma_X\sigma_Y \geq h,$$

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'écrit $4\pi\sigma_X\sigma_Y \geq 1$. 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

$$(2.24) \quad 16\pi^2 K_Y - \mathbb{J}(X) \geq 0 \quad \text{et de façon symétrique} \quad 16\pi^2 K_X - \mathbb{J}(Y) \geq 0,$$

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 (2.24) permettent de passer du principe d'incertitude de WEYL-HEISENBERG à celui de CRAMÉR-RAO et réciproquement.

2.6.3 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 (2.26) (ci-dessous) plus forte que la version euclidienne de l'inégalité de SOBOLEV logarithmique gaussienne (2.11).

Théorème 2.6.4 (Principe d'incertitude de Beckner-Hirschman). *Si X et Y sont deux vecteurs aléatoires associés, alors*

$$(2.25) \quad 16\pi^2 \mathbf{N}(X) \mathbf{N}(Y) \geq 1.$$

En utilisant le maximum gaussien à variance fixé (2.5), on en déduit que

$$16\pi^2 |K_X|^{1/n} |K_Y|^{1/n} \geq 1,$$

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

$$\|\varphi\|_{p'} \leq c_p^{n/2} \|\psi\|_p,$$

où p et p' sont des conjugués de HÖLDER avec $1 < p \leq 2$, et où c_p est la constante intervenant dans l'inégalité de YOUNG (2.5.1). Cette inégalité permet d'ailleurs de retrouver l'inégalité de YOUNG (2.5.1) sur le domaine $1 \leq p, q \leq 2$ et $2 \leq r$ (voir [Bec75]).

Preuve du théorème 2.6.4. Notons $|\varphi|^2$ la densité de X et $|\widehat{\varphi}|^2$ celle de Y . Alors, par l'inégalité de HAUSDORFF-YOUNG, on a

$$\log \|\varphi\|_{p'} - \log \|\widehat{\varphi}\|_p \leq \log c_p^{n/2}$$

pour $1 < p \leq 2$, c'est-à-dire $p' \geq 2$, et avec égalité en $p = p' = 2$. On peut donc dériver cette inégalité en $p' = 2$. On a

$$\partial_{p'=2} \log \|\varphi\|_{p'} = -\frac{1}{4} \mathbf{H}(|\varphi|^2) = -\frac{1}{4} \mathbf{H}(X),$$

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 :

$$\mathbf{H}(X) + \mathbf{H}(Y) \geq n(1 - \log 2).$$

□

Nous allons voir que le principe d'incertitude de BECKNER-HIRSCHMAN (2.25) implique l'inégalité suivante, valable pour tout vecteur aléatoire X à densité :

$$(2.26) \quad \mathbf{N}(X) |\mathbb{J}(X)|^{1/n} \geq 1.$$

Cette inégalité est plus forte que la version euclidienne (2.11) 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 (2.11) au vecteur aléatoire $\mathbb{J}(X)^{1/2} X$ pour obtenir immédiatement la version forte (2.26) (cet argument est dû à [Dem90]). Ainsi, même si à X fixé, l'inégalité forte (2.26) est plus fine que (2.11), ces deux inégalités fonctionnelles sont équivalentes.

Pour prouver l'inégalité (2.26), 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 (2.24). Le maximum gaussien de l'entropie à variance fixée (2.5) permet de faire la majoration

$$16\pi^2 \mathbf{N}(Y) \leq 16\pi^2 |K_Y|^{1/n} = |\mathbb{J}(X)|^{1/n},$$

et enfin le principe d'incertitude de BECKNER-HIRSCHMAN (2.25) permet d'obtenir (2.26).

Nous allons encore renforcer le lien entre l'inégalité de SOBOLEV logarithmique et le principe d'incertitude de BECKNER-HIRSCHMAN. Ce principe d'incertitude (2.25) concerne la mesure de LEBESGUE. En notant $|\varphi|^2$ la densité de X , il s'écrit de façon additive par

$$\mathbf{H}(|\varphi|^2) + \mathbf{H}(|\widehat{\varphi}|^2) \geq n(1 - \log 2).$$

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

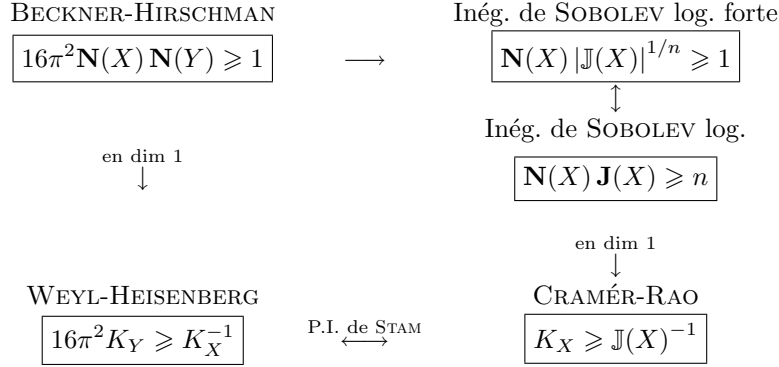
$$(2.27) \quad \mathbf{Ent}_{\mathcal{N}(0, \frac{I_n}{4\pi})}(|f|^2) + \mathbf{Ent}_{\mathcal{N}(0, \frac{I_n}{4\pi})}(|\mathcal{W}(f)|^2) \leq \frac{1}{2\pi} \mathbf{E}_{\mathcal{N}(0, \frac{I_n}{4\pi})}(|\nabla f|^2),$$

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 (2.27), on reconnaît un renforcement de l'inégalité de SOBOLEV logarithmique gaussienne (2.7), 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.



Comme pour le schéma précédent, ces inégalités sont toutes vraies, et les flèches désignent des passages mathématiques clairs.

2.7 Notes

La théorie de l'information, appelée également théorie de la communication¹⁴, 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]¹⁵ et WIENER [Wie48] sur l'entropie, rejoignant ainsi la notion d'entropie thermodynamique comme mesure du « désordre », 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], [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, au signe près, la fonction qui apparaît dans le célèbre « théorème-H » 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'auteurs¹⁶ se sont intéressés alors, dans les décennies qui ont suivi, aux aspects mathématiques de ce que l'on appelle aujourd'hui « théorie de

¹⁴SHANNON lui même semblait préférer cette désignation.

¹⁵HALPHEN 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].

¹⁶On peut citer, parmi les plus connus, par ordre alphabétique : BLACHMAN, FANO, FEINSTEIN, KHINTCHINE, KOLMOGOROV, KULLBACK, Mac MILLAN, PINSKER, SLEPIAN et STAM.

l'information ». 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 2.5.1), 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 (2.13) 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 (2.12), et dont l'extension multidimensionnelle 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 (2.12) à partir de l'inégalité de CRAMÉR-RAO.

L'identité de DEBRUIJN, essentielle pour lier les quantités $\mathbf{H}$ et $\mathbf{J}$, 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'analogie 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 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 $4\pi^2 e^2$ 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

- [ABC⁺00] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO et G. SCHEFFER – *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses, vol. 10, Société Mathématique de France, Paris, 2000. 1, 2, 6, 12, 31, 36, 39, 40, 42, 43, 45, 64, 66, 72
- [App96] D. APPLEBAUM – *Probability and information, an integrated approach*, Cambridge University Press, Cambridge, 1996. 58
- [BCL97] D. BAKRY, D. CONCORDET et M. LEDOUX – « Optimal heat kernel bounds under logarithmic Sobolev inequalities », *ESAIM Probab. Statist.* **1** (1997), p. 391–407 (electronic). 45
- [Bec75] W. BECKNER – « Inequalities in Fourier analysis », *Ann. of Math. (2)* **102** (1975), no. 1, p. 159–182. 46, 54, 57
- [Bec95] —, « Pitt's inequality and the uncertainty principle », *Proc. Amer. Math. Soc.* **123** (1995), no. 6, p. 1897–1905. 53
- [Bec99] —, « Geometric asymptotics and the logarithmic Sobolev inequality », *Forum Math.* **11** (1999), no. 1, p. 105–137. 40, 64
- [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. 58
- [BL76] H. J. BRASCAMP et E. H. LIEB – « Best constants in Young's inequality, its converse, and its generalization to more than three functions », *Adv. Math.* **20** (1976), no. 2, p. 151–173. 46
- [BL00] S. G. BOBKOV et M. LEDOUX – « From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities », *Geom. Funct. Anal.* **10** (2000), no. 5, p. 1028–1052. 48
- [Bla65] N. M. BLACHMAN – « The convolution inequality for entropy powers », *IEEE Trans. Information Theory* **IT-11** (1965), p. 267–271. 41, 57
- [Bri64] L. BRILLOUIN – *Scientific uncertainty, and information*, Academic Press, New York, 1964. 30, 56

- [Car91] E. CARLEN – « Super-additivity of Fisher's information and logarithmic Sobolev inequalities », *J. Funct. Anal.* **101** (1991), p. 194–211. 40, 41, 44, 55, 64
- [CGG89] T. M. COVER, P. GÁCS et R. M. GRAY – « Kolmogorov's contributions to information theory and algorithmic complexity », *Ann. Probab.* **17** (1989), no. 3, p. 840–865. 58
- [CT91] T. M. COVER et J. A. THOMAS – *Elements of information theory*, John Wiley & Sons Inc., New York, 1991, A Wiley-Interscience Publication. 32, 33, 36, 37, 38, 39, 57, 58, 64, 66
- [DCD82] D. DACUNHA-CASTELLE et M. DUFLO – *Probabilités et statistiques. Tome 1*, Masson, Paris, 1982, Problèmes à temps fixe. 51
- [DCT91] A. DEMBO, T. M. COVER et J. A. THOMAS – « Information-theoretic inequalities », *IEEE Trans. Inform. Theory* **37** (1991), no. 6, p. 1501–1518. 37, 44, 45, 46, 47, 48, 57, 58, 64
- [Dem90] A. DEMBO – « Information inequalities and uncertainty principles », Tech. Rep., Dept. of Statist., Stanford Univ., 1990. 55, 57, 64
- [DL84] R. DAUTRAY et 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. 53
- [EE90] P. EHRENFEST et 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. 56
- [Fan61] R. M. FANO – *Transmission of information: A statistical theory of communications.*, The M.I.T. Press, Cambridge, Mass., 1961. 58
- [Fei58] A. FEINSTEIN – *Foundations of information theory*, McGraw-Hill Electrical and Electronic Engineering Series. McGraw-Hill Book Co., Inc., New York-Toronto-London, 1958. 58
- [Fis22] A. FISHER, R. – « On the mathematical foundations of theoretical statistics », *Philos. Trans. Roy. Soc. Ann. Probab., London, Sec. A* (1922), no. 222, p. 309–368. 56
- [Fis25] — , « Theory of statistical estimation », *Proc. Cambridge Phil. Society* (1925), no. 22, p. 700–725. 56
- [FS98] H. G. FEICHTINGER et T. STROHMER (éds.) – *Gabor analysis and algorithms*, Birkhäuser Boston Inc., Boston, MA, 1998, Theory and applications. 56

- [Geo88] H.-O. GEORGII – *Gibbs measures and phase transitions*, Walter de Gruyter & Co., Berlin, 1988. 36
- [Gra90] R. M. GRAY – *Entropy and information theory*, Springer-Verlag, New York, 1990. 58
- [Har28] V. HARTLEY, R. – « Transmission of information », *Bell Sys. Tech. J.* (1928), no. 7, p. 535. 56
- [Hir57] I. I. HIRSCHMAN, JR. – « A note on entropy », *Amer. J. Math.* **79** (1957), p. 152–156. 57
- [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. 56
- [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. 56
- [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. 30, 37, 38, 58
- [KKK87] S. KULLBACK, J. C. KEEGEL et J. H. KULLBACK – *Topics in statistical information theory*, Springer-Verlag, Berlin, 1987. 58
- [Kul97] S. KULLBACK – *Information theory and statistics*, Dover Publications Inc., Mineola, NY, 1997, Retirage de la seconde édition de 1968. 58
- [Lie78] E. H. LIEB – « Proof of an entropy conjecture of Wehrl », *Comm. Math. Phys.* **62** (1978), no. 1, p. 35–41. 47
- [Lie90] —, « Gaussian kernels have only Gaussian maximizers », *Invent. Math.* **102** (1990), no. 1, p. 179–208. 47, 57
- [ME81] N. F. MARTIN et J. W. ENGLAND – « Mathematical theory of entropy. Foreword by James K. Brooks », *Encyclopedia of Mathematics and its Applications*, vol. 12, Addison-Wesley Publishing Company, 1981. 27, 56
- [Nyq24] H. NYQUIST – « Certain factors affecting telegraph speed », *Bell Sys. Tech. J.* (1924), no. 3, p. 324. 56
- [Rom92] S. ROMAN – *Coding and information theory*, Springer-Verlag, New York, 1992. 30, 31, 32, 33, 37, 38, 58
- [Rom97] —, *Introduction to coding and information theory*, Springer-Verlag, New York, 1997. 58
- [Sha48] C. E. SHANNON – « A mathematical theory of communication », *Bell System Tech. J.* **27** (1948), p. 379–423, 623–656. 29, 30, 32, 38, 39, 41, 56, 57, 63, 64, 66

- [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. 58
- [Sta59] A. STAM – « Some inequalities satisfied by the quantities of information of Fisher and Shannon », *Information and Control* **2** (1959), p. 101–112. 41, 42, 53, 57, 64
- [Sto88] J. STORER – *Data compression. Methods and theory*, Computer Science Press, 1988. 33
- [SW49] C. E. SHANNON et W. WEAVER – *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Ill., 1949. 56
- [Tos91] G. TOSCANI – « On Shannon's entropy power inequality », *Ann. Univ. Ferrara Sez. VII (N.S.)* **37** (1991), p. 167–184 (1992). 57
- [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. 58
- [Voi98] D. VOICULESCU – « The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms », *Invent. Math.* **132** (1998), no. 1, p. 189–227. 57
- [Wie48] N. WIENER – *Cybernetics, or Control and Communication in the Animal and the Machine*, Hermann et Cie., Paris, 1948, Actualités Sci. Ind., no. 1053. 56
- [Zam98] R. ZAMIR – « A proof of the Fisher information inequality via a data processing argument », *IEEE Trans. Inform. Theory* **44** (1998), no. 3, p. 1246–1250. 41, 52, 57
- [ZF93] R. ZAMIR et M. FEDER – « A generalization of the entropy power inequality with applications. », *IEEE Trans. Inform. Theory* **39** (1993), no. 5, p. 1723–1728 (English). 57
- [Zin96] M. ZINSMEISTER – *Formalisme thermodynamique et systèmes dynamiques holomorphes*, Panoramas et Synthèses, vol. 4, Société Mathématique de France. Paris, 1996. 27, 56

Chapitre 3

Gaussian maximum of entropy and reversed log-Sobolev inequality

À paraître in « Séminaire de Probabilité » XXXVI (2002).

Abstract

The aim of this note is to connect a reversed form of the Gross logarithmic Sobolev inequality with the Gaussian maximum of Shannon's entropy power. There is thus a complete parallel with the well-known link between logarithmic Sobolev inequalities and their information theoretic counterparts. We moreover provide an elementary proof of the reversed Gross inequality via a two-point inequality and the Central Limit Theorem.

3.1 Shannon's entropy power and Gross's inequality

In the sequel, we denote by $\mathbf{Ent}_\mu(f)$ the entropy of a non-negative integrable function f with respect to a positive measure μ , defined by

$$\mathbf{Ent}_\mu(f) := \int f \log f d\mu - \int f d\mu \log \int f d\mu.$$

The Shannon entropy [Sha48] of an n -variate random vector X with probability density function (pdf) f is given by

$$\mathbf{H}(X) := -\mathbf{Ent}_{\lambda_n}(f) = - \int f \log f dx,$$

where dx denotes the n -dimensional Lebesgue measure on $\mathbb{R}^n$. The Shannon entropy power [Sha48] of X is then given by

$$\mathbf{N}(X) := \frac{1}{2\pi e} \exp \left(\frac{2}{n} \mathbf{H}(X) \right).$$

It is well-known (cf. [Sha48, CT91]) that Gaussians saturates this entropy at fixed covariance. Namely, for any n -variate random vector X with covariance matrix $\mathbf{K}(X)$, one have

$$(3.1) \quad \mathbf{N}(X) \leq |\mathbf{K}(X)|^{1/n},$$

and $|\mathbf{K}|^{1/n}$ is the entropy power of the n -dimensional Gaussian with covariance $\mathbf{K}$.

The logarithmic Sobolev inequality of Gross [Gro75] expresses that for any non-negative smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$

$$(3.2) \quad 2 \mathbf{Ent}_{\gamma_n}(f) \leq \mathbf{E}_{\gamma_n} \left(\frac{|\nabla f|^2}{f} \right),$$

where $\mathbf{E}_{\gamma_n}$ denotes the expectation with respect to γ_n , $|\cdot|$ the Euclidean norm and γ_n the n -dimensional standard Gaussian given by

$$d\gamma_n(x) := (2\pi)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2}} dx.$$

Inequality (3.2) is sharp and the equality is achieved for f of the form $\exp(a \cdot)$.

By performing a change of function and an optimization, Beckner showed [Bec99] (see also [Car91]) that (3.2) is equivalent to the following “Euclidean” logarithmic Sobolev inequality, for any pdf g

$$(3.3) \quad \mathbf{Ent}_{\lambda_n}(g) \leq \frac{n}{2} \log \left[\frac{1}{2\pi en} \int \frac{|\nabla g|^2}{g} dx \right],$$

where λ_n is the n -dimensional Lebesgue measure on $\mathbb{R}^n$. Therefore, for any n -variate random vector X (with pdf g), we have

$$(3.4) \quad \mathbf{N}(X) \mathbf{J}(X) \geq n.$$

This inequality can be obtained by many methods. The most classical ones are via Shannon’s entropy power inequality together with DeBruijn identity, or via Stam’s super-additivity of the Fisher information (cf. [Sta59, Car91, DCT91, ABC⁺00]). Moreover, Dembo showed in [Dem90] that (3.4) is equivalent to

$$(3.5) \quad \mathbf{N}(X) |\mathbb{J}(X)|^{1/n} \geq 1,$$

where $\mathbb{J}(X)$ is the Fisher information matrix of X defined by

$$\mathbb{J}(X) := \int \nabla \log g (\nabla \log g)^\top g dx,$$

where $\nabla \log g$ is a column vector and $(\nabla \log g)^\top$ the associated row vector. Thus, we simply have $\mathbf{J}(X) = \text{Tr } \mathbb{J}(X)$. To deduce (3.5) from (3.4), apply (3.4) to the random vector $X = \mathbf{K}(Y)^{-1/2} Y$. Conversely, use the arithmetic-geometric means inequality

$$(3.6) \quad (a_1 \cdots a_n)^{\frac{1}{n}} \leq \frac{a + \cdots + a_n}{n}$$

on the spectrum of the non-negative symmetric matrix $\mathbb{J}(X)$.

3.2 Reversed Gross's logarithmic Sobolev inequality

The Gross logarithmic Sobolev inequality (3.2) admits a reversed form which states that for any positive smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$

$$(3.7) \quad \frac{|\mathbf{E}_{\gamma_n}(\nabla f)|^2}{\mathbf{E}_{\gamma_n}(f)} \leq 2 \mathbf{Ent}_{\gamma_n}(f).$$

Here again, the 2 constant is optimal and the equality is achieved for f of the form $\exp(a \cdot)$. Alike for (3.2), one can show by a change of function and an optimization that the reverse form (3.7) is equivalent to the following inequality, for any pdf g

$$(3.8) \quad -\mathbf{Ent}_{\lambda_n}(g) \leq \frac{n}{2} \log \left[2\pi e \frac{\text{Tr } \mathbf{K}(g)}{n} \right],$$

where $\mathbf{K}(g)$ is the covariance matrix of the pdf g . Hence, we have for any n -variate random vector X with pdf

$$(3.9) \quad \mathbf{N}(X) \leq \frac{\text{Tr } \mathbf{K}(X)}{n},$$

where $\mathbf{K}(X)$ denotes the covariance matrix of X . This inequality is optimal and is achieved by Gaussians X . Moreover, as we will show, inequality (3.9) is equivalent to (3.1).

Summarizing, we obtain the following statement

Theorem 3.2.1. *The following assertions are true and equivalent*

(i) *For any smooth $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$,*

$$|\mathbf{E}_{\gamma_n}(\nabla f)|^2 \leq 2 \mathbf{Ent}_{\gamma_n}(f) \mathbf{E}_{\gamma_n}(f).$$

(ii) *For any smooth $g : \mathbb{R}^n \rightarrow \mathbb{R}^+$,*

$$-\mathbf{Ent}_{\lambda_n}(g) \leq \frac{n}{2} \log \left[\frac{2\pi e}{n} \text{Tr } \mathbf{K}(g) \right].$$

(iii) *For any n -variate random vector X with smooth pdf,*

$$n\mathbf{N}(X) \leq \text{Tr } \mathbf{K}(X).$$

(iv) *For any n -variate random vector X with smooth pdf,*

$$\mathbf{N}(X) \leq |\mathbf{K}(X)|^{1/n}.$$

Therefore, there is a complete parallel between the equivalence between (3.2), (3.3), (3.4), (3.5) in one hand and the equivalence between (3.7), (3.8), (3.9), (3.1) in the other hand.

3.3 Sketches of proofs

In this section, we present first two proofs of (3.7), then we explain how to deduce (3.8) from (3.7) and (3.1) from (3.9) and vice versa.

The most natural way to establish (3.7) is to start from a two-point. inequality, just like Gross does for the logarithmic Sobolev inequality (3.2) in [Gro75]. Namely, if we denote by β the symmetric Bernoulli measure on $\{-1, +1\}$, one can show easily that for any non-negative function $f : \{-1, +1\} \rightarrow \mathbb{R}^+$,

$$(3.10) \quad (f(+1) - f(-1))^2 \leq \frac{1}{8} \mathbf{Ent}_\beta(f) \mathbf{E}_\beta(f).$$

This inequality is nothing else but the Csiszár-Kullback inequality for β (see [Pin64]). Actually, the optimal constant for the Bernoulli measure of parameter $p = 1 - q$ is $p^2 q^2 (\log q - \log p) / (q - p)$, which resembles the optimal constant for the logarithmic Sobolev inequality (see for example [ABC⁺00]), but here again, only the symmetric case gives the optimal constant of (3.7).

The next step is to establish the following chain rule formula for $\mathbf{Ent}$, which generalizes the classical chain rule formula (cf. [Sha48, CT91]) for $\mathbf{H}$

Proposition 3.3.1. *For any positive measures μ_i and their product μ on the product space, and for any bounded real valued measurable function f on the product space, we have*

$$(3.11) \quad \mathbf{Ent}_\mu(f) \geq \sum_{i=1}^n \mathbf{Ent}_{\mu_i} \left(\mathbf{E}_{\mu_{\setminus i}}(f) \right),$$

where $\mu_{\setminus i}$ denotes the product of the measures μ_j with $j \neq i$.

Finally, inequality (3.7) can then be recovered by the use of the Central Limit Theorem and integration by parts (in both discrete and Gaussian forms). This concludes the first proof of (3.7).

Actually, inequality (3.7) can be recovered by a simple semi-group argument, just like for the logarithmic Sobolev inequality (3.2) (cf. [Led00]). Namely, consider the heat semi-group $(\mathbf{P}_t)_{t \geq 0}$ on $\mathbb{R}^n$, acting on a bounded continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ as follows

$$\mathbf{P}_t(f)(x) := \int_{\mathbb{R}^n} f(x + \sqrt{t} y) d\gamma_n(y).$$

Notice that for any smooth function f , $\nabla \mathbf{P}_t(f) = \mathbf{P}_t(\nabla f)$ and

$$\partial_t \mathbf{P}_t(f) = \frac{1}{2} \Delta \mathbf{P}_t(f) = \frac{1}{2} \mathbf{P}_t(\Delta f).$$

Now, for any smooth positive bounded function $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$, any $t \geq 0$ and any x , we can write, by performing an integration by parts and omitting the x variable

$$\begin{aligned} \mathbf{P}_t(f \log f) - \mathbf{P}_t(f) \log \mathbf{P}_t(f) &= \int_0^t \partial_s [\mathbf{P}_s(\mathbf{P}_{t-s}(f) \log \mathbf{P}_{t-s}(f))] ds \\ &= \frac{1}{2} \int_0^t \mathbf{P}_s \left(\frac{|\nabla \mathbf{P}_{t-s}(f)|^2}{\mathbf{P}_{t-s}(f)} \right) ds. \end{aligned}$$

But by Cauchy-Schwarz inequality we get

$$\mathbf{P}_s \left(\frac{|\nabla \mathbf{P}_{t-s}(f)|^2}{\mathbf{P}_{t-s}(f)} \right) = \mathbf{P}_s \left(\frac{|\mathbf{P}_{t-s}(\nabla f)|^2}{\mathbf{P}_{t-s}(f)} \right) \geq \frac{|\mathbf{P}_s(\mathbf{P}_{t-s}(\nabla f))|^2}{\mathbf{P}_s(\mathbf{P}_{t-s}(f))},$$

which gives

$$\mathbf{P}_t(f \log f) - \mathbf{P}_t(f) \log \mathbf{P}_t(f) \geq \frac{t}{2} \frac{|\mathbf{P}_t(\nabla f)|^2}{\mathbf{P}_t(f)}.$$

Finally, inequality (3.7) follows by taking $(t, x) = (1, 0)$. Notice that this method gives also the logarithmic Sobolev inequality (3.2). Namely, by Cauchy-Schwarz inequality

$$|\mathbf{P}_{t-s}(\nabla f)|^2 \leq \mathbf{P}_{t-s}(|\nabla f|^2) \leq \mathbf{P}_{t-s}(f) \mathbf{P}_{t-s} \left(\frac{|\nabla f|^2}{f} \right),$$

therefore, we obtain

$$\begin{aligned} \mathbf{P}_t(f \log f) - \mathbf{P}_t(f) \log \mathbf{P}_t(f) &= \frac{1}{2} \int_0^t \mathbf{P}_s \left(\frac{|\mathbf{P}_{t-s}(\nabla f)|^2}{\mathbf{P}_{t-s}(f)} \right) ds \\ &\leq \frac{1}{2} \int_0^t \mathbf{P}_s \left(\mathbf{P}_{t-s} \left(\frac{|\nabla f|^2}{f} \right) \right) ds \\ &= \frac{t}{2} \mathbf{P}_t \left(\frac{|\nabla f|^2}{f} \right), \end{aligned}$$

which gives (3.2) by taking here again $(t, x) = (1, 0)$.

To deduce (3.8) from (3.7), just apply (3.7) to

$$f(x) = h(x) (2\pi)^{\frac{n}{2}} e^{\frac{|x|^2}{2}}$$

where h is a compactly supported smooth pdf. One then gets

$$\left| \int x h dx + \int \nabla h dx \right|^2 \leq 2 \int h \log h dx + \int |x|^2 h dx + n \log(2\pi).$$

But we have $\int \nabla h \, dx = 0$. Therefore, by denoting $\mathbf{K}(h)$ the covariance matrix associated with the pdf h , one gets

$$-\mathbf{Ent}_{\lambda_n}(h) \leq \frac{1}{2} \text{Tr } \mathbf{K}(h) + \frac{n}{2} \log(2\pi),$$

which remains true for any smooth pdf h . Finally, by performing the change of function $h = \alpha g(\alpha \cdot)$ and optimizing in α , one obtains

$$-\mathbf{Ent}_{\lambda_n}(h) \leq \frac{n}{2} \log \left[2\pi e \frac{\text{Tr } \mathbf{K}(h)}{n} \right]$$

which is nothing else than (3.9). Conversely, it is easy to see that we can recover (3.7) for any pdf f by approximating f by compactly supported probability density functions.

The equivalence between inequalities (3.9) and (3.1) is obtained as for the equivalence between (3.4) and (3.5). Namely, to deduce (3.1) from (3.9), apply (3.9) to the random vector

$$X = \mathbf{K}(Y)^{-1/2} Y.$$

Conversely, use the arithmetic-geometric means inequality (3.6) on the spectrum of the non-negative symmetric matrix $\mathbf{K}(X)$.

3.4 Remarks

It is well-known that the logarithmic Sobolev inequality (3.2) is a consequence of the Gaussian isoperimetric inequality [Led00]. In contrast, it is shown in [BCEF01] that the reversed form (3.7) is equivalent to a translation property [Bob99]. Namely, for any smooth function $f : \mathbb{R}^n \rightarrow [0, 1]$

$$(3.12) \quad |\mathbf{E}_{\gamma_n}(\nabla f)| \leq \mathbf{I}(\mathbf{E}_{\gamma_n}(f)),$$

where $\mathbf{I}$ is the Gaussian isoperimetric function given by $\mathbf{I} := \Phi' \circ \Phi^{-1}$, where Φ is the Gaussian distribution function given by $\Phi(\cdot) := \gamma((-\infty, \cdot])$. Bobkov's inequality (3.12) expresses that among all measurable sets with fixed Gaussian measure, half spaces have maximum barycenter [BCEF01].

Acknowledgements

The author would like to thanks Michel Ledoux and Cécile Ané for their encouragements and helpful comments.

Bibliography

[ABC⁺00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*,

- Panoramas et Synthèses, vol. 10, Société Mathématique de France, Paris, 2000. 1, 2, 6, 12, 31, 36, 39, 40, 42, 43, 45, 64, 66, 72
- [BCEF01] F. Barthe, D. Cordero-Erausquin, and M. Fradelizi, *Shift inequalities of Gaussian type and norms of barycentres*, *Studia Math.* **146** (2001), no. 3, 245–259. 7, 21, 68
- [BCL97] D. Bakry, D. Concordet, and M. Ledoux, *Optimal heat kernel bounds under logarithmic Sobolev inequalities*, *ESAIM Probab. Statist.* **1** (1997), 391–407 (electronic). 45
- [Bec99] William Beckner, *Geometric asymptotics and the logarithmic Sobolev inequality*, *Forum Math.* **11** (1999), no. 1, 105–137. 40, 64
- [Bla65] N. M. Blachman, *The convolution inequality for entropy powers*, *IEEE Trans. Information Theory* **IT-11** (1965), 267–271. 41, 57
- [Bob99] S. G. Bobkov, *The size of singular component and shift inequalities*, *Ann. Probab.* **27** (1999), no. 1, 416–431. 7, 20, 21, 68
- [Car91] E. Carlen, *Super-additivity of Fisher’s information and logarithmic Sobolev inequalities*, *J. Funct. Anal.* **101** (1991), 194–211. 40, 41, 44, 55, 64
- [CT91] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons Inc., New York, 1991, A Wiley-Interscience Publication. 32, 33, 36, 37, 38, 39, 57, 58, 64, 66
- [DCT91] Amir Dembo, Thomas M. Cover, and Joy A. Thomas, *Information-theoretic inequalities*, *IEEE Trans. Inform. Theory* **37** (1991), no. 6, 1501–1518. 37, 44, 45, 46, 47, 48, 57, 58, 64
- [Dem90] A. Dembo, *Information inequalities and uncertainty principles*, Tech. Rep., Dept. of Statist., Stanford Univ., 1990. 55, 57, 64
- [Gro75] L. Gross, *Logarithmic Sobolev inequalities*, *Amer. J. Math.* **97** (1975), no. 4, 1061–1083. 64, 66
- [Led99] M. Ledoux, *Concentration of measure and logarithmic Sobolev inequalities*, *Séminaire de Probabilités, XXXIII, Lecture Notes in Math.*, Springer, Berlin, 1999, pp. 120–216. 4, 20, 21
- [Led00] —, *The geometry of Markov diffusion generators*, *Ann. Fac. Sci. Toulouse Math.* (6) **9** (2000), no. 2, 305–366. 6, 7, 66, 68

- [Pin64] M. S. Pinsker, *Information and information stability of random variables and processes*, Holden-Day Inc., San Francisco, Calif., 1964, Translated by Amiel Feinstein. 66
- [Sha48] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656. 29, 30, 32, 38, 39, 41, 56, 57, 63, 64, 66
- [Sta59] A. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control **2** (1959), 101–112. 41, 42, 53, 57, 64

Chapitre 4

Glauber versus Kawasaki for spectral gap and logarithmic Sobolev inequalities of some unbounded conservative spin systems

Abstract

Inspired by the recent results of C. Landim, G. Panizo and H.-T. Yau [LPY00] on spectral gap and logarithmic Sobolev inequalities for unbounded conservative spin systems, we study uniform bounds in these inequalities for Glauber dynamics of Hamiltonian of the form

$$\sum_{i=1}^n V(x_i) + V\left(M - \sum_{i=1}^n x_i\right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

Specifically, we examine the case V is strictly convex (or small perturbation of strictly convex) and, following [LPY00], the case V is a bounded perturbation of a quadratic potential. By a simple path counting argument for the standard random walk, uniform bounds for the Glauber dynamics yields, in a transparent way, the classical L^{-2} decay for the Kawasaki dynamics on d -dimensional cubes of length L . The arguments of proofs however closely follow and make heavy use of the conservative approach and estimates of [LPY00], relying in particular on the Lu-Yau martingale decomposition and clever partitionings of the conditional measure.

Introduction

Let Q be a probability measure on $\mathbb{R}^n$. In the sequel, we denote by $\mathbf{E}_Q(f)$ the expectation of f with respect to Q , $\mathbf{Var}_Q(f) := \mathbf{E}_Q(f^2) - \mathbf{E}_Q(f)^2$ the variance of f for Q , and $\mathbf{Ent}_Q(f)$ the entropy of a non negative measurable function f with respect to Q , defined by

$$\mathbf{Ent}_Q(f) := \int f \log f dQ - \int f dQ \log \int f dQ.$$

We say that Q satisfies a Poincaré inequality if there exists a positive constant $\mathcal{P}$ such that for any smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$(4.1) \quad \mathbf{Var}_Q(f) \leq \mathcal{P} \mathbf{E}_Q(|\nabla f|^2),$$

where $|\nabla f|^2 := \sum_{i=1}^n |\partial_i f|^2$. Similarly, we say that Q satisfies a logarithmic Sobolev inequality if there exists a positive constant $\mathcal{L}$ such that for any smooth function f ,

$$(4.2) \quad \mathbf{Ent}_Q(f^2) \leq \mathcal{L} \mathbf{E}_Q(|\nabla f|^2).$$

This inequality strengthens the Poincaré inequality (4.1) since for ε small enough,

$$\mathbf{Ent}_Q((1 + \varepsilon f)^2) = 2\varepsilon^2 \mathbf{Var}_Q(f) + \mathcal{O}(\varepsilon^3),$$

which gives $2\mathcal{P} \leq \mathcal{L}$. Let $\mathcal{H} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ such that

$$Z_{\mathcal{H}} := \int_{\mathbb{R}^n} e^{-\mathcal{H}(x)} dx < +\infty.$$

The probability measure Q defined by $dQ(x) = (Z_{\mathcal{H}})^{-1} \exp(-\mathcal{H}(x)) dx$ is the symmetric invariant measure of the diffusion process $(X_t)_{t \geq 0}$ on $\mathbb{R}^n$ driven by the S.D.E.

$$dX_t = \sqrt{2} dB_t - \nabla \mathcal{H}(X_t) dt,$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion on $\mathbb{R}^n$. In this context, we say that the probability measure Q is associated with the “Hamiltonian” $\mathcal{H}$. It is well known that Q satisfies the Poincaré inequality (4.1) with a constant $\mathcal{P}$ if and only if the infinitesimal generator $\mathbf{L} := \Delta - \nabla \mathcal{H} \cdot \nabla$ possesses a spectral gap greater than $\mathcal{P}^{-1}$. In the other hand, a famous Theorem of Gross states that Q satisfies the logarithmic Sobolev inequality (4.2) if and only if the diffusion semi-group generated by $\mathbf{L}$ is hyper-contractive. A celebrated result of Bakry and Émery ensures that when there exists a constant $\rho > 0$ such that for any $x \in \mathbb{R}^n$,

$$\text{Hess}(\mathcal{H})(x) \geq \rho \mathbf{I}_p$$

as quadratic forms on $\mathbb{R}^n$, i.e. $\mathcal{H}$ is uniformly strictly convex or Q is log-concave, then Q satisfies to (4.1) and (4.2) with constants $\mathcal{P} = \rho^{-1}$ and $\mathcal{L} = 2\rho^{-1}$ respectively. Moreover, Q satisfies to (4.1) with a constant $\mathcal{P}$ if and only if

$$\mathcal{P} \mathbf{E}_Q((\mathbf{L}f)^2) \geq \mathbf{E}_Q(|\nabla f|^2)$$

for any smooth function f . The reader may find an introduction to logarithmic Sobolev inequalities and related fields in [ABC⁺00].

We are interested in the present work to particular Hamiltonians $H : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$. Let $M \in \mathbb{R}$ and define $H_M \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ by

$$(4.3) \quad H_M(x_1, \dots, x_n) := H\left(x_1, \dots, x_n, M - \sum_{i=1}^n x_i\right).$$

Assume that $Z_{H_M} < \infty$. Our aim is to establish Poincaré and logarithmic Sobolev inequalities for probability measures on $\mathbb{R}^n$ of the form

$$(4.4) \quad d\sigma_M(x_1, \dots, x_n) := (Z_{H_M})^{-1} \exp(-H_M(x)) dx_1 \cdots dx_n,$$

with constants $\mathcal{P}$ and $\mathcal{L}$ which does not depend on n and M . This investigation is motivated by the study of certain conditional probability measures. Namely, if the probability measure μ on $\mathbb{R}^{n+1}$ given by

$$(4.5) \quad d\mu(x) := (Z_H)^{-1} \exp(-H(x_1, \dots, x_{n+1})) dx_1 \cdots dx_{n+1}$$

is well-defined, i.e. $Z_H < +\infty$, then for any $M \in \mathbb{R}$, one can define the conditional probability measure μ_M by

$$(4.6) \quad \mu_M := \mu\left(\cdot \mid \sum_{i=1}^{n+1} x_i = M\right),$$

and we get, for any $f \in \mathcal{C}_b(\mathbb{R}^{n+1}, \mathbb{R})$,

$$(4.7) \quad \mathbf{E}_{\mu_M}(f) = \int_{\mathbb{R}^n} f\left(x_1, \dots, x_n, M - \sum_{i=1}^n x_i\right) d\sigma_M(x_1, \dots, x_n).$$

Thus, σ_M can be viewed as the translation of the conditional probability measure μ_M under the affine hyperplane of $\mathbb{R}^{n+1}$ of equation $x_1 + \cdots + x_{n+1} = M$. Alternatively, and following Caputo in [Cap01], the conditional probability measure μ_M can be defined from the probability measure μ given in (4.5) by adding an infinite potential outside of the affine constraint $x_1 + \cdots + x_{n+1} = M$. Namely for any bounded continuous function $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$

$$\mathbf{E}_{\mu_M}(f) = \lim_{\beta \rightarrow +\infty} \mathbf{E}_{\mu_{M,\beta}}(f).$$

where $\mu_{M,\beta}$ denotes the probability measure on $\mathbb{R}^{n+1}$ defined by

$$d\mu_{M,\beta}(x) := \mathcal{Z}_{\mu_{M,\beta}}^{-1} e^{-\beta(M-x_1-\cdots-x_{n+1})^2} d\mu(x).$$

A simple change of variable in $\mathbf{E}_{\mu_{M,\beta}}(f)$ gives that

$$\lim_{\beta \rightarrow +\infty} \mathbf{E}_{\mu_{M,\beta}}(f) = \int_{\mathbb{R}^n} f\left(x_1, \dots, x_n, M - \sum_{i=1}^n x_i\right) d\sigma_M(x_1, \dots, x_n),$$

This weak limit definition of μ_M was used by Caputo in [Cap01] in order to study the case of a convex Hamiltonian H . We do not use it in our approach. Notice that if $f \in \mathcal{C}_b(\mathbb{R}^n, \mathbb{R})$, we get from (4.7) that

$$(4.8) \quad \mathbf{E}_{\sigma_M}(f) = \int_{\mathbb{R}^{n+1}} f(x_1, \dots, x_n) d\mu_M(x_1, \dots, x_{n+1}).$$

Observe that (4.7) gives $\mathbf{E}_{\mu_M}(x_1 + \dots + x_{n+1}) = M$. Thus, when H is a symmetric function, σ_M and μ_M are exchangeable measures, i.e. invariant by any permutation of the coordinates. This holds for example when $H(x) = V(x_1) + \dots + V(x_{n+1})$. Moreover, $M - \sum_{i=1}^n x_i$ and x_j have then the same law under σ_M for any j in $\{1, \dots, n\}$ and we get

$$(4.9) \quad \mathbf{E}_{\mu_M}(x_1) = \dots = \mathbf{E}_{\mu_M}(x_{n+1}) = \mathbf{E}_{\sigma_M}(x_1) = \dots = \mathbf{E}_{\sigma_M}(x_n) = \frac{M}{n+1}.$$

Thus, the mean of μ_M and σ_M does not depend on H in this case.

Let us see now how to translate (4.1) and (4.2) for σ_M in terms of μ_M . One can observe that for any $i \in \{1, \dots, n\}$

$$\begin{aligned} \partial_i \left(f \left(x_1, \dots, x_n, M - \sum_{i=1}^n x_i \right) \right) \\ = (\partial_i f) \left(x_1, \dots, x_n, M - \sum_{i=1}^n x_i \right) - (\partial_{n+1} f) \left(x_1, \dots, x_n, M - \sum_{i=1}^n x_i \right). \end{aligned}$$

By replacing the coordinate x_{n+1} by any of the $x_1, \dots, x_n$ in (4.7), we obtain the following proposition

Proposition 4.0.1. *Let $H : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ and assume that for any permutation π of the coordinates, the probability measure σ_M^π on $\mathbb{R}^n$ defined by (4.4) and associated to $H \circ \pi$ satisfies to Poincaré (resp. logarithmic Sobolev) inequality with a constant $\mathcal{P}$ (resp. $\mathcal{L}$) which does not depend on n , M and π . Then, if μ_M is the associated conditional probability measure defined by (4.6), we get for any smooth $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$*

$$(4.10) \quad \mathbf{Var}_{\mu_M}(f) \leq \frac{\mathcal{P}}{n+1} \mathbf{E}_{\mu_M} \left(\sum_{1 \leq i, j \leq n+1} |\partial_i f - \partial_j f|^2 \right),$$

and respectively

$$(4.11) \quad \mathbf{Ent}_{\mu_M}(f^2) \leq \frac{\mathcal{L}}{n+1} \mathbf{E}_{\mu_M} \left(\sum_{1 \leq i, j \leq n+1} |\partial_i f - \partial_j f|^2 \right).$$

These inequalities leads to constants in L^2 for the “Kawasaki dynamics” associated to μ_M . Namely, consider a finite box $\Lambda := \{1, \dots, L\}^d \subset \mathbb{Z}^d$ on the lattice $\mathbb{Z}^d$ and n such that $\mathbb{R}^\Lambda \simeq \mathbb{R}^{n+1}$ (i.e. $n+1 = |\Lambda| = L^d$). There exists a constant $C > 0$ depending only on d such that for any $a \in \mathbb{R}^\Lambda$

$$(4.12) \quad \frac{1}{|\Lambda|} \sum_{i,j \in \Lambda} (a_i - a_j)^2 \leq C L^2 \sum_{\substack{i,j \in \Lambda \\ |i-j|=1}} (a_i - a_j)^2,$$

Therefore, it is straightforward to deduce from (4.10) and (4.11) that for a constant $C > 0$ which does not depend on n and M , one have

$$(4.13) \quad \mathbf{Var}_{\mu_M}(f) \leq C L^2 \sum_{\substack{k,l \in \Lambda \\ |k-l|=1}} \mathbf{E}_{\mu_M}(|\partial_i f - \partial_j f|^2),$$

and

$$(4.14) \quad \mathbf{Ent}_{\mu_M}(f^2) \leq C L^2 \sum_{\substack{k,l \in \Lambda \\ |k-l|=1}} \mathbf{E}_{\mu_M}(|\partial_i f - \partial_j f|^2).$$

Inequality (4.12) follows from a classical path counting argument (see for example section 4.2 of [SC97]). However, let us gives briefly a proof. For any i, j in Λ , consider the path Γ_{ij} inside Λ joining i and j obtained by adjusting the d coordinates one after the other. We have $|\Gamma_{ij}| \leq dL$ and for each k, l in Λ with $|k-l|=1$, the number of such paths containing the edge (k, l) is bounded above by $c_d L^{d+1}$ where $c_d > 0$ is a constant depending only on d . Now by Cauchy-Schwarz’s inequality

$$(a_i - a_j)^2 = \left[\sum_{(k,l) \in \Gamma_{ij}, |k-l|=1} (a_k - a_l) \right]^2 \leq dL \sum_{(k,l) \in \Gamma_{ij}, |k-l|=1} (a_k - a_l)^2,$$

and therefore

$$\sum_{i,j \in \Lambda} (a_i - a_j)^2 \leq dL \sum_{\substack{|k-l|=1 \\ k,l \in \Lambda}} (a_k - a_l)^2 \sum_{\substack{i,j \in \Lambda \\ \Gamma_{ij} \ni (k,l)}} 1 \leq dc_d L^{d+2} \sum_{\substack{|k-l|=1 \\ k,l \in \Lambda}} (a_k - a_l)^2,$$

which gives the desired result (4.12).

A simple example is given by uniformly strictly convex H with a constant ρ in $\mathbb{R}^{n+1}$. Namely, if there exists a constant $\rho > 0$ such that for any $x \in \mathbb{R}^{n+1}$, $\text{Hess}(H)(x) \geq \rho \mathbf{I}_{n+1}$ as quadratic forms on $\mathbb{R}^{n+1}$, then, an easy calculus gives for any $x \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$

$$\langle (\text{Hess}(H_M))(x)h, h \rangle_{\mathbb{R}^n} \geq \rho \sum_{i=1}^n h_i^2 + \rho \left(- \sum_{i=1}^n h_i \right)^2 \geq \rho \sum_{i=1}^n h_i^2.$$

Thus, H_M is uniformly strictly convex with the same constant ρ , and therefore, by the Bakry-Émery criterion, σ_M satisfies to Poincaré and logarithmic Sobolev inequalities with a constant ρ^{-1} and $2\rho^{-1}$ respectively, which does not depend on n and M . The hypotheses of Proposition 4.0.1 are full-filled since by the same calculus, $(H \circ \pi)_M$ is also uniformly strictly convex with a constant ρ . A more simple example is given by

$$H(x) = V(x_1) + \cdots + V(x_{n+1})$$

where V is in $\mathcal{C}^2(\mathbb{R}, \mathbb{R})$ with $V'' > \rho > 0$. Let us consider now another convex Hamiltonian example on $\mathbb{R}^{n+1}$ defined by

$$H(x) := \frac{1}{2(n+1)} \sum_{i,j=1}^{n+1} V_{\{i,j\}}(x_i - x_j),$$

where $V_{\{i,j\}}$ are in $\mathcal{C}^2(\mathbb{R}, \mathbb{R})$ and even. This is a so called mean-field Hamiltonian when all the $V_{\{i,j\}}$ are equal. We have for any i, j in $\{1, \dots, n+1\}$

$$(n+1) \partial_{ij}^2 H(x) = \begin{cases} \sum_{\substack{k=1 \\ k \neq i}}^{n+1} V''_{\{i,k\}}(x_i - x_k) & \text{if } i = j \\ -V''_{\{i,j\}}(x_i - x_j) & \text{if } i \neq j \end{cases}$$

Therefore, if $V''_{\{i,j\}}(u) \geq 0$ for any $u \in \mathbb{R}$ and any $i, j \in \{1, \dots, n+1\}$, i.e. $V_{\{i,j\}}$ is convex, the Gershgorin-Hadamard theorem implies that for any $x \in \mathbb{R}^{n+1}$, $\text{Hess}(H)(x) \geq 0$ as a quadratic form, and thus H is convex on $\mathbb{R}^{n+1}$. Unfortunately, since $\sum_{j=1}^{n+1} \partial_{ij}^2 H(x) = 0$ for any $i \in \{1, \dots, n+1\}$, the null space of $\text{Hess}(H)$ contains $\mathbf{1}_{n+1}$ and therefore, the measure μ on $\mathbb{R}^{n+1}$ defined by $d\mu(x) := \exp(-H(x)) dx$ cannot be normalised into a probability measure since $Z_\mu := \mu(\mathbb{R}^{n+1}) = +\infty$. Nevertheless, suppose that there exists a constant $\rho > 0$ such that $V''_{\{i,j\}}(u) \geq \rho$ for any $u \in \mathbb{R}$ and any $i, j \in \{1, \dots, n\}$. Then, $u \in \mathbb{R} \mapsto V_{\{i,j\}}(u) - \rho u^2/2$ is convex and the latter implies that

$$\text{Hess}(H)(x) \geq \rho \mathbf{I}_{n+1} - (n+1)^{-1} \rho \mathbf{1}_{n+1},$$

as quadratic forms. Thus, by writing $\mathbb{R}^{n+1} = \mathbb{R} \mathbf{1}_{n+1} \oplus^\perp \mathcal{H}_n$ where $\mathcal{H}_n$ is the hyperplane of equation $h_1 + \cdots + h_{n+1} = 0$, we get that the spectrum of $\text{Hess}(H)(x)$ is of the form

$$\{0 = \lambda_1(x) < \lambda_2(x) \leq \cdots \leq \lambda_{n+1}(x)\}$$

with $\lambda_2(x) \geq n(n+1)^{-1}\rho$. Hence, one can define the probability measure σ_M on $\mathbb{R}^n$ as in (4.4) for any M in $\mathbb{R}$. Moreover σ_M is uniformly log-concave with a constant $n(n+1)^{-1}\rho$ and therefore the conditional measure μ_M can be defined from σ_M as a probability measure by equation (4.7), despite the fact that μ is not a probability measure on $\mathbb{R}^{n+1}$. The particular case $V_{\{i,j\}} = V$ with V even and uniformly convex

is considered for example in [Mal01], in terms of the associated S.D.E., in order to study the granular media equation.

As we seen, when H is uniformly strictly convex with a constant $\rho > 0$, the hypotheses of Proposition 4.0.1 are full-filled and hence, inequalities (4.13) and (4.14) hold. It is quite natural to ask if (4.13) and (4.14) remains true for symmetric but non convex Hamiltonians H . In this direction, the Bakry-Émery criterion allows the following perturbative statement due to Ivan Gentil. The proof, prototype of which can be found in [Led01], is taken from [BH99] and is postponed to section 4.1.

Proposition 4.0.2 (Perturbative result). *Let $H(x) = V(x_1) + \cdots + V(x_{n+1})$ with*

$$V(u) = \frac{u^2}{2} + F(u)$$

where $F : \mathbb{R} \rightarrow \mathbb{R}$, and let σ_M be the probability measure on $\mathbb{R}^n$ defined by (4.4), namely

$$\sigma_M(dx_1, \dots, dx_n) = (Z_{\sigma_M})^{-1} \int_{\mathbb{R}^n} \exp \left(- \sum_{i=1}^n V(x_i) - V(M - \sum_{i=1}^n x_i) \right) dx_1 \cdots dx_n.$$

Then, for $\|F\|_\infty$ small enough, there exists a positive constant $\mathcal{P}$ depending only on $\|F\|_\infty$ such that for any n , any M and any smooth $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$(4.15) \quad \mathbf{Var}_{\sigma_M}(f) \leq \mathcal{P} \mathbf{E}_{\sigma_M}(|\nabla f|^2).$$

Proposition 4.0.2 remains valid if we replace, in the definition of σ_M , the square function $u \mapsto u^2/2$ by a smooth convex function $u \mapsto \Phi(u)$, provided that there exists real constants α and β such that $0 < \alpha \leq \beta \leq 2\alpha$ and $\alpha \leq \Phi''(u) \leq \beta$ for every $u \in \mathbb{R}$. The constant $\mathcal{P}$ becomes in this case $e^{2 \operatorname{osc}(F)} / (2\alpha e^{-2 \operatorname{osc}(F)} - \beta)$ for $\operatorname{osc}(F) < \log \sqrt{2\alpha/\beta}$.

The exchangeability of the underlying measure μ_M indicates that the perturbative approach by mean of Helffer's method which sees σ_M as a quasi-product measure with small interactions is not relevant : any reduction of F in the interaction term

$$V \left(M - \sum_{i=1}^n x_i \right)$$

affects the product term $\sum_{i=1}^n V(x_i)$. Helffer's method was essentially developed for spins systems with *boundary conditions* for which the measure is not exchangeable. For our measure σ_M , one can expect in contrast that the symmetries of H_M induces a stronger result, as for many mean field models. In this direction, Landim, Panizo and Yau have recently established in [LPY00] that μ_M satisfies inequalities (4.13) and (4.14) when H is of the form $H(x) = V(x_1) + \cdots + V(x_{n+1})$ where $V(u) = u^2/2 + F(u)$ with F and F' bounded and Lipschitz. A simple example is given by

$F(x) = P(\sin(Q(x)))$ where P and Q are fixed polynomials in $\mathbb{R}[X]$. Their proof relies on Lu-Yau's Markovian decomposition [LY93] and on Local Central Limit Theorem estimates [KL99].

Following closely [LPY00], we are actually able to show that measure σ_M itself satisfies to (4.1) and (4.2) with a constants which does not depend on n and M , as stated in our main result, which follows.

Theorem 4.0.3. *Let $H(x) = V(x_1) + \cdots + V(x_{n+1})$ with $V(u) = u^2/2 + F(u)$ and let σ_M be the probability measure on $\mathbb{R}^n$ defined by (4.4), namely*

$$\sigma_M(dx_1, \dots, dx_n) = (Z_{\sigma_M})^{-1} \int_{\mathbb{R}^n} \exp \left(- \sum_{i=1}^n V(x_i) - V(M - \sum_{i=1}^n x_i) \right) dx_1 \cdots dx_n.$$

Then, if F is bounded and Lipschitz, there exists a positive constant $\mathcal{P}$ depending only on $\|F\|_\infty$ and $\|F'\|_\infty$ such that for any n and M and any smooth $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$(4.16) \quad \text{Var}_{\sigma_M}(f) \leq \mathcal{P} \mathbf{E}_{\sigma_M}(|\nabla f|^2).$$

Moreover, if F'' is also bounded, there exists a positive constant $\mathcal{L}$ depending only on $\|F\|_\infty$, $\|F'\|_\infty$ and $\|F''\|_\infty$ such that for any n and M and any smooth $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$(4.17) \quad \text{Ent}_{\sigma_M}(f^2) \leq \mathcal{L} \mathbf{E}_{\sigma_M}(|\nabla f|^2).$$

As a Corollary, we recover from Proposition 4.0.1 and (4.12) the L^2 factor for the Kawasaki dynamics (cf. (4.13) and (4.14)) obtained by [LPY00].

The rest of the paper is divided as follows. The first section gives the proof of Proposition 4.0.2, which relies only on the Bakry-Émery criterion. In Section 4.2, we give some preliminaries to the proof of Theorem 4.0.3. Lemma 4.2.1 gives some covariance bounds taken from [LPY00]. This Lemma allows us to derive the “one spin Lemma” 4.2.2 by a simple application of the Bakry-Émery criterion. Section 4.3 is devoted to the derivation of the Poincaré inequality (4.16) and section 4.4 to the derivation of the logarithmic Sobolev counterpart (4.16). The proofs make heavy use of the LCLT based estimates of [LPY00] throughout Lemmas 4.3.1 and 4.4.1, but our induction in n is quite different.

It is quite natural to ask if Theorem 4.0.3 remains valid if we replace the quadratic potential $u^2/2$ by a uniformly strictly convex potential Φ . We believe that this is true, at least for Φ such that $0 < \alpha \leq \Phi''(u) \leq \beta$. In an other direction, one can ask if our method remains valid for discrete spins systems similar to those presented in [Mar99]. It is not clear at all for us.

4.1 Proof of Proposition 4.0.2

We give here a proof of Proposition 4.0.2, which relies only on the Bakry-Émery criterion. Let σ_M^* the probability measure on $\mathbb{R}^n$ defined by

$$(Z_{\sigma_M^*})^{-1} \exp \left(- \sum_{i=1}^n V(x_i) - \frac{1}{2} \left(M - \sum_{i=1}^n x_i \right)^2 \right) dx_1 \cdots dx_n.$$

If σ_M^* satisfies a Poincaré inequality with a constant $c > 0$, then σ_M satisfies a Poincaré inequality with a constant $c \exp(2 \operatorname{osc}(F))$. Now, for any smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\begin{aligned} \mathbf{E}_{\sigma_M^*}((\mathbf{L}f)^2) &= \sum_{i,j=1}^n \mathbf{E}_{\sigma_M^*}(|\partial_{ij}^2 f|^2) \\ &\quad + \mathbf{E}_{\sigma_M^*} \left(\sum_{i=1}^n (1 + F''(x_i)) |\partial_i f|^2 \right) \\ &\quad + \mathbf{E}_{\sigma_M^*} \left(\left(\sum_{i=1}^n \partial_i f \right)^2 \right). \end{aligned}$$

In another hand, for any $i \in \{1, \dots, n\}$ and any $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, the Bakry-Émery criterion gives that the one dimensional probability measure

$$\rho_i(dx_i) := (Z_{\rho_i})^{-1} \exp \left(-V(x_i) - \frac{1}{2} \left(M - \sum_{i=1}^n x_i \right)^2 \right) dx_i$$

satisfies a Poincaré inequality with a constant $(1/2) \exp(2 \operatorname{osc}(F))$, hence, by the Bakry-Émery criterion applied reversely, we get for any smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, by summing over i

$$\sum_{i=1}^n \mathbf{E}_{\rho_i}(|\partial_{ii}^2 f|^2) + \sum_{i=1}^n \mathbf{E}_{\rho_i}((2 + F''(x_i)) |\partial_i f|^2) \geq 2 e^{-2 \operatorname{osc}(F)} \sum_{i=1}^n \mathbf{E}_{\rho_i}(|\partial_i f|^2).$$

Notice that $\rho_i = \operatorname{Law}_{\sigma_M^*}(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Therefore, by taking the expectation with respect to σ_M^* , we get

$$\begin{aligned} \mathbf{E}_{\sigma_M^*}((\mathbf{L}f)^2) &\geq (2 e^{-2 \operatorname{osc}(F)} - 1) \sum_{i=1}^n \mathbf{E}_{\sigma_M^*}(|\partial_i f|^2) \\ &=: (2 e^{-2 \operatorname{osc}(F)} - 1) \mathbf{E}_{\sigma_M^*}(|\nabla f|^2). \end{aligned}$$

Thus, for $\operatorname{osc}(F)$ sufficiently small ($< \log \sqrt{2}$), one can take

$$\mathcal{P} = \frac{e^{2 \operatorname{osc}(F)}}{2 e^{-2 \operatorname{osc}(F)} - 1},$$

which is optimal when $F \equiv 0$ (pure Gaussian case).

4.2 Preliminaries to the proof of Theorem 4.0.3

Let $\gamma_{n,M}$ the Gaussian measure on $\mathbb{R}^n$ of mean $M/(n+1)$ and covariance matrix

$$(\mathbf{I}_n + \mathbf{1}_n)^{-1} = \left(\frac{n}{n+1} \delta_{i=j} - \frac{1}{n+1} \delta_{i \neq j} \right)_{1 \leq i, j \leq n}.$$

If $B(x) := \sum_{i=1}^n F(x_i) + F\left(M - x_1 - \cdots - x_n\right)$, one can write

$$d\sigma_M(x_1, \dots, x_n) = (Z_{n,M})^{-1} \exp(-B(x)) d\gamma_{n,M}(x_1, \dots, x_n).$$

Thus, σ_M is a bounded perturbation of $\gamma_{n,M}$, which is log-concave with a constant ρ equal to 1, and therefore, σ_M satisfies to Poincaré and logarithmic Sobolev inequalities with constants depending only on $\|B\|_\infty$ (i.e. $\|F\|_\infty$ and n). Our goal is to show that the dependence in n can be dropped by taking into account $\|F'\|_\infty$ and $\|F''\|_\infty$. The presence of the bounded part F in V and the non-product nature of σ_M does not allow any direct approach based on the Bakry-Émery criterion.

Observe that $\mathbf{Cov}_{\gamma_{n,M}}(x_1, x_2) = -(n+1)^{-1}$, and we can then expect the same decrease in n for $\mathbf{Cov}_{\sigma_M}(V'(x_1), V'(x_2))$. This is actually the case, as stated in the following Lemma. Notice that since σ_M is exchangeable and since $M - \sum_{i=1}^n x_i$ and x_i have the same law under σ_M , we have $\mathbf{Var}_{\sigma_M}(x_1) = -n\mathbf{Cov}_{\sigma_M}(x_1, x_2)$, as for $\gamma_{n,M}$.

Lemma 4.2.1. *Let σ_M be the probability measure on $\mathbb{R}^n$ ($n > 2$) defined in Theorem 4.0.3 and μ_M the associated conditional measure defined by (4.6). Assume that F and F' are bounded, then there exists a constant $C > 0$ depending only on $\|F\|_\infty$ and $\|F'\|_\infty$ such that for any $M \in \mathbb{R}$*

$$(4.18) \quad |\mathbf{Cov}_{\sigma_M}(V'(x_1), V'(x_2))| = |\mathbf{Cov}_{\mu_M}(V'(x_1), V'(x_2))| \leq \frac{C}{n},$$

and

$$(4.19) \quad \begin{aligned} \mathbf{Var}_{\sigma_M}\left(\sum_{i=1}^n V'(x_i) + V'(M - \sum_{i=1}^n x_i)\right) &= \mathbf{Var}_{\sigma_M}\left(\sum_{i=1}^n F'(x_i) + F'(M - \sum_{i=1}^n x_i)\right) \\ &= \mathbf{Var}_{\mu_M}\left(\sum_{i=1}^{n+1} F'(x_i)\right) \leq nC. \end{aligned}$$

Proof. Inequality (4.19) follows from (4.7) and [LPY00, Corollary 5.4]. For (4.18), just write

$$\begin{aligned} \mathbf{Cov}_{\sigma_M}(V'(x_1), V'(x_2)) &= \mathbf{Cov}_{\sigma_M}(x_1, x_2) \\ &\quad + 2\mathbf{Cov}_{\sigma_M}(x_1, F'(x_2)) \\ &\quad + \mathbf{Cov}_{\sigma_M}(F'(x_1), F'(x_2)), \end{aligned}$$

and use (4.7) and [LPY00, Corollary 5.3] to estimate each term. Actually, one can derive the estimates of $\mathbf{Cov}_{\sigma_M}(x_1, F'(x_2))$ and $\mathbf{Cov}_{\sigma_M}(x_1, x_2)$ directly by using the symmetries of σ_M . $\square$

Inequality (4.18) of Lemma 4.2.1 allows us to establish the following one spin result, which is the first step in our proof of Poincaré and logarithmic Sobolev inequalities for σ_M by induction on n by mean of the Lu-Yau Markovian decomposition. In the other hand, inequality (4.19) will be useful, as we will see in sections 4.3 and 4.4, for the induction itself.

Lemma 4.2.2 (One spin Lemma). *Let σ_M be the probability measure on $\mathbb{R}^n$ defined in Theorem 4.0.3. If F is bounded and Lipschitz, there exists a constant $A > 0$ depending only on $\|F\|_\infty$ and $\|F'\|_\infty$ and not on n and M such that for any n and M and any smooth $f : \mathbb{R} \rightarrow \mathbb{R}$,*

$$\mathbf{Ent}_{\sigma_M}(f(x_1)^2) \leq 2A \mathbf{E}_{\sigma_M}(f'(x_1)^2),$$

and

$$\mathbf{Var}_{\sigma_M}(f(x_1)) \leq A \mathbf{E}_{\sigma_M}(f'(x_1)^2).$$

Proof of Lemma 4.2.2. As we already noticed, it is clear that the desired inequalities are true with a constant depending on n and $\|F\|_\infty$, so we just have to see what happens for large values of n . We have in mind the use of the Bakry-Émery criterion. The Hamiltonian of the probability measure in x_1 is given by

$$\begin{aligned} \varphi_{M,n}(x_1) &:= V(x_1) + \log Z_{M,n} \\ &\quad - \log \int \exp \left(- \sum_{i=2}^n V(x_i) - V(M - \sum_{i=1}^n x_i) \right) dx_2 \cdots dx_n. \end{aligned}$$

We first observe that we can forget the $F(x_1)$ part in $V(x_1)$, which is paid by a factor $\exp(2 \operatorname{osc}(F))$ in A . Hence, we simply have, after an integration by parts

$$\varphi_{M,n}''(x_1) = 1 - \mathbf{Cov}_{\sigma_{M-x_1}(dx_2, \dots, dx_n)}(V'(x_2), V'(x_3)).$$

Now, (4.18) gives $\varphi_{M,n}''(x_1) \geq 1 - Cn^{-1}$, where C is a positive constant depending only on $\|F\|_\infty$ and $\|F'\|_\infty$ and not on n and M . Thus, we are able to apply the Bakry-Émery criterion for large values of n . Hence, the proof is completed, with a constant A depending only on $\|F\|_\infty$ and $\|F'\|_\infty$ and not on M and n . $\square$

Obviously, one can replace x_1 in f and f' by $M - x_1 - \cdots - x_n$ or by any x_i for $i \in \{1, \dots, n\}$. Moreover, according to (4.8), one can replace $\mathbf{E}_{\sigma_M}$ by $\mathbf{E}_{\mu_M}$.

4.3 Derivation of the Poincaré inequality

This section is devoted to the derivation of inequality (4.16) of Theorem 4.0.3. The proof relies on the one spin Lemma 4.2.2 and on the crucial Lemma 4.3.1 which allows us to use the Lu-Yau Markovian decomposition.

Proof of (4.16). As we already noticed, the result is true with a constant depending on n , so that if we denote by $\mathcal{P}_n$ the maximum of best Poincaré constants in dimension less than or equal to n , we just have to show that the non decreasing sequence of constants $(\mathcal{P}_n)_{n \geq 1}$ is bounded.

Let us denote by σ the measure σ_M and by $\sigma^{(k)}$ the measure σ_M given $x_1, \dots, x_k$ for $k \in \{0, \dots, n\}$ and by f_k the conditional expectation

$$\mathbf{E}_\sigma(f|x_1, \dots, x_k) = \mathbf{E}_{\sigma^{(k)}}(f).$$

Notice that $\sigma^{(k)}$ is nothing else but $\sigma_{M-x_1-\dots-x_k}(dx_{k+1}, \dots, dx_n)$. Moreover, $f_n = f$ and by convention $\sigma^{(0)} := \sigma$ and thus $f_0 = \mathbf{E}_\mu(f)$. For a *fixed* function f , we can always choose the order of the coordinates $x_1, \dots, x_n$ such that $\mathbf{E}_\sigma(|\partial_k f|^2)$ becomes a *non increasing* sequence in $k \in \{1, \dots, n\}$. This gives

$$\sum_{i=k+1}^n \frac{1}{n-k} \mathbf{E}_\sigma(|\partial_i f|^2) \leq \mathbf{E}_\sigma(|\partial_{k+1} f|^2).$$

Following Lu-Yau [LY93], we have the following Markovian decomposition of the variance

$$\mathbf{Var}_\sigma(f) := \mathbf{E}_\sigma(f^2) - \mathbf{E}_\sigma(f)^2 = \sum_{k=1}^n \mathbf{E}_\sigma((f_k)^2 - (f_{k-1})^2) = \sum_{k=1}^n \mathbf{E}_\sigma(\mathbf{Var}_{\sigma^{(k-1)}}(f_k)).$$

Since measure $\sigma^{(k-1)}$ integrates coordinates $x_k, \dots, x_n$ and function f_k depends only on coordinates $x_1, \dots, x_k$, the quantity $\mathbf{Var}_{\sigma^{(k-1)}}(f_k)$ is actually a variance for a one spin function. Therefore, by the one spin Lemma 4.2.2, there exists a constant $A > 0$ depending on $\|F\|_\infty$ and $\|F'\|_\infty$ but not on n and M such that

$$\mathbf{Var}_\sigma(f) \leq A \sum_{k=1}^n \mathbf{E}_\sigma(|\partial_k f_k|^2).$$

Our aim is to express the right hand side of the previous inequality in terms of $|\partial_k f|^2$. Notice that the $k = n$ term in the sum is trivial since $f_n = f$. By definition of f_k , we get for any $k \in \{1, \dots, n-1\}$

$$\partial_k f_k = \mathbf{E}_{\sigma^{(k)}}(\partial_k f) - \mathbf{Cov}_{\sigma^{(k)}}\left(f, V'\left(M - \sum_{i=1}^n x_i\right)\right).$$

At this stage, we notice that by $n-k$ integrations by parts, we have

$$\begin{aligned} \mathbf{Cov}_{\sigma^{(k)}}\left(f, V'\left(M - \sum_{i=1}^n x_i\right)\right) &= \frac{1}{n-k} \sum_{i=k+1}^n \mathbf{Cov}_{\sigma^{(k)}}(f, V'(x_i)) \\ &\quad - \frac{1}{n-k} \sum_{i=k+1}^n \mathbf{E}_{\sigma^{(k)}}(\partial_i f). \end{aligned}$$

Therefore, we can write by denoting $S_k := \sum_{i=k+1}^n V'(x_i) + V'(M - \sum_{i=1}^n x_i)$

$$\partial_k f_k = \mathbf{E}_{\sigma^{(k)}}(\partial_k f) - \frac{1}{n-k+1} \mathbf{Cov}_{\sigma^{(k)}}(f, S_k) + \frac{1}{n-k+1} \sum_{i=k+1}^n \mathbf{E}_{\sigma^{(k)}}(\partial_i f).$$

Now, by the Cauchy-Schwarz inequality

$$|\partial_k f_k|^2 \leq 3 \mathbf{E}_{\sigma^{(k)}}(|\partial_k f|^2) + \frac{3}{(n-k)^2} \mathbf{Cov}_{\sigma^{(k)}}(f, S_k)^2 + \frac{3}{n-k} \sum_{i=k+1}^n \mathbf{E}_{\sigma^{(k)}}(|\partial_i f|^2).$$

This gives by summing over all k in $\{1, \dots, n-1\}$ (the case $k = n$ is trivial)

$$\begin{aligned} \sum_{k=1}^{n-1} \mathbf{E}_{\sigma}(|\partial_k f_k|^2) &\leq 3 \mathbf{E}_{\sigma}(|\nabla f|^2) + 3 \sum_{k=1}^{n-1} \frac{1}{(n-k)^2} \mathbf{E}_{\sigma}(\mathbf{Cov}_{\sigma^{(k)}}(f, S_k)^2) \\ &\quad + 3 \sum_{k=1}^{n-1} \frac{1}{n-k} \sum_{i=k+1}^n \mathbf{E}_{\sigma}(|\partial_i f|^2). \end{aligned}$$

The monotonicity of $\mathbf{E}_{\sigma}(|\partial_i f|^2)$ yields

$$\sum_{k=1}^{n-1} \mathbf{E}_{\sigma}(|\partial_k f_k|^2) \leq 6 \mathbf{E}_{\sigma}(|\nabla f|^2) + 3 \sum_{k=1}^{n-1} \frac{1}{(n-k)^2} \mathbf{E}_{\sigma}(\mathbf{Cov}_{\sigma^{(k)}}(f, S_k)^2).$$

By inequality (4.21) of Lemma 4.3.1, there exists a positive constant C depending only on $\|F\|_{\infty}$ and $\|F'\|_{\infty}$ such that for any $\varepsilon > 0$, there exists a positive constant C_{ε} depending only on $\|F\|_{\infty}$, $\|F'\|_{\infty}$ and ε such that for any $k \in \{1, \dots, n-1\}$

$$\mathbf{Cov}_{\sigma^{(k)}}(f, S_k)^2 \leq (C_{\varepsilon} + \varepsilon(n-k)C) \mathbf{Var}_{\sigma^{(k)}}(f) + (n-k)C_{\varepsilon} \sum_{i=k+1}^n \mathbf{E}_{\sigma^{(k)}}(|\partial_i f|^2).$$

Therefore, by the monotonicity of $\mathbf{E}_{\sigma}(|\partial_i f|^2)$ again

$$\begin{aligned} \sum_{k=1}^{n-1} \mathbf{E}_{\sigma}(|\partial_k f_k|^2) &\leq C'_{\varepsilon} \mathbf{E}_{\sigma}(|\nabla f|^2) + C'_{\varepsilon} \sum_{k=1}^{n-1} \frac{1}{(n-k)^2} \mathbf{E}_{\sigma}(\mathbf{Var}_{\sigma^{(k)}}(f)) \\ (4.20) \quad &\quad + \varepsilon C' \sum_{k=1}^{n-1} \frac{1}{n-k} \mathbf{E}_{\sigma}(\mathbf{Var}_{\sigma^{(k)}}(f)). \end{aligned}$$

Recall that $\mathcal{P}_n$ is the maximum of best Poincaré constants in dimension less than or equal to n . The last sum of the right hand side (RHS) of (4.20) can be bounded above as follows

$$\sum_{k=1}^{n-1} \frac{1}{n-k} \mathbf{E}_{\sigma}(\mathbf{Var}_{\sigma^{(k)}}(f)) \leq \mathcal{P}_{n-1} \mathbf{E}_{\sigma}(|\nabla f|^2).$$

It remains to examine the first sum of the RHS of (4.20). The Jensen inequality yields

$$\mathbf{E}_\sigma(\mathbf{Var}_{\sigma^{(k)}}(f)) \leq \mathbf{Var}_\sigma(f),$$

and therefore, we get for any $p \in \{1, \dots, n-1\}$

$$\begin{aligned} \sum_{k=1}^{n-1} \frac{1}{(n-k)^2} \mathbf{E}_\sigma(\mathbf{Var}_{\sigma^{(k)}}(f)) &= \mathbf{Var}_\sigma(f) \sum_{k=1}^{n-p-1} \frac{1}{(n-k)^2} + \sum_{k=n-p}^{n-1} \frac{\mathbf{E}_\sigma(\mathbf{Var}_{\sigma^{(k)}}(f))}{(n-k)^2} \\ &\leq \mathbf{Var}_\sigma(f) \sum_{k=p+1}^{n-1} \frac{1}{k^2} + \sum_{k=1}^p \frac{1}{k^2} \mathbf{E}_\sigma(\mathbf{Var}_{\sigma^{(n-k)}}(f)). \end{aligned}$$

At this stage, we observe that for every k in $\{1, \dots, p\}$,

$$\mathbf{E}_\sigma(\mathbf{Var}_{\sigma^{(n-k)}}(f)) \leq \mathcal{P}_p \sum_{i=n-k+1}^n \mathbf{E}_\sigma(|\partial_i f|^2) \leq p \mathcal{P}_p \mathbf{E}_\sigma(|\partial_{n-p+1} f|^2).$$

We are now able to collect our estimates of the RHS of (4.20). Putting all together, we have obtained that

$$\sum_{k=1}^{n-1} \mathbf{E}_\sigma(|\partial_k f_k|^2) \leq (C'_\varepsilon + p\pi^2 \mathcal{P}_p C'_\varepsilon + \varepsilon C'' \mathcal{P}_{n-1}) \mathbf{E}_\sigma(|\nabla f|^2) + (C'_\varepsilon R_p) \mathbf{Var}_\sigma(f),$$

where $R_p := \sum_{k=p+1}^{n-1} k^{-2}$. Therefore, for some $C''_{p,\varepsilon} > 0$,

$$(1 - AC'_\varepsilon R_p) \mathbf{Var}_\sigma(f) \leq (C''_{p,\varepsilon} + \varepsilon AC'' \mathcal{P}_{n-1}) \mathbf{E}_\sigma(|\nabla f|^2).$$

Now, we may choose $\varepsilon < 1/(AC'')$ and then p large enough (always possible when n is sufficiently large) to ensure that

$$R_p < \min \left(\frac{1}{AC'_\varepsilon}, \frac{1 - \varepsilon AC''}{AC'_\varepsilon} \right).$$

This gives two positive constants α and β with $\beta < 1$ depending only on $\|F\|_\infty$ and $\|F'\|_\infty$ such that for large values of n , one has $\mathcal{P}_n \leq \alpha + \beta \mathcal{P}_{n-1}$, and therefore $\sup_n \mathcal{P}_n < +\infty$. $\square$

Let us give now the crucial Lemma which allows us to use the Markovian decomposition of Lu-Yau, by splitting the covariance term into a variance term and a gradient term. The proof makes heavy use of estimates taken from [LPY00].

Lemma 4.3.1. *Let σ_M be the probability measure on $\mathbb{R}^n$ defined in Theorem 4.0.3. Assume that F is bounded and Lipschitz, then there exists a positive constant C depending only on $\|F\|_\infty$ and $\|F'\|_\infty$ such that for any $\varepsilon > 0$, there exists a positive*

constant C_ε depending only on $\|F\|_\infty$, $\|F'\|_\infty$ and ε such that for any $n \in \mathbb{N}^*$, any $M \in \mathbb{R}$ and any smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$

$$(4.21) \quad \mathbf{Cov}_{\sigma_M}(f, S)^2 \leq (C_\varepsilon + \varepsilon n C) \mathbf{Var}_{\sigma_M}(f) + n C_\varepsilon \mathbf{E}_{\sigma_M}(|\nabla f|^2),$$

where $S := \sum_{i=1}^n V'(x_i) + V'(M - \sum_{i=1}^n x_i)$.

Proof of Lemma 4.3.1. Notice that we just have to study what happens for small values of ε and large values of n , since for any $\varepsilon > 0$ and any $n \leq n_\varepsilon$, we get by the Cauchy-Schwarz inequality and (4.19) that

$$\mathbf{Cov}_{\sigma_M}(f, S)^2 \leq n_\varepsilon C \mathbf{Var}_{\sigma_M}(f) =: C_\varepsilon \mathbf{Var}_{\sigma_M}(f).$$

We have in mind the use of the partitioning result of [LPY00]. If μ_M denotes the conditional measure on $\mathbb{R}^{n+1}$ associated to σ_M as in (4.7), we have

$$\mathbf{Cov}_{\sigma_M}(f, S(x_1, \dots, x_n))^2 = \mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{n+1} F'(x_i)\right)^2.$$

Now, for n large enough, one can then subdivide the set $\{1, \dots, n+1\}$ into ℓ adjacent subsets I_i of size K or $K+1$. We have in mind to take $K^{-1} \leq \varepsilon$, which is always possible when n is large enough. We can write with this decomposition

$$\mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{n+1} F'(x_i)\right) = \mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{\ell} \sum_{k \in I_i} F'(x_k)\right).$$

For any $(i, x) \in \{1, \dots, \ell\} \times \mathbb{R}^{n+1}$, we define the “total spin on I_i ” by $M_i(x) := \sum_{k \in I_i} x_k$. On $\mathbb{R}^{I_i}$, one can define the conditional measure μ_{M_i} with total spin M_i , as in (4.6). To lighten the notations, we denote this measure by $\mu_{(i)}$. We get from the latter by the Cauchy-Schwarz inequality

$$(4.22) \quad \begin{aligned} \mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{n+1} F'(x_i)\right)^2 &\leq 2 \mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{\ell} \sum_{k \in I_i} \left(F'(x_k) - \mathbf{E}_{\mu_{(i)}}(F')\right)\right)^2 \\ &\quad + 2 \mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{\ell} |I_i| \mathbf{E}_{\mu_{(i)}}(F')\right)^2 \end{aligned}$$

By the Cauchy-Schwarz inequality again and by (4.8), the second term of the RHS of (4.22) can be bounded above by

$$\mathbf{Var}_{\sigma_M}(f) \mathbf{Var}_{\mu_M}\left(\sum_{i=1}^{\ell} |I_i| \mathbf{E}_{\mu_{(i)}}(F')\right).$$

Now, according to [LPY00, ineq. (3.10)], the last variance in the RHS is bounded above by nC/K for n sufficiently large, which can be rewritten as εnC . We turn now to the control of the first term of the RHS of (4.22). Since $\mathbf{E}_{\mu_M} = \mathbf{E}_{\mu_M} \circ \mathbf{E}_{\mu_{(i)}}$, we get

$$\mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{\ell} \sum_{k \in I_i} \left(F'(x_k) - \mathbf{E}_{\mu_{(i)}}(F')\right)\right) = \sum_{i=1}^{\ell} \mathbf{E}_{\mu_M}\left(\mathbf{Cov}_{\mu_{(i)}}\left(f, \sum_{k \in I_i} F'(x_k)\right)\right).$$

Thus, the Cauchy-Schwarz inequality yields

$$\mathbf{Cov}_{\mu_M}\left(f, \sum_{i=1}^{\ell} \sum_{k \in I_i} \left(F'(x_k) - \mathbf{E}_{\mu_{(i)}}(F')\right)\right)^2 \leq \ell \sum_{i=1}^{\ell} \mathbf{E}_{\mu_M}\left(\mathbf{Cov}_{\mu_{(i)}}\left(f, \sum_{k \in I_i} F'(x_k)\right)^2\right).$$

Again by the Cauchy-Schwarz inequality, we get

$$\mathbf{Cov}_{\mu_{(i)}}\left(f, \sum_{k \in I_i} F'(x_k)\right)^2 \leq \mathbf{Var}_{\mu_{(i)}}(f) \mathbf{Var}_{\mu_{(i)}}\left(\sum_{k \in I_i} F'(x_k)\right).$$

By virtue of (4.19) applied to $\mu_{(i)}$, we obtain

$$\mathbf{Cov}_{\mu_{(i)}}\left(f, \sum_{k \in I_i} F'(x_k)\right)^2 \leq C |I_i| \mathbf{Var}_{\mu_{(i)}}(f).$$

Now, for any i , let $r_i = \max\{k, k \in I_i\}$ and $J_i := I_i \setminus \{r_i\}$ and $\sigma_{(i)}$ the probability measure on $\mathbb{R}^{J_i}$ associated with the Hamiltonian

$$\sum_{k \in J_i} V(x_k) + V(M_i - \sum_{k \in J_i} x_k).$$

Equation (4.8) simply gives

$$\mathbf{Var}_{\mu_{(i)}}(f) = \mathbf{Var}_{\sigma_{(i)}}(f(\varphi_i(x))),$$

where $\varphi_i : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ is defined by

$$(\varphi_i(x))_k := \begin{cases} x_k & \text{if } k \neq r_i \\ M_i - \sum_{l \in J_i} x_l & \text{if } k = r_i \end{cases}$$

Recall that $\mathcal{P}_K$ is the maximum of the best Poincaré constants for σ_M in dimensions less than or equal to K . We get by definition of $\mathcal{P}_K$ that

$$\begin{aligned} \mathbf{Var}_{\sigma_{(i)}}(f) &\leq \mathcal{P}_K \sum_{k \in J_i} \mathbf{E}_{\sigma_{(i)}}(|(\partial_k f)(\varphi_i) - (\partial_{r_i} f)(\varphi_i)|^2) \\ &= \mathcal{P}_K \sum_{k \in J_i} \mathbf{E}_{\mu_{(i)}}(|\partial_k f - \partial_{r_i} f|^2). \end{aligned}$$

Hence, by the Cauchy-Schwarz inequality, we get

$$\mathbf{Var}_{\sigma_{(i)}}(f) \leq 2\mathcal{P}_K \mathbf{E}_{\mu_{(i)}} \left(\sum_{k \in J_i} |\partial_k f|^2 \right) + 2(|I_i| - 1) \mathcal{P}_K \mathbf{E}_{\mu_{(i)}} (|\partial_{r_i} f|^2),$$

Summarising, since $\mathcal{P}_K$ depends only on K , $\|F\|_\infty$, $\|F'\|_\infty$, we have obtained that the first term of the right hand side of (4.22) is bounded above by

$$nC_K \mathcal{P}_K \mathbf{E}_{\mu_M} (|\nabla f|^2),$$

which can be rewritten by virtue of (4.8) as $nC'_\varepsilon \mathbf{E}_{\sigma_M} (|\nabla f|^2)$. This concludes the proof of (4.21) and Lemma 4.3.1. $\square$

4.4 Derivation of the logarithmic Sobolev inequality

This section is devoted to the derivation of inequality (4.17) of Theorem 4.0.3. As for the Poincaré inequality (4.16), the proof relies on the one spin Lemma 4.2.2 and on a crucial Lemma 4.4.1 which allows us to use the Lu-Yau Markovian decomposition.

Proof of the logarithmic Sobolev inequality (4.17) of Theorem 4.0.3. We follow here the same scheme used for the Poincaré inequality. For any smooth non negative function $g : \mathbb{R}^n \rightarrow \mathbb{R}^+$, we have the following decomposition of the entropy

$$\begin{aligned} \mathbf{Ent}_\sigma(g) &:= \mathbf{E}_\sigma(g \log g) - \mathbf{E}_\sigma(g) \log \mathbf{E}_\sigma(g) \\ &= \sum_{k=1}^n \mathbf{E}_\sigma(g_k \log g_k - g_{k-1} \log g_{k-1}) \\ &= \sum_{k=1}^n \mathbf{E}_\sigma(\mathbf{Ent}_{\sigma^{(k-1)}}(g_k)). \end{aligned}$$

Alike for the variance, measure $\sigma^{(k-1)}$ integrates on $x_k, \dots, x_n$ and function f_k depends only on $x_1, \dots, x_k$, so that $\mathbf{Ent}_{\sigma^{(k-1)}}(g_k)$ is actually an entropy for a one spin function. Therefore, by the one spin Lemma 4.2.2, there exists a positive constant A depending on $\|F\|_\infty$ and $\|F'\|_\infty$ but not on n and M such that

$$\mathbf{Ent}_\sigma(g) \leq 2A \sum_{k=1}^n \mathbf{E}_\sigma \left(\frac{|\partial_k g_k|^2}{4g_k} \right),$$

By taking $g = f^2$ for a smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, we get

$$\mathbf{Ent}_\sigma(f^2) \leq 2A \sum_{k=1}^n \mathbf{E}_\sigma \left(\frac{|\partial_k (f^2)_k|^2}{4(f^2)_k} \right).$$

By imitating the method used for the Poincaré inequality, we get that

$$\begin{aligned} \frac{|\partial_k(f^2)_k|^2}{4(f^2)_k} &\leq 3 \frac{|\mathbf{E}_{\sigma^{(k)}}(f\partial_k f)|^2}{\mathbf{E}_{\sigma^{(k)}}(f^2)} \\ &\quad + \frac{6}{(n-k)^2} \frac{\mathbf{Cov}_{\sigma^{(k)}}(f^2, S_k)^2}{\mathbf{E}_{\sigma^{(k)}}(f^2)} \\ &\quad + \frac{3}{n-k} \sum_{i=k+1}^n \frac{|\mathbf{E}_{\sigma^{(k)}}(f\partial_i f)|^2}{\mathbf{E}_{\sigma^{(k)}}(f^2)}. \end{aligned}$$

The Cauchy-Schwarz inequality yields

$$\frac{|\mathbf{E}_{\sigma^{(k)}}(f\partial_k f)|^2}{\mathbf{E}_{\sigma^{(k)}}(f^2)} \leq \mathbf{E}_{\sigma^{(k)}}(|\partial_k f|^2).$$

Therefore, the Jensen inequality and the monotonicity of $\mathbf{E}_{\sigma}(|\partial_i f|^2)$ yield

$$\sum_{k=1}^{n-1} \mathbf{E}_{\sigma} \left(\frac{|\partial_k(f^2)_k|^2}{4(f^2)_k} \right) \leq 6 \mathbf{E}_{\sigma}(|\nabla f|^2) + 6 \sum_{k=1}^{n-1} \frac{1}{(n-k)^2} \mathbf{E}_{\sigma} \left(\frac{\mathbf{Cov}_{\sigma^{(k)}}(f^2, S_k)^2}{\mathbf{E}_{\sigma^{(k)}}(f^2)} \right).$$

By inequality (4.23) of Lemma (4.4.1), there exists a positive constant C depending only on $\|F\|_{\infty}$, $\|F'\|_{\infty}$ and $\|F''\|_{\infty}$ such that for any $\varepsilon > 0$, there exists a positive constant C_{ε} depending only on $\|F\|_{\infty}$, $\|F'\|_{\infty}$, $\|F''\|_{\infty}$ and ε such that for any n and M

$$\frac{\mathbf{Cov}_{\sigma^{(k)}}(f^2, S_k)^2}{\mathbf{E}_{\sigma^{(k)}}(f^2)} \leq (C_{\varepsilon} + \varepsilon(n-k)C) \mathbf{Ent}_{\sigma^{(k)}}(f^2) + (n-k)C_{\varepsilon} \sum_{i=k+1}^n \mathbf{E}_{\sigma^{(k)}}(|\partial_i f|^2).$$

Hence, we are now able to proceed as the same way as for the Poincaré inequality. $\square$

As for the derivation of the Poincaré inequality, we give now the crucial Lemma which allows us to use the Markovian decomposition of Lu-Yau.

Lemma 4.4.1. *Let σ_M be the probability measure on $\mathbb{R}^n$ defined in Theorem 4.0.3. Assume that F , F' and F'' are bounded, then there exists a positive constant C depending only on $\|F\|_{\infty}$, $\|F'\|_{\infty}$ and $\|F''\|_{\infty}$ such that for any $\varepsilon > 0$, there exists a positive constant C_{ε} depending only on $\|F\|_{\infty}$, $\|F'\|_{\infty}$, $\|F''\|_{\infty}$ and ε such that for any $n \in \mathbb{N}^*$, any $M \in \mathbb{R}$ and any smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $\mathbf{E}_{\sigma_M}(f^2) = 1$*

$$(4.23) \quad \mathbf{Cov}_{\sigma_M}(f^2, S)^2 \leq (C_{\varepsilon} + \varepsilon n C) \mathbf{Ent}_{\sigma_M}(f^2) + n C_{\varepsilon} \mathbf{E}_{\sigma_M}(|\nabla f|^2),$$

where $S(x) := \sum_{i=1}^n V'(x_i) + V'(M - \sum_{i=1}^n x_i)$.

Proof of Lemma 4.4.1. We follow the same scheme as for (4.21), but this time, we have to replace sometimes the Cauchy-Schwarz inequality by the entropy inequality. Since f^2 is a density with respect to σ_M , we can write

$$\mathbf{Cov}_{\sigma_M}(f^2, S) = \mathbf{E}_{\sigma_M}((S - \mathbf{E}_{\sigma_M}(S)) f^2),$$

and hence, we get by the entropy inequality that for any $\beta > 0$

$$\mathbf{Cov}_{\sigma_M}(f^2, S) \leq \beta^{-1} \log \mathbf{E}_{\sigma_M}(\exp(\beta(S - \mathbf{E}_{\sigma_M}(S)))) + \beta^{-1} \mathbf{E}_{\sigma_M}(f^2 \log f^2).$$

By (4.7) and [LPY00, Lemma 6.1], the first term of the right hand side is bounded above by $nC\beta$ where C depends only on $\|F\|_\infty$, and $\|F''\|_\infty$. This yields by considering the minimum in $\beta > 0$

$$\mathbf{Cov}_{\sigma_M}(f^2, S)^2 \leq nC \mathbf{E}_{\sigma_M}(f^2 \log f^2).$$

Thus, for any fixed $\varepsilon > 0$, we just have to study what happens for large values of n since $nC \leq n_\varepsilon C =: C_\varepsilon$ for $n \leq n_\varepsilon$. After rewriting (4.23) in terms of μ_M , we get by Cauchy-Schwarz's inequality

$$(4.24) \quad \begin{aligned} \mathbf{Cov}_{\mu_M}\left(f^2, \sum_{i=1}^{n+1} F'(x_i)\right)^2 &\leq 2 \mathbf{Cov}_{\mu_M}\left(f^2, \sum_{i=1}^{\ell} \sum_{k \in I_i} \left(F'(x_k) - \mathbf{E}_{\mu_{(i)}}(F')\right)\right)^2 \\ &\quad + 2 \mathbf{Cov}_{\mu_M}\left(f^2, \sum_{i=1}^{\ell} |I_i| \mathbf{E}_{\mu_{(i)}}(F')\right)^2. \end{aligned}$$

Let us treat the first term of the right hand side of (4.24). It can be rewritten as

$$2 \sum_{i=1}^{\ell} \mathbf{E}_{\mu_M}\left(\mathbf{E}_{\mu_{(i)}}(f^2) \mathbf{Cov}_{\mu_{(i)}}\left(f_i^2, \sum_{k \in I_i} F'(x_k)\right)\right),$$

where $f_i^2 := f^2 / \mathbf{E}_{\mu_{(i)}}(f^2)$. Thus, by the Cauchy-Schwarz inequality, the first term of the RHS of (4.24) is bounded above by

$$2 \ell \sum_{i=1}^{\ell} \mathbf{E}_{\mu_M}\left(\mathbf{E}_{\mu_{(i)}}(f^2) \mathbf{Cov}_{\mu_{(i)}}\left(f_i^2, \sum_{k \in I_i} F'(x_k)\right)^2\right),$$

where we used the Jensen inequality with respect to the density $\mathbf{E}_{\mu_{(i)}}(f^2)$. Now, by the entropy inequality and by [LPY00, Lemma 6.1]

$$\mathbf{E}_{\mu_{(i)}}(f^2) \mathbf{Cov}_{\mu_{(i)}}\left(f_i^2, \sum_{k \in I_i} F'(x_k)\right)^2 \leq C |I_i| \mathbf{Ent}_{\mu_{(i)}}(f^2).$$

At this stage, the argument used for the Poincaré inequality can be rewritten exactly in the same way, by replacing the variance by the entropy and $\mathcal{P}_K$ by $\mathcal{L}_K$. It gives finally that the first term of the RHS of (4.24) is bounded above by

$$nC_K \mathcal{L}_K \mathbf{E}_{\mu_M}(|\nabla f|^2).$$

The latter can be rewritten by virtue of (4.8) as $nC'_\varepsilon \mathbf{E}_{\sigma_M}(|\nabla f|^2)$. It remains to bound the last term of the RHS of (4.24). Let β_0 as in [LPY00, Lemma 6.5] and $\delta \in (0, 2)$. By a simple rewriting of [LPY00, Lemma 4.5], one gets that if $\mathbf{Ent}_{\mu_M}(f^2) \leq \delta(n+1)\beta_0^2$ with n and K large enough

$$\mathbf{Cov}_{\mu_M}\left(f^2, \sum_{i=1}^{\ell} |I_i| \mathbf{E}_{\mu_{(i)}}(F')\right)^2 \leq \delta n C \mathbf{Ent}_{\mu_M}(f^2).$$

In another hand, if $\mathbf{Ent}_{\mu_M}(f^2) \geq \delta(n+1)\beta_0^2$, one gets

$$\mathbf{Cov}_{\mu_M}\left(f^2, \sum_{i=1}^{\ell} |I_i| \mathbf{E}_{\mu_{(i)}}(F')\right)^2 \leq \delta n C_K \mathbf{Ent}_{\mu_M}(f^2) + C_{K,\delta} n \mathbf{E}_{\mu_M}(|\nabla f|^2).$$

This last estimate is based on a simple rewriting of [LPY00, Lemma 4.5] together with the following straightforward but essential version of [LPY00, Lemma 4.6] :

$$\begin{aligned} \mathbf{E}_{\nu_{I_i \cup I_j, M}}((m_i - m_j)^2 f^2) &\leq C_1(K) \mathbf{E}_{\nu_{I_i \cup I_j, M}}(f^2) \\ &\quad + C_2(K) \mathcal{L}_{2K} \mathbf{E}_{\nu_{I_i \cup I_j, M}}\left(\sum_{k \in I_i \cup I_j} |\partial_k f|^2\right), \end{aligned}$$

where $\nu_{I_i \cup I_j, M}$ is the conditional measure on $I_i \cup I_j$, $m_i = |I_i|^{-1} \sum_{k \in I_i}$, and $C_1(K)$ tends to 0 when $K \rightarrow +\infty$.

Summarising, we get that for any $\delta \in (0, 2)$ and for n and K large enough, the last term of the RHS of (4.24) is bounded above as follows

$$\mathbf{Cov}_{\mu_M}\left(f^2, \sum_{i=1}^{\ell} |I_i| \mathbf{E}_{\mu_{(i)}}(F')\right)^2 \leq n\delta C_K \mathbf{Ent}_{\mu_M}(f^2) + nC_{K,\delta} \mathcal{L}_{2K} \mathbf{E}_{\mu_M}(|\nabla f|^2),$$

which can be rewritten by virtue of (4.8) as $\varepsilon n C'' \mathbf{Ent}_{\sigma_M}(f^2) + nC''_\varepsilon \mathbf{E}_{\sigma_M}(|\nabla f|^2)$. This achieves the proof of (4.23) and Lemma 4.4.1. $\square$

Acknowledgements

The author would like to warmly acknowledge Prof. Michel Ledoux for helpful discussions and encouraging comments, and Doct. Ivan Gentil for some discussions at the beginning of this work.

Bibliography

- [ABC⁺00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses, vol. 10, Société Mathématique de France, Paris, 2000. 1, 2, 6, 12, 31, 36, 39, 40, 42, 43, 45, 64, 66, 72
- [BH99] T. Bodineau and B. Helffer, *The log-Sobolev inequality for unbounded spin systems*, J. Funct. Anal. **166** (1999), no. 1, 168–178. 77
- [Cap01] P. Caputo, *A remark on spectral gap and logarithmic sobolev inequalities for conservative spin systems*, Preprint archived as mp_arc 01-71, 2001. 73, 74
- [KL99] C. Kipnis and C. Landim, *Scaling limit of interacting particle systems*, Grundlehren der Mathematischen Wissenschaften, vol. 320, Springer, Berlin, New-York, 1999. 78
- [Led01] M. Ledoux, *Logarithmic Sobolev inequalities for unbounded spin systems revisited*, Séminaire de Probabilités, XXXV, Springer, Berlin, 2001, pp. 167–194. 77
- [LPY00] C. Landim, G. Panizo, and H. T. Yau, *Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems*, Preprint, to appear in Ann. Inst. H. Poincaré, 2000. 2, 15, 16, 71, 77, 78, 80, 84, 85, 86, 89, 90
- [LY93] S. L. Lu and H.-T. Yau, *Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics*, Comm. Math. Phys. **156** (1993), no. 2, 399–433. 78, 82
- [Mal01] F. Malrieu, *Convergence to equilibrium for granular media equations and their Euler schemes*, preprint, to appear in “Annals of Applied Probability”, 2001. 77
- [Mar99] F. Martinelli, *Lectures on Glauber dynamics for discrete spin models*, Lectures on probability theory and statistics. École d’été de probabilités de St-Flour 1997, Lecture Notes in Math., vol. 1717, Springer, Berlin, 1999, pp. 93–191. 78
- [SC97] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics. École d’été de probabilités de St-Flour 1996, Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413. 75

- [SZ92] D. Stroock and B. Zegarliński, *The logarithmic Sobolev inequality for continuous spin systems on a lattice*, J. Funct. Anal. **104** (1992), no. 2, 299–326.

Bibliographie générale

- [ABC⁺00] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO et G. SCHEFFER – *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses, vol. 10, Société Mathématique de France, Paris, 2000. 1, 2, 6, 12, 31, 36, 39, 40, 42, 43, 45, 64, 66, 72
- [App96] D. APPLEBAUM – *Probability and information, an integrated approach*, Cambridge University Press, Cambridge, 1996. 58
- [BAL93] G. BEN AROUS et M. LEDOUX – « Schilder’s large deviation principle without topology », *Asymptotic problems in probability theory: Wiener functionals and asymptotics* (Sanda/Kyoto, 1990), Longman Sci. Tech., Harlow, 1993, p. 107–121. 20, 23
- [BCEF01] F. BARTHE, D. CORDERO-ERAUSQUIN et M. FRADELIZI – « Shift inequalities of Gaussian type and norms of barycentres », *Studia Math.* **146** (2001), no. 3, p. 245–259. 7, 21, 68
- [BCL97] D. BAKRY, D. CONCORDET et M. LEDOUX – « Optimal heat kernel bounds under logarithmic Sobolev inequalities », *ESAIM Probab. Statist.* **1** (1997), p. 391–407 (electronic). 45
- [Bec75] W. BECKNER – « Inequalities in Fourier analysis », *Ann. of Math. (2)* **102** (1975), no. 1, p. 159–182. 46, 54, 57
- [Bec95] —, « Pitt’s inequality and the uncertainty principle », *Proc. Amer. Math. Soc.* **123** (1995), no. 6, p. 1897–1905. 53
- [Bec99] —, « Geometric asymptotics and the logarithmic Sobolev inequality », *Forum Math.* **11** (1999), no. 1, p. 105–137. 40, 64
- [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. 58
- [BH99] T. BODINEAU et B. HELFFER – « The log-Sobolev inequality for unbounded spin systems », *J. Funct. Anal.* **166** (1999), no. 1, p. 168–178. 77

- [BL76] H. J. BRASCAMP et E. H. LIEB – « Best constants in Young’s inequality, its converse, and its generalization to more than three functions », *Adv. Math.* **20** (1976), no. 2, p. 151–173. 46
- [BL96] D. BAKRY et M. LEDOUX – « Lévy-Gromov’s isoperimetric inequality for an infinite-dimensional diffusion generator », *Invent. Math.* **123** (1996), no. 2, p. 259–281. 20, 21, 23, 24
- [BL00] S. G. BOBKOV et M. LEDOUX – « From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities », *Geom. Funct. Anal.* **10** (2000), no. 5, p. 1028–1052. 48
- [Bla65] N. M. BLACHMAN – « The convolution inequality for entropy powers », *IEEE Trans. Information Theory* **IT-11** (1965), p. 267–271. 41, 57
- [Bob97] S. G. BOBKOV – « An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space », *Ann. Probab.* **25** (1997), no. 1, p. 206–214. 3, 20
- [Bob99] — , « The size of singular component and shift inequalities », *Ann. Probab.* **27** (1999), no. 1, p. 416–431. 7, 20, 21, 68
- [Bri64] L. BRILLOUIN – *Scientific uncertainty, and information*, Academic Press, New York, 1964. 30, 56
- [Cap01] P. CAPUTO – « A remark on spectral gap and logarithmic sobolev inequalities for conservative spin systems », Preprint archived as mp_arc 01-71, 2001. 73, 74
- [Cap02] — , « Uniform Poincaré inequalities for unbounded conservative spin systems: The non-interacting case », Preprint archived as mp_arc 02-46, 2002. 17
- [Car91] E. CARLEN – « Super-additivity of Fisher’s information and logarithmic Sobolev inequalities », *J. Funct. Anal.* **101** (1991), p. 194–211. 40, 41, 44, 55, 64
- [CGG89] T. M. COVER, P. GÁCS et R. M. GRAY – « Kolmogorov’s contributions to information theory and algorithmic complexity », *Ann. Probab.* **17** (1989), no. 3, p. 840–865. 58
- [Cha02a] D. CHAFAÏ – « Gaussian maximum of entropy and reversed logarithmic sobolev inequality », To appear in “Séminaire de Probabilité XXXVI, 2002”, october 2002. 1
- [Cha02b] — , « Glauber versus Kawasaki for spectral gap and logarithmic Sobolev inequalities of some unbounded conservative spin systems », preprint archived as mp_arc 02-30, january 2002. 1
- [CL99] D. CHAFAÏ et M. LEDOUX – « Méthodes fonctionnelles pour des grandes déviations quasi-gaussiennes. (On Gaussian-like large deviations by functional methods). », *C. R. Acad. Sci., Paris, Ser. I, Math.* **329** (1999), no. 6, p. 523–526 (Français). 1

- [CT91] T. M. COVER et J. A. THOMAS – *Elements of information theory*, John Wiley & Sons Inc., New York, 1991, A Wiley-Interscience Publication. 32, 33, 36, 37, 38, 39, 57, 58, 64, 66
- [DCD82] D. DACUNHA-CASTELLE et M. DUFLO – *Probabilités et statistiques. Tome 1*, Masson, Paris, 1982, Problèmes à temps fixe. 51
- [DCT91] A. DEMBO, T. M. COVER et J. A. THOMAS – « Information-theoretic inequalities », *IEEE Trans. Inform. Theory* **37** (1991), no. 6, p. 1501–1518. 37, 44, 45, 46, 47, 48, 57, 58, 64
- [Dem90] A. DEMBO – « Information inequalities and uncertainty principles », Tech. Rep., Dept. of Statist., Stanford Univ., 1990. 55, 57, 64
- [DGI00] J.-D. DEUSCHEL, G. GIACOMIN et D. IOFFE – « Large deviations and concentration properties for ∇_φ interface models. », *Probab. Theor. Relat. Fields* **117** (2000), p. 49–111. 17
- [DL84] R. DAUTRAY et 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. 53
- [EE90] P. EHRENFEST et 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. 56
- [Fan61] R. M. FANO – *Transmission of information: A statistical theory of communications.*, The M.I.T. Press, Cambridge, Mass., 1961. 58
- [Fei58] A. FEINSTEIN – *Foundations of information theory*, McGraw-Hill Electrical and Electronic Engineering Series. McGraw-Hill Book Co., Inc., New York-Toronto-London, 1958. 58
- [Fis22] A. FISHER, R. – « On the mathematical foundations of theoretical statistics », *Philos. Trans. Roy. Soc. Ann. Probab., London, Sec. A* (1922), no. 222, p. 309–368. 56
- [Fis25] — , « Theory of statistical estimation », *Proc. Cambridge Phil. Society* (1925), no. 22, p. 700–725. 56
- [FS98] H. G. FEICHTINGER et T. STROHMER (éds.) – *Gabor analysis and algorithms*, Birkhäuser Boston Inc., Boston, MA, 1998, Theory and applications. 56
- [FZ99] S. FANG et T. S. ZHANG – « On the small time behavior of Ornstein-Uhlenbeck processes with unbounded linear drifts », *Probab. Theory Related Fields* **114** (1999), no. 4, p. 487–504. 24

- [Geo88] H.-O. GEORGII – *Gibbs measures and phase transitions*, Walter de Gruyter & Co., Berlin, 1988. 36
- [GOS01] G. GIACOMIN, S. OLLA et H. SPOHN – « Equilibrium fluctuations for $\nabla\varphi$ interface model », Preprint, 2001. 17
- [Gra90] R. M. GRAY – *Entropy and information theory*, Springer-Verlag, New York, 1990. 58
- [Gro75] L. GROSS – « Logarithmic Sobolev inequalities », *Amer. J. Math.* **97** (1975), no. 4, p. 1061–1083. 64, 66
- [Har28] V. HARTLEY, R. – « Transmission of information », *Bell Sys. Tech. J.* (1928), no. 7, p. 535. 56
- [Hir57] I. I. HIRSCHMAN, JR. – « A note on entropy », *Amer. J. Math.* **79** (1957), p. 152–156. 57
- [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. 56
- [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. 56
- [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. 30, 37, 38, 58
- [KKK87] S. KULLBACK, J. C. KEEGEL et J. H. KULLBACK – *Topics in statistical information theory*, Springer-Verlag, Berlin, 1987. 58
- [KL99] C. KIPNIS et C. LANDIM – *Scaling limit of interacting particle systems*, Grundlehren der Mathematischen Wissenschaften, vol. 320, Springer, Berlin, New-York, 1999. 78
- [Kul97] S. KULLBACK – *Information theory and statistics*, Dover Publications Inc., Mineola, NY, 1997, Retirage de la seconde édition de 1968. 58
- [Led99] M. LEDOUX – « Concentration of measure and logarithmic Sobolev inequalities », Séminaire de Probabilités, XXXIII, Lecture Notes in Math., Springer, Berlin, 1999, p. 120–216. 4, 20, 21
- [Led00] — , « The geometry of Markov diffusion generators », *Ann. Fac. Sci. Toulouse Math. (6)* **9** (2000), no. 2, p. 305–366. 6, 7, 66, 68
- [Led01] — , « Logarithmic Sobolev inequalities for unbounded spin systems revisited », Séminaire de Probabilités, XXXV, Springer, Berlin, 2001, p. 167–194. 77
- [Lie78] E. H. LIEB – « Proof of an entropy conjecture of Wehrl », *Comm. Math. Phys.* **62** (1978), no. 1, p. 35–41. 47

- [Lie90] — , « Gaussian kernels have only Gaussian maximizers », *Invent. Math.* **102** (1990), no. 1, p. 179–208. 47, 57
- [LPY00] C. LANDIM, G. PANIZO et H. T. YAU – « Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems », Preprint, to appear in *Ann. Inst. H. Poincaré*, 2000. 2, 15, 16, 71, 77, 78, 80, 84, 85, 86, 89, 90
- [LY93] S. L. LU et H.-T. YAU – « Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics », *Comm. Math. Phys.* **156** (1993), no. 2, p. 399–433. 78, 82
- [Mal01] F. MALRIEU – « Convergence to equilibrium for granular media equations and their Euler schemes », preprint, to appear in “Annals of Applied Probability”, 2001. 77
- [Mar99] F. MARTINELLI – « Lectures on Glauber dynamics for discrete spin models », 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. 78
- [ME81] N. F. MARTIN et J. W. ENGLAND – « Mathematical theory of entropy. Foreword by James K. Brooks », *Encyclopedia of Mathematics and its Applications*, vol. 12, Addison-Wesley Publishing Company, 1981. 27, 56
- [Nyq24] H. NYQUIST – « Certain factors affecting telegraph speed », *Bell Sys. Tech. J.* (1924), no. 3, p. 324. 56
- [Pin64] M. S. PINSKER – *Information and information stability of random variables and processes*, Holden-Day Inc., San Francisco, Calif., 1964, Translated by Amiel Feinstein. 66
- [Rom92] S. ROMAN – *Coding and information theory*, Springer-Verlag, New York, 1992. 30, 31, 32, 33, 37, 38, 58
- [Rom97] — , *Introduction to coding and information theory*, Springer-Verlag, New York, 1997. 58
- [SC97] L. SALOFF-COSTE – « Lectures on finite Markov chains », 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. 75
- [Sha48] C. E. SHANNON – « A mathematical theory of communication », *Bell System Tech. J.* **27** (1948), p. 379–423, 623–656. 29, 30, 32, 38, 39, 41, 56, 57, 63, 64, 66
- [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. 58

- [Sta59] A. STAM – « Some inequalities satisfied by the quantities of information of Fisher and Shannon », *Information and Control* **2** (1959), p. 101–112. 41, 42, 53, 57, 64
- [Sto88] J. STORER – *Data compression. Methods and theory*, Computer Science Press, 1988. 33
- [SW49] C. E. SHANNON et W. WEAVER – *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Ill., 1949. 56
- [SZ92] D. STROOCK et B. ZEGARLIŃSKI – « The logarithmic Sobolev inequality for continuous spin systems on a lattice », *J. Funct. Anal.* **104** (1992), no. 2, p. 299–326.
- [Tos91] G. TOSCANI – « On Shannon’s entropy power inequality », *Ann. Univ. Ferrara Sez. VII (N.S.)* **37** (1991), p. 167–184 (1992). 57
- [Var67] S. R. S. VARADHAN – « Diffusion processes in a small time interval », *Comm. Pure Appl. Math.* **20** (1967), p. 659–685. 9, 24
- [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. 58
- [Voi98] D. VOICULESCU – « The analogues of entropy and of Fisher’s information measure in free probability theory. V. Noncommutative Hilbert transforms », *Invent. Math.* **132** (1998), no. 1, p. 189–227. 57
- [Wie48] N. WIENER – *Cybernetics, or Control and Communication in the Animal and the Machine*, Hermann et Cie., Paris, 1948, Actualités Sci. Ind., no. 1053. 56
- [Zam98] R. ZAMIR – « A proof of the Fisher information inequality via a data processing argument », *IEEE Trans. Inform. Theory* **44** (1998), no. 3, p. 1246–1250. 41, 52, 57
- [ZF93] R. ZAMIR et M. FEDER – « A generalization of the entropy power inequality with applications. », *IEEE Trans. Inform. Theory* **39** (1993), no. 5, p. 1723–1728 (English). 57
- [Zin96] M. ZINSMEISTER – *Formalisme thermodynamique et systèmes dynamiques holomorphes*, Panoramas et Synthèses, vol. 4, Société Mathématique de France. Paris, 1996. 27, 56

On logarithmic Sobolev inequalities in Information Theory and for conservative spin systems in Statistical Mechanics

Abstract

In a first part, it is shown that some Gaussian functional inequalities have simple generalisations to some Gaussian-like cases. They allow us to establish Gaussian-like Large Deviations Principles and bounds via Gaussian concentration and shift inequalities for certain families of BOLTZMANN measures and laws of diffusion semi groups in short time. Beyond the results themselves, we would like to emphasize here the method and the symmetry of the arguments used for upper and lower bounds by means of the functional inequalities.

The second 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.

The last part is inspired by the recent results of C. Landim, G. Panizo and H.-T. Yau [LPY] on spectral gap and logarithmic Sobolev inequalities for unbounded conservative spin systems in Statistical Mechanics. We study uniform bounds in these inequalities for Glauber dynamics of Hamiltonian of the form

$$\sum_{i=1}^n V(x_i) + V\left(M - \sum_{i=1}^n x_i\right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

Specifically, we examine the case V is strictly convex (or small perturbation of strictly convex) and, following [LPY], the case V is a bounded perturbation of a quadratic potential. By a simple path counting argument for the standard random walk, uniform bounds for the Glauber dynamics yields, in a transparent way, the classical L^{-2} decay for the Kawasaki dynamics on d -dimensional cubes of length L . The arguments of proofs however closely follow and make heavy use of the conservative approach and estimates of [LPY].

Keywords : Logarithmic Sobolev Inequality, Poincaré Inequality, Entropy, Statistical Mechanics, Large Deviations, Markov semigroup, Information Theory.

Field : Mathematics, Probability Theory.

Auteur : Djalil CHAFAÏ

Titre : Sur les inégalités de SOBOLEV logarithmiques en théorie de l'information et pour des systèmes de spins conservatifs en mécanique statistique.

Date et lieu de soutenance : le 17 mai 2002 à l'Université Paul SABATIER (Toulouse III).

Résumé

Dans une première partie, nous montrons comment certaines inégalités fonctionnelles gaussiennes possédant des généralisations simples dans des cas quasi-gaussiens permettent d'établir des principes et bornes de grandes déviations (PGD) quasi-gaussiens via des inégalités de concentration et de translation gaussiennes pour certaines familles de mesures de BOLTZMANN et lois de certains semi-groupes diffusifs en temps petit. Au-delà des résultats, l'accent est mis ici sur la méthode et la symétrie des arguments utilisés pour les bornes inférieures et supérieures.

La seconde partie consiste en une synthèse sur les liens existant entre les mathématiques construites autour des inégalités de SOBOLEV logarithmiques, et certaines inégalités faisant intervenir l'entropie de SHANNON ou l'information de FISHER. Ces dernières sont liées également à d'autres inégalités en analyse mathématique (BRUNN-MINKOWSKI, YOUNG), en statistique (inégalité de CRAMÉR-RAO), et en physique (principes d'incertitude de BECKNER-HIRSCHMAN et de WEYL-HEISENBERG). Nous faisons également le lien entre une forme inverse de l'inégalité de SOBOLEV logarithmique et le maximum gaussien de l'entropie exponentielle de SHANNON.

Enfin, la dernière partie consiste en une étude des inégalités de Poincaré et de Sobolev logarithmique pour des mesures de probabilités liées à des modèles conservatifs à spins continus en mécanique statistique. Inspiré des résultats récents de C. LANDIM, G. PANIZO et H.-T. YAU [LPY] sur le trou spectral et l'inégalité de Sobolev logarithmique pour des systèmes conservatifs de spins non bornés, nous étudions des bornes *uniformes* dans ces inégalités pour la dynamique de GLAUBER associée à des hamiltoniens de la forme $\sum_{i=1}^n V(x_i) + V(M - x_1 - \dots - x_n)$. Plus précisément, nous examinons le cas où V est strictement convexe et, en suivant [LPY], le cas où V est une perturbation bornée d'un potentiel quadratique. Par un argument de chemins élémentaire pour la marche aléatoire simple, les bornes uniformes pour la dynamique de GLAUBER entraînent, de façon limpide, la décroissance classique en L^{-2} pour la dynamique de KAWASAKI sur des cubes d'arête L d'un réseau de dimension d . Cependant, les arguments de nos preuves suivent de près et utilisent abondamment les estimées et la démarche de [LPY].

Mots clés : Inégalité de SOBOLEV logarithmique - Inégalité de POINCARÉ - Entropie - Mécanique statistique - Grandes déviations - Semi-groupe de MARKOV - Théorie de l'information.

Discipline : Mathématiques, Probabilités.