CONCENTRATION OF MEASURE
AND LOGARITHMIC SOBOLEV INEQUALITIES

Berlin, 3-7 November, 1997

Michel Ledoux

Institut de Mathématiques
Université Paul-Sabatier
31062, Toulouse (France)
ledoux@math.ups-tlse

TABLE OF CONTENTS

INTRODUCTION

The concentration of measure phenomenon was put forward in the seventies by V. D. Milman in the local theory of Banach spaces. Of isoperimetric inspiration, it is of powerful interest in applications, in particular in probability theory (probability in Banach spaces, empirical processes, geometric probabilities, statistical mechanics...) One main example is the Gaussian concentration property which expresses that, whenever A is a Borel set in $\mathbb{R}^n$ of canonical Gaussian measure $\gamma(A) \geq \frac{1}{2}$, for every $r \geq 0$,

$$\gamma(A_r) \geq 1 - \mathrm{e}^{-r^2/2}$$

where A_r is the r-th Euclidean neighborhood of A. As r increases, the enlargement A_r thus gets very rapidly a measure close to one. This Gaussian concentration property can be described equivalently on functions. If F is a Lipschitz map on $\mathbb{R}^n$ with $\|F\|_{\mathrm{Lip}} \leq 1$, for every $r \geq 0$,

$$\gamma\big(F \geq \textstyle\int F d\gamma + r\big) \leq \mathrm{e}^{-r^2/2}.$$

Together with the same inequality for $-F$, the Lipschitz function F is seen to be concentrated around some mean value with very high probability. These quantitative estimates are dimension free and extend to arbitrary infinite dimensional Gaussian measures. As such, they are a main tool in the study of Gaussian processes and measures.

Simultaneously, hypercontractive estimates and logarithmic Sobolev inequalities came up in quantum field theory with the contributions of E. Nelson and L. Gross. In particular, L. Gross proved in 1975 a Sobolev inequality for Gaussian measures of logarithmic type. Namely, for all smooth functions f on $\mathbb{R}^n$,

$$\int f^2 \log f^2 d\gamma - \int f^2 d\gamma \log \int f^2 d\gamma \leq 2 \int |\nabla f|^2 d\gamma.$$

This inequality is again independent of the dimension and proved to be a substitute of the classical Sobolev inequalities in infinite dimensional settings. Logarithmic Sobolev inequalities have been used extensively in the recent years as a way to measure the smoothing properties (hypercontractivity) of Markov semigroups. In particular, they are a basic ingredient in the investigation of the time to equilibrium.

One of the early questions on logarithmic Sobolev inequalities was to determine which measures, on $\mathbb{R}^n$, satisfy an inequality similar to the one for Gaussian measures. To this question, raised by L. Gross, I. Herbst (in an unpublished letter to L.

Gross) found the following necessary condition: if μ is a probability measure such that for some $C > 0$ and every smooth function f on $\mathbb{R}^n$,

$$\int f^2 \log f^2 d\mu - \int f^2 d\mu \log \int f^2 d\mu \leq C \int |\nabla f|^2 d\mu,$$

then,

$$\int e^{\alpha |x|^2} d\mu(x) < \infty$$

for every $\alpha < \frac{1}{C}$. Furthermore, for any Lipschitz function F on $\mathbb{R}^n$ with $\|F\|_{\mathrm{Lip}} \leq 1$, and every real λ,

$$\int e^{\lambda F} d\mu \leq e^{\lambda \int F d\mu + C\lambda^2/4}.$$

By a simple use of Chebyshev's inequality, the preceding thus relates in an essential way to the Gaussian concentration phenomenon.

Herbst's result was mentioned in the early eighties by E. Davies and B. Simon, and has been revived recently by S. Aida, T. Masuda and I. Shigekawa. It was further developed and refined by S. Aida, S. Bobkov, F. Götze, L. Gross, O. Rothaus, D. Stroock and the author. Following these authors and their contributions, the aim of these notes is to present a complete account on the applications of logarithmic Sobolev inequalities to the concentration of measure phenomenon. We exploit Herbst's original argument to deduce from the logarithmic Sobolev inequalities some differential inequalities on the Laplace transforms of Lipschitz functions. According to the family of entropy-energy inequalities we are dealing with, these differential inequalities yield various behaviors of the Laplace transforms of Lipschitz functions and of their concentration properties. In particular, the basic product property of entropy allows us to investigate with this tool concentration properties in product spaces. The principle is rather simple minded, and as such convenient for applications.

The first part of this set of notes includes a introduction to isoperimetry and concentration for Gaussian and Boltzmann measures. The second part then presents spectral gap and logarithmic Sobolev inequalities, and describes Herbst's basic Laplace transform argument. In the third part, we investigate by this method deviation and concentration inequalities for product measures. While concentration inequalities do not necessarily tensorize, we show that they actually follow from stronger logarithmic Sobolev inequalities. We thus recover most of M. Talagrand's recent results on isoperimetric and concentration inequalities in product spaces. We briefly mention there the information theoretic inequalities by K. Marton which provide an alternate approach to concentration also based on entropy, and which seems to be well suited to dependent structures. We then develop the subject of modified logarithmic Sobolev inequalities investigated recently in joint works with S. Bobkov. We examine in this way concentration properties for the product measure of the exponential distribution, as well as, more generally, of measures satisfying a Poincaré inequality. In the next section, the analogous questions for discrete gradients are addressed, with particular emphasis on Bernoulli and Poisson measures. We then present some applications to large deviation upper bounds and to tail estimates for

Brownian motion on a manifold. In the final part, we discuss some recent results on the logarithmic Sobolev constant in Riemannian manifolds with non-negative Ricci curvature. The last section is an addition of L. Saloff-Coste on the logarithmic Sobolev constant and the diameter for Markov chains. We sincerely thank him for this contribution.

It is a pleasure to thank the organizers (in particular M. Scheutzow) and the participants of the "Graduierten- kolleg" course which was held in Berlin in November 1997 for the opportunity to present, and to prepare, these notes. These notes would not exist without the collaboration with S. Bobkov which led to the concept of modified logarithmic Sobolev inequality and whose joint work form most of Parts 4 and 5. Thanks are also due to S. Kwapień for numerous exchanges over the years on the topic of these notes. D. Piau and D. Steinsaltz were very helpful with their comments and corrections on the manuscript.

With respect to the version published in the Séminaire de Probabilités XXXIII, these notes benefited from several corrections by L. Miclo that we warmly thank for his help.

1. ISOPERIMETRIC AND CONCENTRATION INEQUALITIES

In this first part, we present the Gaussian isoperimetric inequality as well as a Gaussian type isoperimetric inequality for a class of Boltzmann measures with a sufficiently convex potential. Isoperimetry is a natural way to introduce to the concentration of measure phenomenon. For completness, we propose a rather short, self-contained proof of these isoperimetric inequalities following the recent contributions [Bob4], [Ba-L]. Let us mention however that our first goal in these notes is to produce simpler, more functional arguments to derive concentration properties. We then present the concentration of measure phenomenon, and discuss a few of its first properties.

1.1 Introduction

The classical isoperimetric inequality in Euclidean space states that among all subsets with fixed finite volume, balls achieve minimal surface area. In probabilistic, and also geometric, applications one is often interested in finite measure space, such as the unit sphere S^n in $\mathbb{R}^{n+1}$ equipped with its normalized invariant measure σ^n. On S^n, (geodesic) balls, or caps, are again the extremal sets, that is achieve minimal surface measure among sets with fixed measure.

The isoperimetric inequality on the sphere was used by V. D. Milman in the early seventies as a tool to prove the famous Dvoretzky theorem on Euclidean sections of convex bodies (cf. [Mi], [M-S]). Actually, V. D. Milman is using the isoperimetric property as a concentration property. Namely, in its integrated version, the isoperimetry inequality states that whenever $\sigma^n(A) = \sigma^n(B)$ where B is a ball on S^n, for every $r \geq 0$,

$$\sigma^n(A_r) \geq \sigma^n(B_r) \tag{1.1}$$

where A_r (resp. B_r) is the neighborhood of order r of A (resp. B) for the geodesic metric on the sphere. Since, for a set A on S^n with smooth boundary ∂A, the surface measure σ_s^n of ∂A can be described by the Minkowski content formula as

$$\sigma_s^n(\partial A) = \liminf_{r \to 0} \frac{1}{r} \left[\sigma^n(A_r) - \sigma^n(A)\right],$$

(1.1) is easily seen to be equivalent to the isoperimetric statement. Now, the measure of a cap may be estimated explicitely. For example, if $\sigma^n(A) \geq \frac{1}{2}$, it follows from

(1.1) that

$$\sigma^n(A_r) \geq 1 - \sqrt{\tfrac{\pi}{8}}\, \mathrm{e}^{-(n-1)r^2/2} \tag{1.2}$$

for every $r \geq 0$. Therefore, if the dimension is large, only a small increase of r (of the order of $\frac{1}{\sqrt{n}}$) makes the measure of A_r close to 1. In a sense, the measure σ^n is concentrated around the equator, and (1.2) describes the so-called concentration of measure phenomenon of σ^n. One significant aspect of this concentration phenomenon is that the enlargements are not infinitesimal as for isoperimetry, and that emphasis is not on extremal sets. These notes will provide a sample of concentration properties with the functional tool of logarithmic Sobolev inequalities.

1.2 Isoperimetric inequalities for Gaussian and Boltzmann measures

It is well known that uniform measures on n-dimensional spheres with radius $\sqrt{n}$ approximate (when projected on a finite number of coordinates) Gaussian measures (Poincaré's lemma). In this sense, the isoperimetric inequality on spheres gives rise to an isoperimetric inequality for Gaussian measures (cf. [Le3]). Extremal sets are then half-spaces (which may be considered as balls with centers at infinity). Let, more precisely, $\gamma = \gamma^n$ be the canonical Gaussian measure on $\mathbb{R}^n$ with density

$$(2\pi)^{-n/2} \exp(-|x|^2/2)$$

with respect to Lebesgue measure. Define the Gaussian surface measure of a Borel set A in $\mathbb{R}^n$ as

$$\gamma_s(\partial A) = \liminf_{r \to 0} \frac{1}{r} \big[\gamma(A_r) - \gamma(A)\big] \tag{1.3}$$

where $A_r = \{x \in \mathbb{R}^n; d_2(x, A) < r\}$ is the r-Euclidean open neighborhood of A. Then, if H is a half-space in $\mathbb{R}^n$, that is $H = \{x \in \mathbb{R}^n; \langle x, u\rangle < a\}$, where $|u| = 1$ and $a \in [-\infty, +\infty]$, and if $\gamma(A) = \gamma(H)$, then

$$\gamma_s(\partial A) \geq \gamma_s(\partial H).$$

Let $\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2} dx$, $t \in [-\infty, +\infty]$, be the distribution function of the canonical Gaussian measure in dimension one and let $\varphi = \Phi'$. Then $\gamma(H) = \Phi(a)$ and $\gamma_s(\partial H) = \varphi(a)$ so that,

$$\gamma_s(\partial A) \geq \varphi(a) = \varphi \circ \Phi^{-1}\big(\gamma(A)\big). \tag{1.4}$$

Moreover, half-spaces are the extremal sets in this inequality. In this form, the Gaussian isoperimetric inequality is dimension free.

In applications, the Gaussian isoperimetric inequality is often used in its integrated version. Namely, if $\gamma(A) = \gamma(H) = \Phi(a)$ (or only $\gamma(A) \geq \Phi(a)$), then, for every $r \geq 0$,

$$\gamma(A_r) \geq \gamma(H_r) = \Phi(a + r). \tag{1.5}$$

In particular, if $\gamma(A) \geq \frac{1}{2}$ $(= \Phi(0))$,

$$\gamma(A_r) \geq \Phi(r) \geq 1 - \mathrm{e}^{-r^2/2}. \tag{1.6}$$

To see that (1.4) implies (1.5), we may assume, by a simple approximation, that A is given by a finite union of open balls. The family of such sets A is closed under the operation $A \mapsto A_r$, $r \geq 0$. Then, the lim inf in (1.3) is a true limit. Actually, the boundary ∂A of A is a finite union of piecewise smooth $(n-1)$-dimensional surfaces in $\mathbb{R}^n$ and $\gamma_s(\partial A)$ is given by the integral of the Gaussian density along ∂A with respect to Lebesgue measure on ∂A. Now, by (1.4), the function $v(r) = \Phi^{-1} \circ \gamma(A_r)$, $r \geq 0$, satisfies

$$v'(r) = \frac{\gamma_s(\partial A_r)}{\varphi \circ \Phi^{-1}(\gamma(A_r))} \geq 1$$

so that $v(r) = v(0) + \int_0^r v'(s)ds \geq v(0) + r$, which is (1.5). (Alternatively, see [Bob3].)

The Euclidean neighborhood A_r of a Borel set A can be viewed as the Minkowski sum $A + rB_2 = \{a + rb\,; a \in A, b \in B_2\}$ with B_2 the Euclidean open unit ball. If γ is any (centered) Gaussian measure on $\mathbb{R}^n$, B_2 has to be replaced by the ellipsoid associated to the covariance structure of γ. More precisely, denote by $\Gamma = M\,{}^t\!M$ the covariance matrix of the Gaussian measure γ on $\mathbb{R}^n$. Then γ is the image of the canonical Gaussian measure by the linear map $M = (M_{ij})_{1 \leq i,j \leq n}$. Set $\mathcal{K} = M(B_2)$. Then, if $\gamma(A) \geq \Phi(a)$, for every $r \geq 0$,

$$\gamma\big(A + r\mathcal{K}\big) \geq \Phi(a + r). \tag{1.7}$$

In this formulation, the Gaussian isoperimetric inequality extends to infinite dimensional (centered) Gaussian measures, the set $\mathcal{K}$ being the unit ball of the reproducing kernel Hilbert space $\mathcal{H}$ (the Cameron-Martin space for Wiener measure for example). Cf. [Bor], [Le3].

To see moreover how (1.6) or (1.7) may be used in applications, let for example $X = (X_t)_{t \in T}$ be a centered Gaussian process indexed by some, for simplicity, countable parameter set T. Assume that $\sup_{t \in T} X_t < \infty$ almost surely. Fix $t_1, \ldots, t_n$ in T and consider the distribution γ of the sample $(X_{t_1}, \ldots, X_{t_n})$. Choose m finite such that $\mathbb{P}\{\sup_{t \in T} X_t \leq m\} \geq \frac{1}{2}$. In particular, if

$$A = \Big\{\max_{1 \leq i \leq n} X_{t_i} \leq m\Big\},$$

then $\gamma(A) \geq \frac{1}{2}$. Therefore, by (1.7) (with $a = 0$), for every $r \geq 0$,

$$\gamma\big(A + r\mathcal{K}\big) \geq \Phi(r) \geq 1 - \mathrm{e}^{-r^2/2}.$$

Now, for any h in $\mathcal{K} = M(B_2)$,

$$\max_{1 \leq i \leq n} h_i \leq \max_{1 \leq i \leq n} \bigg(\sum_{j=1}^n M_{ij}^2\bigg)^{1/2} = \max_{1 \leq i \leq n} \big(\mathbb{E}(X_{t_i}^2)\big)^{1/2}$$

by the Cauchy-Schwarz inequality, so that

$$A + r\mathcal{K} \subset \Big\{\max_{1 \leq i \leq n} X_{t_i} \leq m + r \max_{1 \leq i \leq n} \big(\mathbb{E}(X_{t_i}^2)\big)^{1/2}\Big\}.$$

Set $\sigma = \sup_{t\in T}(\mathbb{E}(X_t^2))^{1/2}$. (It is easily seen that σ is always finite under the assumption $\sup_{t\in T} X_t < \infty$. Let indeed m' be such that $\mathbb{P}\{\sup_{t\in T} X_t \le m'\} \ge \frac{3}{4}$. Then, if $\sigma_t = (\mathbb{E}(X_t^2))^{1/2}$, $\frac{m'}{\sigma_t} \ge \Phi^{-1}(\frac{3}{4}) > 0$.) It follows from the preceding that

$$\mathbb{P}\Big\{\max_{1\le i\le n} X_{t_i} \le m + \sigma r\Big\} \ge 1 - \mathrm{e}^{-r^2/2}.$$

By monotone convergence, and taking complements, for every $r \ge 0$,

$$\mathbb{P}\Big\{\sup_{t\in T} X_t \ge m + \sigma r\Big\} \le \mathrm{e}^{-r^2/2}. \tag{1.8}$$

This inequality describes the strong integrability properties of almost surely bounded Gaussian processes. It namely implies in particular (cf. Proposition 1.2 below) that for every $\alpha < \frac{1}{2\sigma^2}$,

$$\mathbb{E}\Big(\exp\big(\alpha(\sup\nolimits_{t\in T} X_t)^2\big)\Big) < \infty. \tag{1.9}$$

Equivalently, in a large deviation formulation,

$$\lim_{r\to\infty} \frac{1}{r^2}\log \mathbb{P}\Big\{\sup_{t\in T} X_t \ge r\Big\} = -\frac{1}{2\sigma^2}\,. \tag{1.10}$$

(The lower bound in (1.10) is just that

$$\mathbb{P}\Big\{\sup_{t\in T} X_t \ge r\Big\} \ge \mathbb{P}\{X_t \ge r\} = 1 - \Phi\Big(\frac{r}{\sigma_t}\Big) \ge \frac{\mathrm{e}^{-r^2/2\sigma_t^2}}{\sqrt{2\pi}(1 + (r/\sigma_t))}$$

for every $t \in T$ and $r \ge 0$.) But inequality (1.8) actually contains more information than just this integrability result. (For example, if X^n is a sequence of Gaussian processes as before, and if we let $\|X^n\| = \sup_{t\in T} X_t^n$, $n \in \mathbb{N}$, then $\|X^n\| \to 0$ almost surely as soon as $\mathbb{E}(\|X^n\|) \to 0$ and $\sigma^n\sqrt{\log n} \to 0$ where $\sigma^n = \sup_{t\in T}(\mathbb{E}((X_t^n)^2))^{1/2}$.) (1.8) describes a sharp deviation inequality in terms of two parameters, m and σ. In this sense, it belongs to the concentration of measure phenomenon which will be investigated in these notes (cf. Section 1.3). Note that (1.8), (1.9), (1.10) hold similarly with $\sup_{t\in T} X_t$ replaced by $\sup_{t\in T}|X_t|$ (under the assumption $\sup_{t\in T}|X_t| < \infty$ almost surely).

The Gaussian isoperimetric inequality was established in 1974 independently by C. Borell [Bor] and V. N. Sudakov and B. S. Tsirel'son [S-T] on the basis of the isoperimetric inequality on the sphere and Poincaré's lemma. A proof using Gaussian symmetrizations was developed by A. Ehrhard in 1983 [Eh]. We present here a short and self-contained proof of this inequality. Our approach will be functional. Denote by $\mathcal{U} = \varphi \circ \Phi^{-1}$ the Gaussian isoperimetric function in (1.4). In a recent striking paper, S. Bobkov [Bob4] showed that for every smooth enough function f with values in the unit interval $[0,1]$,

$$\mathcal{U}\Big(\int f\,d\gamma\Big) \le \int \sqrt{\mathcal{U}^2(f) + |\nabla f|^2}\,d\gamma \tag{1.11}$$

where $|\nabla f|$ denotes the Euclidean length of the gradient ∇f of f. It is easily seen that (1.11) is a functional version of the Gaussian isoperimetric inequality (1.4). Namely, if (1.11) holds for all smooth functions, it holds for all Lipschitz functions with values in $[0,1]$. Assume again that the set A in (1.4) is a finite union of non-empty open balls. In particular, $\gamma(\partial A) = 0$. Apply then (1.11) to $f_r(x) = (1 - \frac{1}{r}\, d_2(x,A))^+$ (where d_2 is the Euclidean distance function). Then, as $r \to 0$, $f_r \to I_A$ and $\mathcal{U}(f_r) \to 0$ almost everywhere since $\gamma(\partial A) = 0$ and $\mathcal{U}(0) = \mathcal{U}(1) = 0$. Moreover, $|\nabla f_r| = 0$ on A and on the complement of the closure of A_r, and $|\nabla f_r| \leq \frac{1}{r}$ everywhere. Note that the sets $\partial(A_r)$ are of measure zero for every $r \geq 0$. Therefore

$$\mathcal{U}\big(\gamma(A)\big) \leq \liminf_{r\to 0} \int |\nabla f_r| d\gamma \leq \liminf_{r\to 0} \frac{1}{r}\, \big[\gamma(A_r) - \gamma(A)\big] = \gamma_s(\partial A).$$

To prove (1.11), S. Bobkov first establishes the analogous inequality on the two-point space and then uses the central limit theorem, very much as L. Gross in his proof of the Gaussian logarithmic Sobolev inequality [Gr1] (cf. Section 2.2). The proof below is direct. Our main tool will be the so-called Ornstein-Uhlenbeck or Hermite semigroup with invariant measure the canonical Gaussian measure γ. For every f, in $\mathrm{L}^1(\gamma)$ say, set

$$P_t f(x) = \int_{\mathbb{R}^n} f\big(\mathrm{e}^{-t/2}x + (1-\mathrm{e}^{-t})^{1/2}y\big) d\gamma(y), \quad x \in \mathbb{R}^n, \quad t \geq 0. \tag{1.12}$$

The operators P_t are contractions on all $\mathrm{L}^p(\gamma)$-spaces, and are symmetric and invariant with respect to γ. That is, for any sufficiently integrable functions f and g, and every $t \geq 0$, $\int f P_t g d\gamma = \int g P_t f d\gamma$. The family $(P_t)_{t\geq 0}$ is a semigroup ($P_s \circ P_t = P_{s+t}$). P_0 is the identity operator whereas $P_t f$ converges in $\mathrm{L}^2(\gamma)$ towards $\int f d\gamma$ as t tends to infinity. All these properties are immediately checked on the preceding integral representation of P_t together with the elementary properties of Gaussian measures. The infinitesimal generator of the semigroup $(P_t)_{t\geq 0}$, that is the operator L such that

$$\frac{d}{dt}\, P_t f = P_t \mathrm{L} f = \mathrm{L} P_t f,$$

acts on all smooth functions f on $\mathbb{R}^n$ by

$$\mathrm{L}f(x) = \tfrac{1}{2}\, \Delta f(x) - \tfrac{1}{2}\, \langle x, \nabla f(x)\rangle.$$

In other words, L is the generator of the Ornstein-Uhlenbeck diffusion process $(X_t)_{t\geq 0}$, the solution of the stochastic differential equation $dX_t = dB_t - \frac{1}{2} X_t dt$ where $(B_t)_{t\geq 0}$ is standard Brownian motion in $\mathbb{R}^n$. Moreover, the integration by parts formula for L indicates that, for f and g smooth enough on $\mathbb{R}^n$,

$$\int f(-\mathrm{L}g) d\gamma = \frac{1}{2} \int \langle \nabla f, \nabla g\rangle d\gamma. \tag{1.13}$$

Let now f be a fixed smooth function on $\mathbb{R}^n$ with values in $[0,1]$. It might actually be convenient to assume throughout the argument that $0 < \varepsilon \leq f \leq 1 - \varepsilon$

and let then ε tend to zero. Recall $\mathcal{U} = \varphi \circ \Phi^{-1}$. To prove (1.11) it will be enough to show that the function

$$J(t) = \int \sqrt{\mathcal{U}^2(P_t f) + |\nabla P_t f|^2}\, d\gamma$$

is non-increasing in $t \geq 0$. Indeed, if this is the case, $J(\infty) \leq J(0)$, which, together with the elementary properties of P_t recalled above, amounts to (1.11). Towards this goal, we first emphasize the basic property of the Gaussian isoperimetric function $\mathcal{U}$ that will be used in the argument, namely that $\mathcal{U}$ satisfies the fundamental differential equality $\mathcal{U}\mathcal{U}'' = -1$ (exercise). We now have

$$\frac{dJ}{dt} = \int \frac{1}{\sqrt{\mathcal{U}^2(P_t f) + |\nabla P_t f|^2}} \Big[\mathcal{U}\mathcal{U}'(P_t f) \mathrm{L} P_t f + \langle \nabla(P_t f), \nabla(\mathrm{L} P_t f)\rangle\Big] d\gamma.$$

To ease the notation, write f for $P_t f$. We also set $K(f) = \mathcal{U}^2(f) + |\nabla f|^2$. Therefore,

$$\frac{dJ}{dt} = \int \frac{1}{\sqrt{K(f)}} \Big[\mathcal{U}\mathcal{U}'(f) \mathrm{L} f + \langle \nabla f, \nabla(\mathrm{L} f)\rangle\Big] d\gamma. \tag{1.14}$$

For simplicity in the exposition, let us assume that the dimension n is one, the general case being entirely similar, though notationally a little bit heavier. By the integration by parts formula (1.13),

$$\begin{aligned}
\int \frac{1}{\sqrt{K(f)}}\, \mathcal{U}\mathcal{U}'(f)\, \mathrm{L} f d\gamma &= -\frac{1}{2} \int \bigg(\frac{\mathcal{U}\mathcal{U}'(f)}{\sqrt{K(f)}}\bigg)' f' d\gamma \\
&= -\frac{1}{2} \int \frac{1}{\sqrt{K(f)}} \big[\mathcal{U}'^2(f) - 1\big] f'^2 d\gamma \\
&\quad + \frac{1}{2} \int \frac{\mathcal{U}\mathcal{U}'(f) f'}{K(f)^{3/2}} \big[\mathcal{U}\mathcal{U}'(f) f' + f' f''\big] d\gamma
\end{aligned}$$

where we used that $\mathcal{U}\mathcal{U}'' = -1$ and that

$$K(f)' = 2\mathcal{U}\mathcal{U}'(f) f' + \big(f'^2\big)' = 2\mathcal{U}\mathcal{U}'(f) f' + 2f' f''. \tag{1.15}$$

In order to handle the second term on the right-hand side of (1.14), let us note that

$$\langle \nabla f, \nabla(\mathrm{L} f)\rangle = \tfrac{1}{2}\, f' \big(f'' - x f'\big)' = -\tfrac{1}{2}\, f'^2 + f' \mathrm{L} f'.$$

Hence, again by the integration by parts formula (1.13), and by (1.15),

$$\begin{aligned}
\int \frac{1}{\sqrt{K(f)}} \langle \nabla f, \nabla(\mathrm{L} f)\rangle d\gamma &= -\frac{1}{2} \int \frac{f'^2}{\sqrt{K(f)}}\, d\gamma + \int \frac{f'}{\sqrt{K(f)}}\, \mathrm{L} f' d\gamma \\
&= -\frac{1}{2} \int \frac{f'^2}{\sqrt{K(f)}}\, d\gamma - \frac{1}{2} \int \frac{f''^2}{\sqrt{K(f)}}\, d\gamma \\
&\quad + \frac{1}{2} \int \frac{f' f''}{K(f)^{3/2}} \big[\mathcal{U}\mathcal{U}'(f) f' + f' f''\big] d\gamma.
\end{aligned}$$

Putting these equations together, we get, after some algebra,

$$\frac{dJ}{dt} = -\frac{1}{2}\int \frac{1}{K(f)^{3/2}}\Big[\mathcal{U}'^2(f)f'^4 - 2\mathcal{U}\mathcal{U}'(f)f'^2 f'' + \mathcal{U}^2(f)f''^2\Big]d\gamma$$

and the result follows since

$$\mathcal{U}'^2(f)f'^4 - 2\mathcal{U}\mathcal{U}'(f)f'^2 f'' + \mathcal{U}^2(f)f''^2 = \Big(\mathcal{U}'(f)f'^2 - \mathcal{U}(f)f''\Big)^2 \geq 0.$$

The preceding proof of the Gaussian isoperimetric inequality came up in the joint work [Ba-L] with D. Bakry. The argument is developed there in an abstract framework of Markov diffusion generators and semigroups and applies to a large class of invariant measures of diffusion generators satisfying a curvature assumption. We present here this result for some concrete class of Boltzmann measures for which a Gaussian-like isoperimetric inequality holds.

Let us consider a smooth (C^2 say) function W on $\mathbb{R}^n$ such that e^{-W} is integrable with respect to Lebesgue measure. Define the so-called Boltzmann measure as the probability measure

$$d\mu(x) = Z^{-1}\mathrm{e}^{-W(x)}dx$$

where Z is the normalization factor. As is well-known, μ may be described as the invariant measure of the generator $\mathrm{L} = \frac{1}{2}\Delta - \frac{1}{2}\nabla W \cdot \nabla$. Alternatively, L is the generator of the Markov semigroup $(P_t)_{t\geq 0}$ of the Kolmogorov process $X = (X_t)_{t\geq 0}$ solution of the stochastic differential Langevin equation

$$dX_t = dB_t - \tfrac{1}{2}\,\nabla W(X_t)dt.$$

The choice of $W(x) = \frac{1}{2}|x|^2$ with invariant measure the canonical Gaussian measure corresponds to the Ornstein-Uhlenbeck process. Denote by $W''(x)$ the Hessian of W at $x \in \mathbb{R}^n$.

Theorem 1.1. *Assume that, for some $c > 0$, $W''(x) \geq c\,\mathrm{Id}$ as symmetric matrices, uniformly in $x \in \mathbb{R}^n$. Then, whenever A is a Borel set in $\mathbb{R}^n$ with $\mu(A) \geq \Phi(a)$, for any $r \geq 0$,*

$$\mu(A_r) \geq \Phi(a + \sqrt{c}\,r).$$

As in the Gaussian case, the inequality of Theorem 1.1 is equivalent to its infinitesimal version

$$\mu_s(\partial A) \geq \sqrt{c}\,\mathcal{U}\big(\mu(A)\big)$$

with the corresponding notion of surface measure and to the functional inequality

$$\mathcal{U}\Big(\int f d\mu\Big) \leq \int \sqrt{\mathcal{U}^2(f) + \tfrac{1}{c}\,|\nabla f|^2}\,d\mu$$

which is the result we established (at least in one direction) in the proof as before. Before turning to this proof, let us comment on the Gaussian aspect of the theorem.

Let F be a Lipschitz map on $\mathbb{R}^n$ with Lipschitz coefficient $\|F\|_{\mathrm{Lip}} \leq \sqrt{c}$. Then, the image measure ν of μ by F is a contraction of the canonical Gaussian measure on $\mathbb{R}$. Indeed, we may assume by some standard regularization procedure (cf. [Ba-L]) that ν is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$ with a strictly positive density. Set $\nu(r) = \nu((-\infty, r])$ so that the measure ν has density ν'. For $r \in \mathbb{R}$, apply Theorem 1.1, or rather its infinitesimal version, to $A = \{F \leq r\}$ to get $\mathcal{U}(\nu(r)) \leq \nu'(r)$. Then, setting $k = \nu^{-1} \circ \Phi$ and $x = \Phi^{-1} \circ \nu(r)$, $k'(x) \leq 1$ so that ν is the image of the canonical Gaussian measure on $\mathbb{R}$ by the contraction k. In particular, in dimension one, every measure satisfying the hypothesis of Theorem 1.1 is a Lipschitz image of the canonical Gaussian measure.

Proof of Theorem 1.1. It is entirely similar to the proof of the Gaussian isoperimetric inequality in Section 1.1. Denote thus by $(P_t)_{t\geq 0}$ the Markov semigroup with generator $\mathrm{L} = \frac{1}{2}\Delta - \frac{1}{2}\nabla W \cdot \nabla$. The integration by parts formula for L reads

$$\int f(-\mathrm{L}g)d\mu = \frac{1}{2}\int \langle \nabla f, \nabla g\rangle d\mu$$

for smooth functions f and g. Fix a smooth function f on $\mathbb{R}^n$ with $0 \leq f \leq 1$. As in the Gaussian case, we aim to show that, under the assumption on W,

$$J(t) = \int \sqrt{\mathcal{U}^2(P_t f) + \tfrac{1}{c}\,|\nabla P_t f|^2}\, d\mu$$

is non-increasing in $t \geq 0$. Remaining as before in dimension one for notational simplicity, the argument is the same than in the Gaussian case with now $K(f) = \mathcal{U}^2(f) + \frac{1}{c}|\nabla f|^2$ so that

$$K(f)' = 2\mathcal{U}\mathcal{U}'(f)f' + \tfrac{2}{c}\, f'f''.$$

Similarly,

$$\langle \nabla f, \nabla(\mathrm{L}f)\rangle = f'\big(\tfrac{1}{2}\, f'' - \tfrac{1}{2}\, W'f'\big)' = -\tfrac{1}{2}\, W''f'^2 + f'\mathrm{L}f'.$$

Hence, again by the integration by parts formula,

$$\begin{aligned}\int \frac{1}{\sqrt{K(f)}}\,\langle \nabla f, \nabla(\mathrm{L}f)\rangle d\gamma &= -\frac{1}{2}\int \frac{W''f'^2}{\sqrt{K(f)}}\,d\mu + \int \frac{f'}{\sqrt{K(f)}}\,\mathrm{L}f'd\mu \\ &= -\frac{1}{2}\int \frac{W''f'^2}{\sqrt{K(f)}}\,d\mu - \frac{1}{2}\int \frac{f''^2}{\sqrt{K(f)}}\,d\mu \\ &\qquad + \frac{1}{2}\int \frac{f'f''}{K(f)^{3/2}}\Big[\mathcal{U}\mathcal{U}'(f)f' + \frac{1}{c}\,f'f''\Big]d\mu.\end{aligned}$$

In the same way, we then get

$$\begin{aligned}\frac{dJ}{dt} &= -\frac{1}{2c}\int \frac{1}{K(f)^{3/2}}\Big[\mathcal{U}'^2(f)f'^4 - 2\mathcal{U}\mathcal{U}'(f)f'^2f'' + \mathcal{U}^2(f)f''^2\Big]d\mu \\ &\qquad - \frac{1}{2}\int \frac{f'^2}{K(f)^{3/2}}\Big(\frac{W''}{c} - 1\Big)\Big[\mathcal{U}^2(f) + \frac{1}{c}\,f'^2\Big]d\mu.\end{aligned}$$

Since $W'' \geq c$, the conclusion follows. The proof of Theorem 1.1 is complete. $\square$

1.3 Some general facts about concentration

As we have seen in (1.6), one corollary of Gaussian isoperimetry is that whenever A is a Borel set in $\mathbb{R}^n$ with $\gamma(A) \geq \frac{1}{2}$ for the canonical Gaussian measure γ, then, for every $r \geq 0$,

$$\gamma(A_r) \geq 1 - \mathrm{e}^{-r^2/2}. \tag{1.16}$$

In other words, starting with a set with positive measure ($\frac{1}{2}$ here), its (Euclidean) enlargement or neighborhood gets very rapidly a mass close to one (think for example of $r = 5$ or 10). We described with (1.2) a similar property on spheres. While true isoperimetric inequalities are usually quite difficult to establish, in particular identification of extremal sets, concentration properties like (1.2) or (1.16) are milder, and may be established by a variety of arguments, as will be illustrated in these notes.

The concentration of measure phenomenon, put forward most vigorously by V. D. Milman in the local theory of Banach spaces (cf. [Mi], [M-S]), may be described for example on a metric space (X, d) equipped with a probability measure μ on the Borel sets of (X, d). One is then interested in the concentration function

$$\alpha(r) = \sup\big\{1 - \mu(A_r); \mu(A) \geq \tfrac{1}{2}\big\}, \quad r \geq 0,$$

where $A_r = \{x \in X; d(x, A) < r\}$. As a consequence of (1.16), $\alpha(r) \leq \mathrm{e}^{-r^2/2}$ in case of the canonical Gaussian measure γ on $\mathbb{R}^n$ with respect to the Euclidean metric. The important feature of this definition is that several measures, as we will see, do have very small concentration functions $\alpha(r)$ as r becomes "large". We will mainly be interested in Gaussian (or at least exponential) concentration functions throughout these notes. Besides Gaussian measures, Haar measures on spheres were part of the first examples (1.2). Martingale inequalities also yield family of examples (cf. [Mau1], [M-S], [Ta7]). In this work, we will encounter further examples, in particular in the context of product measures.

The concentration of measure phenomenon may also be described on functions. Let F be a Lipschitz map on X with $\|F\|_{\mathrm{Lip}} \leq 1$ (by homogeneity) and let m be a median of F for μ. Then, since $\mu(F \leq m) \geq \frac{1}{2}$, and $\{F \leq m\}_r \subset \{F \leq m + r\}$, we see that for every $r \geq 0$,

$$\mu(F \geq m + r) \leq \alpha(r). \tag{1.17}$$

When such an inequality holds, we will speak of a deviation inequality for F. Together with the same inequality for $-F$,

$$\mu\big(|F - m| \geq r\big) \leq 2\alpha(r). \tag{1.18}$$

We then speak of a concentration inequality for F. In particular, the Lipschitz map F concentrates around some fixed mean value m with a probability estimated by α. According to the smallness of α as r increases, F may be considered as almost constant on almost all the space. Note that these deviation or concentration inequalities on (Lipschitz) functions are actually equivalent to the corresponding statement on sets. Let A be a Borel set in (X, d) with $\mu(A) \geq \frac{1}{2}$. Set $F(x) = d(x, A)$ where $r > 0$. Clearly $\|F\|_{\mathrm{Lip}} \leq 1$ while

$$\mu(F > 0) = \mu\big(x; d(x, A) > 0\big) \leq 1 - \mu(A) \leq \tfrac{1}{2}\,.$$

Hence, 0 is a median of F and thus, by (1.17),

$$1 - \mu(A_r) \le \mu(F \ge r) \le \alpha(r). \tag{1.19}$$

In the Gaussian case, for every $r \ge 0$,

$$\gamma(F \ge m + r) \le \mathrm{e}^{-r^2/2} \tag{1.20}$$

when $\|F\|_{\mathrm{Lip}} \le 1$ and

$$\gamma(F \ge m + r) \le \mathrm{e}^{-r^2/2\|F\|_{\mathrm{Lip}}^2}$$

for arbitrary Lipschitz functions, extending thus the simple case of linear functions. These inequalities emphasize the two main parameters in a concentration property, namely some deviation or concentration value m, mean or median, and the Lipschitz coefficient $\|F\|_{\mathrm{Lip}}$ of F. An example of this type already occured in (1.8) which may be shown to follow equivalently from (1.20) (consider $F(x) = \max_{1 \le i \le n}(Mx)_i$). As a consequence of Theorem 1.1, if μ is a Boltzmann measure with $W''(x) \ge c\,\mathrm{Id}$ for every $x \in \mathbb{R}^n$, and if F is Lipschitz with $\|F\|_{\mathrm{Lip}} \le 1$, we get similarly that for every $r \ge 0$,

$$\mu(F \ge m + r) \le \mathrm{e}^{-r^2/2c}. \tag{1.21}$$

Although this last bound covers an interesting class of measures, it is clear that its application is fairly limited. It is therefore of interest to investigate new tools, other than isoperimetric inequalities, to derive concentration inequalities for large families of measures. This is the task of the next chapters.

It might be worthwhile to note that while we deduced the preceding concentration inequalities from isoperimetry, one may also adapt the semigroup arguments to give a direct, simpler, proof of these inequalities. To outline the argument in case of (1.20), let F on $\mathbb{R}^n$ be smooth and such that $\int F d\gamma = 0$ and $\|F\|_{\mathrm{Lip}} \le 1$. For fixed $\lambda \in \mathbb{R}$, set $H(t) = \int \mathrm{e}^{\lambda P_t F} d\gamma$ where $(P_t)_{t \ge 0}$ is the Ornstein-Uhlenbeck semigroup (1.12). Since $H(\infty) = 1$, we may write, for every $t \ge 0$,

$$\begin{aligned} H(t) &= 1 - \int_t^\infty H'(s) ds \\ &= 1 - \lambda \int_t^\infty \left(\int \mathrm{L} P_s F \mathrm{e}^{\lambda P_s F} d\gamma \right) ds \\ &= 1 + \frac{\lambda^2}{2} \int_t^\infty \left(\int |\nabla P_s F|^2 \mathrm{e}^{\lambda P_s F} d\gamma \right) ds \end{aligned}$$

by the integration by parts formula (1.13). Since $\|F\|_{\mathrm{Lip}} \le 1$, $|\nabla F| \le 1$ almost everywhere, so that

$$|\nabla P_s F|^2 = \big|\mathrm{e}^{-s/2} P_s(\nabla F)\big|^2 \le \mathrm{e}^{-s} P_s\big(|\nabla F|^2\big) \le \mathrm{e}^{-s}$$

almost everywhere. Hence, for $t \ge 0$,

$$H(t) \le 1 + \frac{\lambda^2}{2} \int_t^\infty \mathrm{e}^{-s} H(s) ds.$$

By Gronwall's lemma,

$$H(0) = \int \mathrm{e}^{\lambda F} d\gamma \leq \mathrm{e}^{\lambda^2/2}.$$

To deduce the deviation inequality (1.20) from this result, simply apply Chebyshev's inequality: for every $\lambda \in \mathbb{R}$ and $r \geq 0$,

$$\gamma(F \geq r) \leq \mathrm{e}^{-\lambda r + \lambda^2/2}.$$

Minimizing in λ ($\lambda = r$) yields

$$\gamma(F \geq r) \leq \mathrm{e}^{-r^2/2},$$

where we recall that F is smooth and such that $\int F d\gamma = 0$ and $\|F\|_{\mathrm{Lip}} \leq 1$. By a simple approximation procedure, we therefore get that, for every Lipschitz function F on $\mathbb{R}^n$ such that $\|F\|_{\mathrm{Lip}} \leq 1$ and all $r \geq 0$,

$$\gamma\big(F \geq \textstyle\int F d\gamma + r\big) \leq \mathrm{e}^{-r^2/2}. \tag{1.22}$$

The same argument would apply for the Boltzmann measures of Theorem 1.1 to produce (1.21) with the mean instead of a median. We note that this direct proof of (1.22) is shorter than the proof of the full isoperimetric inequality.

Inequality (1.22) may be used to investigate supremum of Gaussian processes as (1.7) or (1.20). As before, let $(X_t)_{t \in T}$ be a centered Gaussian process indexed by some countable set T, and assume that $\sup_{t \in T} X_t < \infty$ almost surely. Fix $t_1, \ldots, t_n$ and denote by $\Gamma = M\,{}^tM$ the covariance matrix of the centered Gaussian sample $(X_{t_1}, \ldots, X_{t_n})$. This sample thus has distribution Mx under $\gamma(dx)$. Let $F(x) = \max_{1 \leq i \leq n} (Mx)_i$, $x \in \mathbb{R}^n$. Then F is Lipschitz with

$$\|F\|_{\mathrm{Lip}} = \max_{1 \leq i \leq n} \big(\mathbb{E}(X_{t_i}^2)\big)^{1/2} \leq \sigma$$

where $\sigma = \sup_{t \in T} (\mathbb{E}(X_t^2))^{1/2}$. Therefore, by (1.22), for every $r \geq 0$,

$$\mathbb{P}\Big\{\max_{1 \leq i \leq n} X_{t_i} \geq \mathbb{E}\big(\max_{1 \leq i \leq n} X_{t_i}\big) + \sigma r\Big\} \leq \mathrm{e}^{-r^2/2}. \tag{1.23}$$

Similarly for $-F$,

$$\mathbb{P}\Big\{\max_{1 \leq i \leq n} X_{t_i} \leq \mathbb{E}\big(\max_{1 \leq i \leq n} X_{t_i}\big) - \sigma r\Big\} \leq \mathrm{e}^{-r^2/2}.$$

Choose now m such that $\mathbb{P}\{\sup_{t \in T} X_t \leq m\} \geq \frac{1}{2}$ and r_0 such that $\mathrm{e}^{-r_0^2/2} < \frac{1}{2}$. Then

$$\mathbb{P}\Big\{\max_{1 \leq i \leq n} X_{t_i} \leq m\Big\} \geq \tfrac{1}{2}\,.$$

Intersecting with the preceding probability, we get

$$\mathbb{E}\big(\max_{1 \leq i \leq n} X_{t_i}\big) \leq m + \sigma r_0$$

independently of $t_1, \dots, t_n$ in T. In particular, $\mathbb{E}(\sup_{t\in T} X_t) < \infty$, and by monotone convergence in (1.23),

$$\mathbb{P}\big\{\sup_{t\in T} X_t \geq \mathbb{E}\big(\sup_{t\in T} X_t\big) + \sigma r\big\} \leq \mathrm{e}^{-r^2/2}. \tag{1.24}$$

This inequality is the analogue of (1.8) with the mean instead of the median. Note that the condition $\mathbb{E}(\sup_{t\in T} X_t) < \infty$ came for free in the argument. It thus also implies (1.9) and (1.10).

This approximation argument may be used in the same way on infinite dimensional Gaussian measures γ with respect to their reproducing kernel Hilbert space $\mathcal{H}$. If F is Lipschitz with respect to $\mathcal{H}$ in the sense that

$$\big|F(x) - F(y)\big| \leq |x-y|_{\mathcal{H}},$$

then

$$\gamma(F \geq m + r) \leq \mathrm{e}^{-r^2/2} \tag{1.25}$$

for all $r \geq 0$ with m either the mean or a median of F for γ. See [Le3].

The inequalities (1.20) and (1.22) yield deviation inequalities for either a median or the mean of a Lipschitz function. Up to numerical constants, these are actually equivalent ([M-S], p. 142). One example was the inequalities (1.8) and (1.24) for supremum of Gaussian processes, and also (1.25). Let us describe the argument in some generality for exponential concentration functions. The argument clearly extends to sufficiently small concentration functions. (We will use this remark in the sequel.)

Let F be a measurable function on some probability space $(X, \mathcal{B}, \mu)$ such that, for some $0 < p < \infty$, some $a \in \mathbb{R}$ and some constants $c, d > 0$,

$$\mu\big(|F-a| \geq r\big) \leq 2c\,\mathrm{e}^{-r^p/d} \tag{1.26}$$

for all $r \geq 0$. Then, first of all,

$$\int |F-a| d\mu = \int_0^\infty \mu\big(|F-a| \geq r\big) dr \leq \int_0^\infty 2c\,\mathrm{e}^{-r^p/d} dr \leq C_p c d^{1/p}$$

where $C_p > 0$ only depends on p. In particular, $|\int F d\mu - a| \leq C_p c d^{1/p}$. Therefore, for $r \geq 0$,

$$\mu\big(F \geq \textstyle\int F d\mu + r\big) \leq \mu(F \geq a - C_p c d^{1/p} + r).$$

According as $r \leq 2C_p c d^{1/p}$ or $r \geq 2C_p c d^{1/p}$ we easily get that

$$\mu\big(F \geq \textstyle\int F d\mu + r\big) \leq c' \mathrm{e}^{-r^p/d'}$$

where $c' = \max(2c, \mathrm{e}^{C_p^p c^p})$ and $d' = 2^p d$. Together with the same inequality for $-F$, (1.26) thus holds with a the mean of F (and c' and d'). Similary, if we choose in (1.26) $r = r_0$ so that

$$2c\,\mathrm{e}^{-r^p/d} < \tfrac{1}{2}\,,$$

for example $r_0^p = d\log(8c)$, we see that $\mu(|F-a| \geq r_0) < \frac{1}{2}$. Therefore a median m of F for μ will satisfy

$$a - r_0 \leq m \leq a + r_0.$$

It is then easy to conclude as previously that, for every $r \geq 0$,

$$\mu(F \geq m + r) \leq c' \mathrm{e}^{-r^p/d'}$$

where $c' = 8c$ and $d' = 2^p d$. We can therefore also choose for a in (1.26) a median of F.

An alternate argument may be given on the concentration function. For a probability measure μ on the Borel sets of a metric space (X, d), assume that for some non-increasing function α on $\mathbb{R}_+$,

$$\mu\big(F \geq \mathrm{E}_\mu(F) + r\big) \leq \alpha(r) \tag{1.27}$$

for every F with $\|F\|_{\mathrm{Lip}} \leq 1$ and every $r \geq 0$. Let A with $\mu(A) > 0$ and fix $r > 0$. Set $F(x) = \min(d(x,A), r)$. Clearly $\|F\|_{\mathrm{Lip}} \leq 1$ and

$$\mathrm{E}_\mu(F) \leq \big(1 - \mu(A)\big) r.$$

Applying (1.27),

$$1 - \mu(A_r) = \mu(F \geq r) \leq \mu\big(F \geq \mathrm{E}_\mu(F) + \mu(A)r\big) \leq \alpha\big(\mu(A)r\big). \tag{1.28}$$

In particular, if $\mu(A) \geq \frac{1}{2}$,

$$\mu(A_r) \geq 1 - \alpha\big(\tfrac{r}{2}\big).$$

We conclude this section by emphasizing that a concentration inequality of such as (1.26) of course implies strong integrability properties of the Lipschitz function F. This is the content of the simple proposition which immediately follows by integration in $r \geq 0$.

Proposition 1.2. *Let F be a measurable function on $(X, \mathcal{B}, \mu)$ such that for some $0 < p < \infty$, some $a \in \mathbb{R}$ and some constants $c, d > 0$,*

$$\mu\big(|F - a| \geq r\big) \leq 2c\,\mathrm{e}^{-r^p/d}$$

for every $r \geq 0$. Then

$$\int \mathrm{e}^{\alpha |F|^p} d\mu < \infty$$

for every $\alpha < \frac{1}{d}$.

Proof. From the hypothesis, for every $r \geq |a|$,

$$\mu\big(|F| \geq r\big) \leq \mu\big(|F - a| \geq r - |a|\big) \leq 2c\,\mathrm{e}^{-(r-|a|)^p/d}.$$

Now, by Fubini's theorem,

$$\begin{aligned}
\int \mathrm{e}^{\alpha|F|^p} d\mu &= 1 + \int_0^\infty p\,\alpha\, r^{p-1} \mu\big(|F| \geq r\big) \mathrm{e}^{\alpha r^p} dr \\
&\leq \mathrm{e}^{\alpha|a|^p} + \int_{|a|}^\infty p\,\alpha\, r^{p-1} \mu\big(|F| \geq r\big) \mathrm{e}^{\alpha r^p} dr \\
&\leq \mathrm{e}^{\alpha|a|^p} + \int_{|a|}^\infty p\,\alpha\, r^{p-1} 2c\,\mathrm{e}^{-(r-|a|)^p/d} \mathrm{e}^{\alpha r^p} dr
\end{aligned}$$

from which the conclusion follows. □

2. SPECTRAL GAP AND LOGARITHMIC SOBOLEV INEQUALITIES

We present in this section the basic simple argument that produces Gaussian concentration under a logarithmic Sobolev inequality. We try to deal with a rather general framework in order to include several variations developed in the literature. Herbst's original argument, mentioned in [D-S], has been revived recently by S. Aida, T. Masuda and I. Shigekawa [A-M-S]. Since then, related papers by S. Aida and D. Stroock [A-S], S. Bobkov and F. Götze [B-G], L. Gross and O. Rothaus [G-R], O. Rothaus [Ro3] and the author [Le1] further developed the methods and results. Most of the results presented in these notes are taken from these works. We will mainly be concerned with Herbst's original differential argument on the Laplace transform. The papers [A-S], [Ro3] and [G-R] also deal with moment growth.

We present in the first paragraph a general setting dealing with logarithmic Sobolev and Poincaré inequalities. We then turn to Herbst's basic argument which yields Gaussian concentration under a logarithmic Sobolev inequality. We discuss next more general entropy-energy inequalities and exponential integrability under spectral gap inequalities.

2.1 Abstract functional inequalities

In order to develop the functional approach to concentration, we need to introduce a convenient setting in which most of the known results may be considered. We will go from a rather abstract and informal framework to more concrete cases and examples.

Let $(X, \mathcal{B}, \mu)$ be a probability space. We denote by E_μ integration with respect to μ, and by $(\mathrm{L}^p(\mu), \|\cdot\|_\infty)$ the Lebesgue spaces over $(X, \mathcal{B}, \mu)$. For any function f in $\mathrm{L}^2(\mu)$, we further denote by

$$\mathrm{Var}_\mu(f) = \mathrm{E}_\mu(f^2) - \big(\mathrm{E}_\mu(f)\big)^2$$

the variance of f. If f is a non-negative function on E such that $\mathrm{E}_\mu(f \log^+ f) < \infty$, we introduce the entropy of f with respect to μ as

$$\mathrm{Ent}_\mu(f) = \mathrm{E}_\mu(f \log f) - \mathrm{E}_\mu(f) \log \mathrm{E}_\mu(f).$$

(Actually, since the function $x \log x$ is bounded below, $\mathrm{Ent}_\mu(f) < \infty$ if and only if $\mathrm{E}_\mu(f \log^+ f) < \infty$.) Note that $\mathrm{Ent}_\mu(f) \geq 0$ and that $\mathrm{Ent}_\mu(\alpha f) = \alpha \mathrm{Ent}_\mu(f)$ for $\alpha \geq 0$. We write E, Var, Ent when there is no confusion with respect to the measure.

On some subset $\mathcal{A}$ of measurable functions f on X, consider now a map, or energy, $\mathcal{E} : \mathcal{A} \to \mathbb{R}_+$. We say that μ satisfies a spectral gap or Poincaré inequality with respect to $\mathcal{E}$ (on $\mathcal{A}$) if there exists $C > 0$ such that

$$\mathrm{Var}_\mu(f) \leq C\mathcal{E}(f) \tag{2.1}$$

for every function $f \in \mathcal{A}$ in $\mathrm{L}^2(\mu)$. We say that μ satisfies a logarithmic Sobolev inequality with respect to $\mathcal{E}$ (on $\mathcal{A}$) if there exists $C > 0$ such that

$$\mathrm{Ent}_\mu(f^2) \leq 2C\mathcal{E}(f) \tag{2.2}$$

for every function $f \in \mathcal{A}$ with $\mathrm{E}_\mu(f^2 \log^+ f^2) < \infty$. (The choice of the normalization in (2.2) will become clear with Proposition 2.1 below.) By extension, the integrability properties on f will be understood when speaking of inequalities (2.1) and (2.2) for all f in $\mathcal{A}$.

These abstract definitions include a number of cases of interest. For example, if (X, d) is a metric space equipped with its Borel σ-field $\mathcal{B}$, one may consider the natural generalization of the modulus of the usual gradient

$$|\nabla f(x)| = \limsup_{d(x,y)\to 0} \frac{|f(x) - f(y)|}{d(x,y)} \tag{2.3}$$

(with $|\nabla f(x)| = 0$ for isolated points x in X). In this case, one may define, for a probability measure μ on $(X, \mathcal{B})$,

$$\mathcal{E}(f) = \mathrm{E}_\mu(|\nabla f|^2) \tag{2.4}$$

on the class $\mathcal{A}$ of all, say, (bounded) Lipschitz functions on X. One important feature of this situation is that ∇ is a derivation in the sense that for a C^∞ function ψ on $\mathbb{R}$, and $f \in \mathcal{A}$, $\psi(f) \in \mathcal{A}$ and

$$\big|\nabla(\psi(f))\big| = |\nabla f|\big|\psi'(f)\big|. \tag{2.5}$$

In particular,

$$\mathcal{E}\big(\psi(f)\big) \leq \|\nabla f\|_\infty^2 \mathrm{E}_\mu\big(\psi'(f)^2\big). \tag{2.6}$$

For example,

$$\mathcal{E}(\mathrm{e}^{f/2}) \leq \frac{1}{4}\,\|\nabla f\|_\infty^2 \mathrm{E}_\mu(\mathrm{e}^f).$$

Another setting of interest, following [A-S] and [G-R], consists of the gradients and Dirichlet forms associated to (symmetric) Markov semigroups. On a probability space $(X, \mathcal{B}, \mu)$, let $p_t(x, \cdot)$ be a Markov transition probability function on $(X, \mathcal{B})$. Assume that $p_t(x, dy)\mu(dx)$ is symmetric in x and y and that, for each bounded measurable function f on X,

$$P_t f(x) = \int f(y) p_t(x, dy)$$

converges to f in $\mathrm{L}^2(\mu)$ as t goes to 0. Denote also by P_t the unique bounded extension of P_t to $\mathrm{L}^2(\mu)$. Then $(P_t)_{t\geq 0}$ defines a strongly continuous semigroup on $\mathrm{L}^2(\mu)$ with Dirichlet form the quadratic form

$$\mathcal{E}(f,f) = \lim_{t\to 0}\frac{1}{2t}\int\!\!\int \big(f(x)-f(y)\big)^2 p_t(x,dy)\mu(dx). \tag{2.7}$$

Let $\mathcal{D}(\mathcal{E})$ be the domain of $\mathcal{E}$ (the space of $f\in\mathrm{L}^2(\mu)$ for which $\mathcal{E}(f,f)<\infty$). On the algebra $\mathcal{A}$ of bounded measurable functions f of $\mathcal{D}(\mathcal{E})$, one may then consider $\mathcal{E}(f)=\mathcal{E}(f,f)$. This energy functional does not necessarily satisfy a chain rule formula of the type of (2.6). However, as was emphasized in [A-S], we still have that, for every f in $\mathcal{A}$,

$$\mathcal{E}(\mathrm{e}^{f/2}) \leq \frac{1}{2}\,|||f|||_\infty^2 \mathrm{E}_\mu(\mathrm{e}^f). \tag{2.8}$$

Here

$$|||f|||_\infty^2 = \sup\big\{\mathcal{E}(gf,f) - \tfrac{1}{2}\,\mathcal{E}(g,f^2); g\in\mathcal{A}, \|g\|_1\leq 1\big\}$$

that may be considered as a generalized norm of a gradient. To establish (2.8), note that, by symmetry,

$$\begin{aligned}\int\!\!\int &\big(\mathrm{e}^{f(x)/2}-\mathrm{e}^{f(y)/2}\big)^2 p_t(x,dy)\mu(dx)\\ &= 2\int\!\!\int_{\{f(x)<f(y)\}} \big(\mathrm{e}^{f(x)/2}-\mathrm{e}^{f(y)/2}\big)^2 p_t(x,dy)\mu(dx)\\ &\leq \frac{1}{2}\int\!\!\int \big(f(x)-f(y)\big)^2 \mathrm{e}^{f(y)} p_t(x,dy)\mu(dx).\end{aligned}$$

Now, for every g in $\mathcal{A}$,

$$\lim_{t\to 0}\frac{1}{2t}\int\!\!\int g(x)\big(f(x)-f(y)\big)^2 p_t(x,dy)\mu(dx) = \mathcal{E}(gf,f) - \tfrac{1}{2}\,\mathcal{E}(g,f^2)$$

from which (2.8) follows.

Examples fitting this general framework are numerous. Let $X=\mathbb{R}^n$ and write ∇f for the usual gradient of a smooth function f on $\mathbb{R}^n$. Let

$$M:\mathbb{R}^n\to \text{invertible matrices } \ \{n\times n\}$$

be measurable and locally bounded and let $d\mu(x)=w(x)dx$ be a probability measure on $\mathbb{R}^n$ with $w>0$. For every C^∞ compactly supported function f on $\mathbb{R}^n$, set

$$\mathcal{E}(f,f) = \int_{\mathbb{R}^n} \langle M(x)\nabla f(x), M(x)\nabla f(x)\rangle d\mu(x).$$

We need not be really concerned here with the semigroup induced by this Dirichlet form. Ignoring questions on the closure of $\mathcal{E}$, it readily follows that in this case

$$|||f|||_\infty = \sup\big\{\big|M(x)\nabla f(x)\big|; x\in\mathbb{R}^n\big\}$$

where $|\cdot|$ is Euclidean length. More generally, if μ is a probability measure on a Riemannian manifold X, and if $\mathcal{E}(f,f) = \int_M |\nabla f|^2 d\mu$, then one has $|||f|||_\infty = \|\nabla f\|_\infty$.

With this class of examples, we of course rejoin the generalized moduli of gradients (2.3). In this case, the Dirichlet form $\mathcal{E}$ is actually local, that is, it satisfies the chain rule formula (2.6). In particular, (2.8) holds in this case with constant $\frac{1}{4}$ (and $|||f|||_\infty = \|\nabla f\|_\infty$). We freely use this observation throughout these notes.

Covering in another way the two preceding settings, one may also consider the abstract Markov semigroup framework of [Ba1] in which, given a Markov generator L on some nice algebra $\mathcal{A}$ of functions, one defines the carré du champ operator as

$$\Gamma(f,g) = \tfrac{1}{2}\big(\mathrm{L}(fg) - f\mathrm{L}g - g\mathrm{L}f\big).$$

For example, if L is the Laplace-Beltrami operator on a manifold M, then $\Gamma(f,g) = \nabla f \cdot \nabla g$. One may then define

$$\mathcal{E}(f) = \mathrm{E}_\mu\big(\Gamma(f,f)\big)$$

on the class $\mathcal{A}$. If L is symmetric, one shows that $|||f|||_\infty = \|\Gamma(f,f)\|_\infty$. Provided L is a diffusion (that is, it satisfies the change of variables formula $\mathrm{L}\psi(f) = \psi'(f)\mathrm{L}f + \psi''(f)\Gamma(f,f)$) $\mathcal{E}$ will satisfy (2.6). A further discussion may be found in [Ba1].

We turn to discrete examples. Let X be a finite or countable set. Let $K(x,y) \geq 0$ satisfy

$$\sum_{y\in X} K(x,y) = 1$$

for every $x \in X$. Asssume furthermore that there is a symmetric invariant probability measure μ on X, that is $K(x,y)\mu(x)$ is symmetric in x and y and $\sum_x K(x,y)\mu(x) = \mu(y)$ for every $y \in X$. In other words, (K,μ) is a symmetric Markov chain. Define

$$\mathcal{E}(f,f) = \frac{1}{2}\sum_{x,y\in X}\big(f(x) - f(y)\big)^2 K(x,y)\mu\big(\{x\}\big).$$

In this case,

$$|||f|||_\infty^2 = \frac{1}{2}\sup_{x\in X}\sum_{y\in X}\big(f(x) - f(y)\big)^2 K(x,y).$$

It might be worthwhile noting that if we let

$$\|\nabla f\|_\infty = \sup\big\{\big|f(x) - f(y)\big|; K(x,y) > 0\big\},$$

then, since $\sum_y K(x,y) = 1$,

$$|||f|||_\infty^2 \leq \frac{1}{2}\,\|\nabla f\|_\infty^2.$$

It should be clear that the definition of the $|||\cdot|||_\infty$-norm tries to be as close as possible to the sup-norm of a gradient in a continuous setting. As such however, it does not always reflect accurately discrete situations. Discrete gradients may actually be examined in another way. If f is a function on $\mathbb{Z}$, set

$$Df(x) = f(x+1) - f(x), \quad x \in \mathbb{Z}. \tag{2.9}$$

One may then consider

$$\mathcal{E}(f) = \mathrm{E}_\mu(|Df|^2) \tag{2.10}$$

for a measure μ on $\mathbb{Z}$. This energy will not satisfy (2.6) but satisfies (2.8). For reals $m(x)$, $x \in \mathbb{Z}$, let

$$\mathcal{E}(f,f) = \sum_{x\in\mathbb{Z}} Df(x)^2 m(x)^2 \mu(\{x\}).$$

One can check that for this Dirichlet form

$$|||f|||_\infty^2 = \sup_{x\in\mathbb{Z}} \frac{1}{2}\bigg(m(x)^2 Df(x)^2 + m(x-1)^2 \,\frac{\mu(\{x-1\})}{\mu(\{x\})}\, Df(x-1)^2 \bigg). \tag{2.11}$$

As will be seen in Part 5, this uniform norm of the gradient is actually of little use in specific examples, such as Poisson measures. It will be more fruitful to consider $\sup_{x\in\mathbb{Z}^d}|Df(x)|$. The lack of chain rule (for example, $|D(\mathrm{e}^f)| \leq |Df|\mathrm{e}^{|Df|}\mathrm{e}^f$ only in general) will then have to be handled by other means. The norm $|||\cdot|||_\infty$ is in fact only well adapted to produce Gaussian bounds as we will see in Section 2.3. It is actually defined in such a way to produce results similar to those which follows from a chain rule formula. As such, this norm is not suited to a number of discrete examples (see also [G-R]).

The preceding example may be further generalized to $\mathbb{Z}^d$. Similarly, in the context of statistical mechanics, set $X = \{-1,+1\}^{\mathbb{Z}^d}$ and let

$$\big|Df(\omega)\big| = \bigg(\sum_{k\in\mathbb{Z}^d} \big|\partial_k f(\omega)\big|^2\bigg)^{1/2} \tag{2.12}$$

where $\partial_k f(\omega) = f(\omega^k) - f(\omega)$ where ω^k is the element of X obtained from ω by replacing the k-th coordinate with $-\omega_k$.

Logarithmic Sobolev inequalities were introduced to describe smoothing properties of Markov semigroups, especially in infinite dimensional settings. The key argument was isolated by L. Gross [Gr1] who showed how a logarithmic Sobolev inequality is actually equivalent to hypercontractivity of a Markov generator. Precisely, if $(P_t)_{t\geq 0}$ is a symmetric Markov semigroup with invariant measure μ and Dirichlet form $\mathcal{E}$, then the logarithmic Sobolev inequality

$$\mathrm{Ent}(f^2) \leq 2C\,\mathcal{E}(f,f), \quad f \in \mathcal{A},$$

is equivalent to saying that, whenever $1 < p < q < \infty$ and $t > 0$ are such that $\mathrm{e}^{2t/C} \geq (q-1)/(p-1)$, we have

$$\|P_t f\|_q \leq \|f\|_p$$

for every $f \in \mathcal{A}$ in $\mathrm{L}^p(\mu)$ (cf. [Gr1], [Ba1] for the precise statement). Hypercontractivity is an important tool in deriving sharp estimates on the time to equilibrium of P_t [S-Z], [St], [D-S] etc.

Now, we mention a simple comparison between spectral and logarithmic Sobolev inequalities. The hypothesis on $\mathcal{E}$ is straightforward in all the previous examples.

Proposition 2.1. *Assume that μ satisfies the logarithmic Sobolev inequality*

$$\mathrm{Ent}_\mu(f^2) \le 2\,\mathcal{E}(f), \quad f \in \mathcal{A},$$

and that $af + b \in \mathcal{A}$ and $\mathcal{E}(af+b) = a^2\mathcal{E}(f)$ for every $f \in \mathcal{A}$ and $a, b \in \mathbb{R}$. Then μ satisfies the spectral gap inequality

$$\mathrm{Var}_\mu(f^2) \le \mathcal{E}(f), \quad f \in \mathcal{A}.$$

Proof. Fix f with $\mathrm{E}_\mu(f) = 0$ and $\mathrm{E}_\mu(f^2) = 1$ and apply the logarithmic Sobolev inequality to $1 + \varepsilon f$. As ε goes to 0, a Taylor expansion of $\log(1+\varepsilon f)$ yields the conclusion. □

It might be worthwhile mentioning that the converse to Proposition 2.1 is not true in general, even within constants. We will have the opportunity to encounter a number of such cases throughout these notes (cf. Sections 4.1, 5.1 and 7.3).

One important feature of both variance and entropy is their product property. Assume we are given probability spaces $(X_i, \mathcal{B}_i, \mu_i)$, $1 \le i \le n$. Denote by P the product probability measure $P = \mu_1 \otimes \cdots \otimes \mu_n$ on the product space $X = X_1 \times \cdots \times X_n$ equipped with the product σ-field $\mathcal{B}$. Given f on the product space, we write furthermore f_i, $1 \le i \le n$, for the function on X_i defined by

$$f_i(x_i) = f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n),$$

with $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ fixed.

Proposition 2.2. *Under appropriate integrability conditions,*

$$\mathrm{Var}_P(f) \le \sum_{i=1}^n \mathrm{E}_P\big(\mathrm{Var}_{\mu_i}(f_i)\big)$$

and

$$\mathrm{Ent}_P(f) \le \sum_{i=1}^n \mathrm{E}_P\big(\mathrm{Ent}_{\mu_i}(f_i)\big).$$

Proof. Let us prove the assertion concerning entropy, the one for variance being (simpler and) similar. Recall first that for a non-negative function f on $(X, \mathcal{B}, \mu)$,

$$\mathrm{Ent}_\mu(f) = \sup\big\{\mathrm{E}_\mu(fg); \mathrm{E}_\mu(\mathrm{e}^g) \le 1\big\}. \tag{2.13}$$

Indeed, assume by homogeneity that $\mathrm{E}_\mu(f) = 1$. By Young's inequality

$$uv \le u\log u - u + \mathrm{e}^v, \quad u \ge 0, \quad v \in \mathbb{R},$$

we get, for $\mathrm{E}_\mu(\mathrm{e}^g) \leq 1$,

$$\mathrm{E}_\mu(fg) \leq \mathrm{E}_\mu(f\log f) - 1 + \mathrm{E}_\mu(\mathrm{e}^g) \leq \mathrm{E}_\mu(f\log f).$$

The converse is obvious.

To prove Proposition 2.2, given g on $(X, \mathcal{B}, P)$ such that $\mathrm{E}_P(\mathrm{e}^g) \leq 1$, set, for every $i = 1\ldots, n$,

$$g^i(x_i, \ldots, x_n) = \log\left(\frac{\int \mathrm{e}^{g(x_1,\ldots,x_n)} d\mu_1(x_1)\cdots d\mu_{i-1}(x_{i-1})}{\int \mathrm{e}^{g(x_1,\ldots,x_n)} d\mu_1(x_1)\cdots d\mu_i(x_i)}\right).$$

Then $g \leq \sum_{i=1}^n g^i$ and $\mathrm{E}_{\mu_i}(\mathrm{e}^{(g^i)_i}) = 1$. Therefore,

$$\mathrm{E}_P(fg) \leq \sum_{i=1}^n \mathrm{E}_P(fg^i) = \sum_{i=1}^n \mathrm{E}_P\big(\mathrm{E}_{\mu_i}\big(f_i(g^i)_i\big)\big) \leq \sum_{i=1}^n \mathrm{E}_P\big(\mathrm{Ent}_{\mu_i}(f_i)\big)$$

which is the result. Proposition 2.2 is established. □

What Proposition 2.2 will tell us in applications is that, whenever the energy on the product space is the sum of the energies on each coordinates, in order to establish a Poincaré or logarithmic Sobolev inequality in product spaces, it will be enough to deal with the dimension one. In particular, these inequalities will be independent of the dimension of the product space. This is why logarithmic Sobolev inequalities are such powerful tools in infinite dimensional analysis.

2.2 Examples of logarithmic Sobolev inequalities

The first examples of logarithmic Sobolev inequalities were discovered by L. Gross in 1975 [Gr1]. They concerned the two-point space and the canonical Gaussian measure. For the two point space $\{0,1\}$ with uniform (Bernoulli) measure $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, L. Gross showed that for every f on $\{0,1\}$,

$$\mathrm{Ent}_\mu(f^2) \leq \frac{1}{2}\,\mathrm{E}_\mu\big(|Df|^2\big) \tag{2.14}$$

where $Df(x) = f(1) - f(0)$, $x \in \{0,1\}$. The constant is optimal. In its equivalent hypercontractive form, this inequality actually goes back to A. Bonami [Bon]. Due to Proposition 2.2, if μ^n is the n-fold product measure of μ on $\{0,1\}^n$, for every f on $\{0,1\}^n$,

$$\mathrm{Ent}_{\mu^n}(f^2) \leq \frac{1}{2}\,\mathrm{E}_{\mu^n}\left(\sum_{i=1}^n |D_i f|^2\right)$$

where, for $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ and $i = 1, \ldots, n$, $D_i f(x) = Df_i(x_i)$. Applying this inequality to

$$f(x_1, \ldots, x_n) = \varphi\Big(\frac{x_1 + \cdots + x_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}}\Big)$$

for some smooth φ on $\mathbb{R}$, L. Gross deduced, with the classical central limit theorem, a logarithmic Sobolev inequality for the canonical Gaussian measure γ on $\mathbb{R}$ in the form of

$$\mathrm{Ent}_\gamma(\varphi^2) \le 2\,\mathrm{E}_\gamma\big({\varphi'}^2\big).$$

By the product property of entropy, if γ is the canonical Gaussian measure on $\mathbb{R}^n$, for every f on $\mathbb{R}^n$ with gradient in $\mathrm{L}^2(\gamma)$,

$$\mathrm{Ent}_\gamma(f^2) \le 2\,\mathrm{E}_\gamma\big(|\nabla f|^2\big). \tag{2.15}$$

Inequality (2.15) may be considered as the prototype of logarithmic Sobolev inequalities. The constant in (2.15) is optimal as can be checked for example on exponential functions $\mathrm{e}^{\lambda x}$, which actually saturate this inequality. This observation is a first indication on the Laplace transform approach we will develop next. Several simple, alternative proofs of this inequality have been developed in the literature. For our purposes, it might be worthwhile noting that it may be seen as consequence of the Gaussian isoperimetric inequality itself. This has been noticed first in [Le1] but recently, W. Beckner [Be] kindly communicated to the author a simple direct argument on the basis of the functional inequality (1.11). Namely, let g be smooth with $\int g^2 d\gamma = 1$ and apply (1.11) to $f = \varepsilon g^2$ with $\varepsilon \to 0$. We get that

$$1 \le \int \sqrt{\frac{\mathcal{U}^2(\varepsilon g^2)}{\mathcal{U}^2(\varepsilon)} + \frac{4\varepsilon^2}{\mathcal{U}^2(\varepsilon)}\, g^2 |\nabla g|^2 d\gamma}.$$

Noticing that $\mathcal{U}^2(\varepsilon) \sim \varepsilon^2 \log(\frac{1}{\varepsilon^2})$ as $\varepsilon \to 0$, we see that

$$1 \le \int g^2 \sqrt{1 - \frac{1}{M}\,\log g^2 + \frac{2}{M}\,\frac{|\nabla g|^2}{g^2} + o\Big(\frac{1}{M}\Big)}\, d\gamma$$

where $M = M(\varepsilon) = \log(\frac{1}{\varepsilon}) \to \infty$ as $\varepsilon \to 0$. Hence

$$1 \le \int g^2 \Big(1 - \frac{1}{2M}\,\log g^2 + \frac{1}{M}\,\frac{|\nabla g|^2}{g^2}\Big) d\gamma + o\Big(\frac{1}{M}\Big)$$

from which the Gaussian logartihmic Sobolev inequality (2.15) follows. The same argument works for the Boltzmann measures of Theorem 1.1. On the other hand, the semigroup arguments leading to the Gaussian isoperimetric inequality may also be adapted to give a direct, simpler proof of the logarithmic Sobolev inequality (2.15) [Ba1], [Le3]. To briefly sketch the argument (following the same notation), let f be smooth and non-negative on $\mathbb{R}$. Then write

$$\mathrm{Ent}_\gamma(f) = -\int_0^\infty \frac{d}{dt}\,\mathrm{E}_\gamma(P_t f \log P_t f)dt$$

(with $(P_t)_{t\ge 0}$ the Ornstein-Uhlenbeck semigroup (1.12)). By the chain rule formula,

$$\frac{d}{dt}\,\mathrm{E}_\gamma(P_t f \log P_t f) = \mathrm{E}_\gamma(\mathrm{L}P_t f \log P_t f) + \mathrm{E}_\gamma(\mathrm{L}P_t f) = -\frac{1}{2}\,\mathrm{E}_\gamma\Big(\frac{(P_t f)'^2}{P_t f}\Big)$$

since γ is invariant under the action of P_t and thus $\mathrm{E}_\gamma(\mathrm{L}P_t f) = 0$. Now, $(P_t f)' = \mathrm{e}^{-t/2} P_t f'$ so that, by the Cauchy-Schwarz inequality for P_t,

$$(P_t f')^2 \le P_t f \, P_t\Big(\frac{f'^2}{f}\Big).$$

Summarizing,

$$\mathrm{Ent}_\gamma(f) \le \frac{1}{2}\int_0^\infty \mathrm{e}^{-t}\,\mathrm{E}_\gamma\bigg(P_t\Big(\frac{f'^2}{f}\Big)\bigg)dt = \frac{1}{2}\,\mathrm{E}_\gamma\Big(\frac{f'^2}{f}\Big)$$

which, by the change of f into f^2, is (2.15) in dimension one.

The preceding proof may be shown to imply in the same way the Poincaré inequality for Gaussian measures

$$\mathrm{Var}_\gamma(f) \le \mathrm{E}_\gamma\big(|\nabla f|^2\big). \tag{2.16}$$

(Write, in dimension one for simplicity,

$$\begin{aligned}\mathrm{Var}_\gamma(f) = -\int_0^\infty \frac{d}{dt}\,\mathrm{E}_\gamma\big((P_t f)^2\big)dt &= -2\int_0^\infty \mathrm{E}(P_t f \mathrm{L} P_t f)dt\\ &= \int_0^\infty \mathrm{E}_\gamma\big({(P_t f)'}^2\big)dt\\ &= \int_0^\infty \mathrm{e}^{-t}\mathrm{E}_\gamma\big((P_t f')^2\big)dt\\ &\le \int_0^\infty \mathrm{e}^{-t}\mathrm{E}_\gamma\big(f'^2\big)dt = \mathrm{E}_\gamma\big(f'^2\big).)\end{aligned}$$

It may also be seen as a consequence of the logarithmic Sobolev inequality (2.15) by Proposition 2.1. Actually, (2.16) is a straigthforward consequence of a series expansion in Hermite polynomials, and may be found, in this form, in the physics literature of the thirties.

Both the (dimension free) logarithmic Sobolev and Poincaré inequalities (2.15) and (2.16) extend to infinite dimensional Gaussian measures replacing the gradient by the Gross-Malliavin derivatives along the directions of the reproducing kernel Hilbert space. This is easily seen by a finite dimensional approximation (cf. [Le3]).

The preceding semigroup proofs also apply to Boltzmann measures as studied in Section 1.2. In particular, under the curvature assumption of Theorem 1.1, these measures satisfy the logarithmic Sobolev inequality

$$\mathrm{Ent}_\mu(f^2) \le \frac{2}{c}\,\mathrm{E}_\mu\big(|\nabla f|^2\big). \tag{2.17}$$

As we have seen, this inequality may also be shown to follow from Theorem 1.1 (cf. also [Ba-L]). We discuss in Section 7.1 logarithmic Sobolev inequalities for a more general class of potentials.

Further logarithmic Sobolev inequalities have been established and studied throughout the literature, mainly for their hypercontractive content. We refer to the survey [Gr2] for more information. We investigate here logarithmic Sobolev inequalities for their applications to the concentration of measure phenomenon.

2.3 The Herbst argument

In this section, we illustrate how concentration properties may follow from a logarithmic Sobolev inequality. Although rather elementary, this observation is a powerful scheme which allows us to establish some new concentration inequalities. Indeed, as illustrated in particular in the next chapter, convexity of entropy allows to tensorize one-dimensional inequalities to produce concentration properties in product spaces whereas concentration itself does not usually tensorize.

To clarify the further developments, we first present Herbst's argument (or what we believe Herbst's argument was) in the original simple case. Let thus μ be a probability measure on $\mathbb{R}^n$ such that for some $C > 0$ and all smooth f on $\mathbb{R}^n$,

$$\mathrm{Ent}(f^2) \leq 2C\,\mathrm{E}\big(|\nabla f|^2\big) \tag{2.18}$$

(where ∇f is the usual gradient of f). Let now F be smooth (and bounded) such that $\|F\|_{\mathrm{Lip}} \leq 1$. In particular, since we assume F to be regular enough, we can have that $|\nabla F| \leq 1$ at every point. Apply now (2.18) to $f^2 = \mathrm{e}^{\lambda F}$ for every $\lambda \in \mathbb{R}$. We have

$$\mathrm{E}\big(|\nabla f|^2\big) = \frac{\lambda^2}{4}\,\mathrm{E}\big(|\nabla F|^2 \mathrm{e}^{\lambda F}\big) \leq \frac{\lambda^2}{4}\,\mathrm{E}(\mathrm{e}^{\lambda F}).$$

Setting $H(\lambda) = \mathrm{E}_\mu(\mathrm{e}^{\lambda F})$, $\lambda \in \mathbb{R}$, we get by the definition of entropy,

$$\lambda H'(\lambda) - H(\lambda)\log H(\lambda) \leq \frac{C\lambda^2}{2}\,H(\lambda).$$

In other words, if $K(\lambda) = \frac{1}{\lambda}\log H(\lambda)$ (with $K(0) = H'(0)/H(0) = \mathrm{E}_\mu(F)$),

$$K'(\lambda) \leq \frac{C}{2}$$

for every λ. Therefore,

$$K(\lambda) = K(0) + \int_0^\lambda K'(u)du \leq \mathrm{E}_\mu(F) + \frac{C\lambda}{2}$$

and hence, for every λ,

$$H(\lambda) = \mathrm{E}_\mu(\mathrm{e}^{\lambda F}) \leq \mathrm{e}^{\lambda \mathrm{E}_\mu(F) + C\lambda^2/2}. \tag{2.19}$$

Replacing F by a smooth convolution, (2.19) extends to all Lipschitz functions with $\|F\|_{\mathrm{Lip}} \leq 1$ (see below). By Chebyshev's inequality, for every $\lambda, r \geq 0$,

$$\mu\big(F \geq \mathrm{E}_\mu(F) + r\big) \leq \mathrm{e}^{-\lambda r + Cr^2/2}$$

and optimizing in λ, for every $r \geq 0$,

$$\mu\big(F \geq \mathrm{E}_\mu(F) + r\big) \leq \mathrm{e}^{-r^2/2C}.$$

The same inequality holds for $-F$.

The next proposition is some abstract formulation on the preceding argument. It aims to cover several situations at once so that it may look akward at first. The subsequent results will take a simpler form. At this point, they all yield Gaussian concentration under logarithmic Sobolev inequalities. In the next section, we study non-Gaussian tails which arise from more general entropy-energy inequalities, or from the lack of chain rule for discrete gradients (cf. Part 5).

Let $(X, \mathcal{B}, \mu)$ be a probability space. We write E for E_μ, and similarly Var, Ent. Let $\mathcal{A}$ be a subset of $\mathrm{L}^1(\mu)$. For every f in $\mathcal{A}$, let $N(f) \geq 0$. Typically $N(f)$ will be our Lipschitz norm or generalized sup-norm of the gradient. For example, $N(f) = \|\nabla f\|_\infty$ in (2.4), or $|||f|||_\infty$ in (2.8), or $\sup_{x \in \mathbb{Z}} |Df(x)|$ in (2.10).

Proposition 2.3. *Let $\mathcal{A}$ and N be such that, for every $f \in \mathcal{A}$ and $\lambda \in \mathbb{R}$, $\lambda f \in \mathcal{A}$, $E(\mathrm{e}^{\lambda f}) < \infty$ and $N(\lambda f) = |\lambda| N(f)$. Assume that for every $f \in \mathcal{A}$,*

$$\mathrm{Ent}(\mathrm{e}^f) \leq \frac{1}{2}\, N(f)^2\, \mathrm{E}(\mathrm{e}^f).$$

Then, whenever F in $\mathcal{A}$ is such that $N(F) \leq 1$, then

$$\mathrm{E}(\mathrm{e}^{\lambda F}) \leq \mathrm{e}^{\lambda \mathrm{E}(F) + \lambda^2/2} \tag{2.20}$$

for every $\lambda \in \mathbb{R}$. Furthermore, for every $r \geq 0$,

$$\mu\big(F \geq \mathrm{E}(F) + r\big) \leq \mathrm{e}^{-r^2/2}, \tag{2.21}$$

and similarly for $-F$.

Proof. It just reproduces the proof of (2.19). Fix $F \in \mathcal{A}$ with $N(F) \leq 1$ and write $H(\lambda) = \mathrm{E}(\mathrm{e}^{\lambda F})$, $\lambda \geq 0$. Similarly, set $K(\lambda) = \frac{1}{\lambda} \log H(\lambda)$, $K(0) = \mathrm{E}(F)$. Applying the logarithmic Sobolev inequality of the statement to λF, $\lambda \geq 0$, we get $K'(\lambda) \leq \frac{1}{2}$ for $\lambda \geq 0$. Therefore,

$$K(\lambda) = K(0) + \int_0^\lambda K'(u)du \leq \mathrm{E}(F) + \frac{\lambda}{2}$$

and hence, for every $\lambda \geq 0$,

$$H(\lambda) \leq \mathrm{e}^{\lambda \mathrm{E}(F) + \lambda^2/2}.$$

Changing F into $-F$ yields (2.20). The proof is completed similarly. $\square$

We begin by adding several comments to Proposition 2.3.

If $N(F) \leq c$ in Proposition 2.3, then, by homogeneity,

$$\mu\big(F \geq \mathrm{E}(F) + r\big) \leq \mathrm{e}^{-r^2/2c^2}, \quad r \geq 0.$$

Sometimes the class $\mathcal{A}$ in Proposition 2.3 only includes λf when $f \in \mathcal{A}$ and $\lambda \geq 0$. The proof above was written so as to show that (2.20) then only holds for all $\lambda \geq 0$. Such a modification can be proposed similarly on the subsequent statements. We use these remarks freely throughout this work.

Very often, the logarithmic Sobolev inequality is only available on a class $\mathcal{A}$ densely defined in some larger, more convenient, class. The class of cylindrical functions on an abstract Wiener space is one typical and important example. In particular, this class might consist of bounded functions, so that the integrability assumptions in Proposition 2.3 are immediate. The conclusions however are only of interest for unbounded functions. Rather than extend the logarithmic Sobolev inequality itself, one may note that the corresponding concentration inequality easily extends. Let us agree that a function f on X satisfies $N(f) \leq 1$ if there is a sequence of functions $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with $N(f_n) \leq 1$ (or, more generally, $N(f_n) \leq 1 + \frac{1}{n}$) that converge μ-almost everywhere to f. For example, under some stability properties of $\mathcal{A}$, f_n could be $f_n = \max(-n, \min(f, n))$ which thus define a sequence of bounded functions converging to f. Dirichlet forms associated to Markov semigroups are stable by Lipschitz functions and $\mathcal{E}(f_n, f_n) \leq \mathcal{E}(f, f)$, thus falling into this case. Energies given by generalized moduli of gradients (2.4) may also be considered. Then, if F on X is such that $N(F) \leq 1$, F is integrable and the conclusions of Proposition 2.3 holds. To see this, let $(F_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ with $N(F_n) \leq 1$ such that $F_n \to F$ almost everywhere. By Proposition 2.3, for every n and $r \geq 0$,

$$\mu\big(\big|F_n - \mathrm{E}(F_n)\big| \geq r\big) \leq 2\,\mathrm{e}^{-r^2/2}. \tag{2.22}$$

Let m be large enough that $\mu(|F| \leq m) \geq \frac{3}{4}$. Then, for some n_0 and every $n \geq n_0$, $\mu(|F_n| \leq m+1) \geq \frac{1}{2}$. Choose furthermore $r_0 > 0$ with $2\mathrm{e}^{-r_0^2/2} < \frac{1}{2}$. Therefore, intersecting the sets $\{|F_n| \leq m+1\}$ and $\{|F_n - \mathrm{E}(F_n)| \geq r_0\}$, we see that

$$\big|\mathrm{E}(F_n)\big| \leq r_0 + m + 1$$

for every $n \geq n_0$ thus. Hence, by (2.22) again,

$$\mu\big(|F_n| \geq r + r_0 + m + 1\big) \leq 2\,\mathrm{e}^{-r^2/2}$$

for every $r \geq 0$ and $n \geq n_0$. In particular $\sup_n \mathrm{E}(F_n^2) < \infty$ so that, by uniform integrability, $\mathrm{E}(|F|) < \infty$ and $\mathrm{E}(F_n) \to \mathrm{E}(F)$. Then, by Fatou's lemma, for every $\lambda \in \mathbb{R}$,

$$\mathrm{E}(\mathrm{e}^{\lambda F}) \leq \liminf_{n\to\infty} \mathrm{E}(\mathrm{e}^{\lambda F_n}) \leq \liminf_{n\to\infty} \mathrm{e}^{\lambda \mathrm{E}(F_n) + \lambda^2/2} = \mathrm{e}^{\lambda \mathrm{E}(F) + \lambda^2/2}.$$

One then concludes as in Proposition 2.3. We emphasize that the integrability of F came for free. A similar reasoning was used in (1.24)

Note furthermore that, in the preceding setting, if $N(F) \leq 1$, then

$$\mathrm{E}\big(\mathrm{e}^{\alpha F^2}\big) < \infty \tag{2.23}$$

for every $\alpha < \frac{1}{2}$. As will be seen below, this condition is optimal. (2.23) is a consequence of Proposition 1.2. A beautiful alternate argument in this case was suggested by L. Gross (cf. [A-M-S]) on the basis of (2.20). If γ is the canonical Gaussian measure on $\mathbb{R}$, by Fubini's theorem,

$$\begin{aligned}
\mathrm{E}\big(\mathrm{e}^{\alpha F^2}\big) &= \mathrm{E}\bigg(\int_{\mathbb{R}} \mathrm{e}^{\sqrt{2\alpha}xF} d\gamma(x)\bigg) \\
&\leq \int_{\mathbb{R}} \mathrm{e}^{\sqrt{2\alpha}x\mathrm{E}(F)+\alpha x^2} d\gamma(x) \\
&= \frac{1}{\sqrt{1-2\alpha}}\, \mathrm{e}^{\alpha \mathrm{E}(F)^2/(1-2\alpha)}.
\end{aligned}$$

The bound is optimal as can be seen from the example $F(x) = x$ (with respect to γ).

We now show how the preceding statement may be applied to the settings presented in Section 2.1 for logarithmic Sobolev inequalities in their more classical form. The results below are taken from [A-M-S], [A-S], [G-R], [Le1], [Ro3].

In the context of Dirichlet forms (2.7) associated to Markov semigroup, let $\mathcal{A}$ be the algebra of bounded functions on $(X, \mathcal{B})$ in the domain $\mathcal{D}(\mathcal{E})$ of the Dirichlet form. Take $N(f) = |||\cdot|||_\infty$, and let us agree, as above, that a measurable function f on X is such that $|||\cdot|||_\infty \leq 1$ if there is a sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with $|||f_n|||_\infty \leq 1$ that converge μ-almost everywhere to f.

Corollary 2.4. *Assume that for some $C > 0$ and every f in $\mathcal{A}$*

$$\mathrm{Ent}(f^2) \leq 2C\mathcal{E}(f).$$

Then, whenever F is such that $|||F|||_\infty \leq 1$, we have $\mathrm{E}(|F|) < \infty$ and, for every $r \geq 0$,

$$\mu\big(F \geq \mathrm{E}(F) + r\big) \leq \mathrm{e}^{-r^2/4C}.$$

Proof. Apply the logarithmic Sobolev inequality to $\mathrm{e}^{f/2}$ to get, according to (2.8),

$$\mathrm{Ent}(\mathrm{e}^f) \leq C|||f|||_\infty^2\, \mathrm{E}(\mathrm{e}^f).$$

The conclusion then follows from Proposition 2.3 (and homogeneity). □

As a second set of examples, consider an operator Γ on some class $\mathcal{A}$ such that $\Gamma(f) \geq 0$ and $\Gamma(\lambda f) = \lambda^2\Gamma(f)$ for every f in $\mathcal{A}$. As a typical example, $\Gamma(f) = |\nabla f|^2$ for a generalized modulus of gradient, or $\Gamma(f) = \Gamma(f, f)$ for a more general carré du champ. One may also choose $\Gamma(f) = |Df|^2$ for a discrete gradient such as (2.9). Keeping with the preceding comments, we agree that a function f on X is such that $N(f) = \|\Gamma(f)\|_\infty \leq 1$ if there is a sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ converging to f such that $\|\Gamma(f_n)\|_\infty \leq 1$ for every n. The following corollary to Proposition 2.3 is immediate.

Corollary 2.5. *Let $\mathcal{A}$ be such that, for every $f \in \mathcal{A}$ and $\lambda \in \mathbb{R}$, $\lambda f \in \mathcal{A}$, $E(\mathrm{e}^{\lambda f}) < \infty$. Assume that for some $C > 0$ and every $f \in \mathcal{A}$,*

$$\mathrm{Ent}(\mathrm{e}^f) \leq \frac{C}{2}\, \mathrm{E}\big(\Gamma(f)\, \mathrm{e}^f\big).$$

Then, whenever F is such that $\|\Gamma(f)\|_\infty \leq 1$, we have $\mathrm{E}(|F|) < \infty$ *and*

$$\mu\big(F \geq \mathrm{E}(F) + r\big) \leq \mathrm{e}^{-r^2/2C}$$

for every $r \geq 0$.

In case of a local gradient operator $\Gamma(f) = |\nabla f|^2$ (2.3) on a metric space $X, d)$ satisfying the chain rule formula (2.5), a logarithmic Sobolev inequality of the type

$$\mathrm{Ent}(f^2) \leq 2C\,\mathrm{E}\big(|\nabla f|^2\big)$$

is actually equivalent to the logarithmic Sobolev inequality

$$\mathrm{Ent}(\mathrm{e}^f) \leq \frac{C}{2}\,\mathrm{E}\big(|\nabla f|^2 \mathrm{e}^f\big) \tag{2.24}$$

of Corollary 2.5 (on some appropriate class of functions $\mathcal{A}$ stable by the operations required for this equivalence to hold). As we will see, this is no more true for non-local gradients. Even in case of a local gradient, it may also happen that (2.20) holds for some class of functions for which the classical logarithmic Sobolev inequality is not satisfied. In the next statement, we do not specify the stability properties on $\mathcal{A}$.

Corollary 2.6. *Assume that for some* $C > 0$ *and all* f *in* $\mathcal{A}$

$$\mathrm{Ent}(f^2) \leq 2C\,\mathrm{E}\big(|\nabla f|^2\big).$$

Then, whenever F is such that $\|\nabla F\|_\infty \leq 1$, we have $\mathrm{E}(|F|) < \infty$ *and, for every* $r \geq 0$,

$$\mu\big(F \geq \mathrm{E}(F) + r\big) \leq \mathrm{e}^{-r^2/2C}.$$

Together with (1.28), for every set A with $\mu(A) > 0$,

$$\mu(A_r) \geq 1 - \mathrm{e}^{-\mu(A)^2 r^2/2C} \tag{2.25}$$

for every $r \geq 0$.

Let us consider, for example, in Corollary 2.6, the Gaussian measure γ on $\mathbb{R}^n$. The logarithmic Sobolev inequality (2.15) holds for all almost everywhere differentiable functions with gradients in $\mathrm{L}^2(\gamma)$. Let $\mathcal{A}$ be the class of bounded Lipschitz functions on $\mathbb{R}^n$. Let F be a Lipschitz function on $\mathbb{R}^n$. For any $n \in \mathbb{N}$, set $F_n = \max(-n, \min(F, n))$. Then F_n is bounded Lipschitz and converges almost everywhere to F. Moreover, if $\|F\|_{\mathrm{Lip}} \leq 1$, $\|F_n\|_{\mathrm{Lip}} \leq 1$ for every n. By Rademacher's theorem, F_n is almost everywhere differentiable with $|\nabla F_n| \leq 1$ almost everywhere. Therefore, as an application of Corollary 2.6, we thus recover that for any Lipschitz F with $\|F\|_{\mathrm{Lip}} \leq 1$,

$$\gamma\big(F \geq \mathrm{E}_\gamma(F) + r\big) \leq \mathrm{e}^{-r^2/2}, \quad r \geq 0,$$

which is the concentration property (1.22). In particular, the optimal constant in the exponent has been preserved throughout this procedure. We thus see how a logarithmic Sobolev inequality always determines a Gaussian concentration of isoperimetric nature.

The previous comment applies exactly similarly for the class of Boltzmann measures investigated in Theorem 1.1 (see also (2.17)). Moreover, the approximation procedure just described may be performed similarly for generalized gradients, on manifolds for example. Similarly, a cylindrical approximation would yield (1.25) for an infinite dimensional Gaussian measure from the Gaussian logarithmic Sobolev inequality. (1.25) would also follow from the logarithmic Sobolev inequality for infinite dimensional Gaussian measures, although the extension scheme is much simpler at the level of concentration inequalities.

We present next an application in a non-local setting following [A-S]. Recall the "gradient" (2.12) for a function f on $X = \{-1,+1\}^{\mathbb{Z}^d}$. Let μ be a Gibbs state on X corresponding to a finite range potential $\mathcal{J}$. It was shown by D. Stroock and B. Zegarlinski [S-Z] that the Dobrushin-Shlosman mixing condition ensures a logarithmic Sobolev inequality for μ

$$\mathrm{Ent}_\mu(f^2) \le 2C\,\mathrm{E}\big(|Df|^2\big)$$

for some $C > 0$. Assume moreover that $\mathcal{J}$ is shift-invariant. Let ψ be a continuous function on X for which $\mathrm{E}_\mu(\psi) = 0$ and

$$\beta = \sum_{k\in\mathbb{Z}^d} \|\partial_k\psi\|_\infty < \infty.$$

Let finally $(a_k)_{k\in\mathbb{Z}^d}$ be a sequence of real numbers with

$$\alpha^2 = \sum_{k\in\mathbb{Z}^d} a_k^2 < \infty.$$

For S^j the natural shift on $\mathbb{Z}^d$ (defined by $S^j(\omega_k) = \omega_{j+k}$ for all k), consider then a function F of the form

$$F = \sum_{j\in\mathbb{Z}^d} a_j \psi \circ S^j.$$

Such a function is actually defined as the limit in quadratic mean of the partial sums. As such, it is easily seen that

$$|||F|||_\infty \le \alpha\beta.$$

The preceding results (Corollary 2.4) apply to yield concentration and integrability properties of such functions F. In particular, for every $r \ge 0$,

$$\mu\big(F \ge \mathrm{E}_\mu(F) + r\big) \le \mathrm{e}^{-r^2/4C\alpha^2\beta^2}.$$

These results are thus very similar to the ones one gets in the non-interacting case (that is when μ is a product measure on $\{-1,+1\}^{\mathbb{Z}^d}$).

Before turning to variations of the previous basic argument to non-Gaussian tails in the next section, we present a recent result of S. Bobkov and F. Götze [B-G] which bounds, in this context, the Laplace transform of a function f in terms of some integral of its gradient. Up to numerical constants, this is an improvement upon the preceding statements. The proof however relies on the same ideas.

Let us consider, as in Corollary 2.5, an operator Γ on some class $\mathcal{A}$ in $\mathrm{L}^1(\mu)$ such that $\Gamma(\lambda f) = \lambda^2 \Gamma(f) \geq 0$ for every $f \in \mathcal{A}$ and $\lambda \in \mathbb{R}$.

Theorem 2.7. *Let $\mathcal{A}$ be such that, for every $f \in \mathcal{A}$ and $\lambda \in \mathbb{R}$ (or only $\lambda \in [-1, +1]$), $\lambda f \in \mathcal{A}$, $E(\mathrm{e}^{\lambda f}) < \infty$ and $\mathrm{E}(\mathrm{e}^{\lambda \Gamma(f)}) < \infty$. Assume that for every $f \in \mathcal{A}$,*

$$\mathrm{Ent}(\mathrm{e}^f) \leq \frac{1}{2}\, \mathrm{E}\big(\Gamma(f)\, \mathrm{e}^f\big).$$

Then, for every $f \in \mathcal{A}$

$$\mathrm{E}(\mathrm{e}^{f - \mathrm{E}(f)}) \leq \mathrm{E}\big(\mathrm{e}^{\Gamma(f)}\big).$$

Proof. Let, for every f, $g = \Gamma(f) - \log \mathrm{E}(\mathrm{e}^{\Gamma(f)})$, so that $\mathrm{E}(\mathrm{e}^g) = 1$. By (2.13),

$$\mathrm{E}\big(\Gamma(f)\, \mathrm{e}^f\big) - \mathrm{E}(\mathrm{e}^f) \log \mathrm{E}\big(\mathrm{e}^{\Gamma(f)}\big) \leq \mathrm{Ent}(\mathrm{e}^f).$$

Together with the hypothesis $\mathrm{E}(\Gamma(f)\mathrm{e}^f) \geq 2\, \mathrm{Ent}(\mathrm{e}^f)$, we get, for every f in $\mathcal{A}$,

$$\mathrm{Ent}(\mathrm{e}^f) \leq \mathrm{E}(\mathrm{e}^f) \log \mathrm{E}\big(\mathrm{e}^{\Gamma(f)}\big).$$

Apply this inequality to λf for every λ. With the notation of the proof of Proposition 2.3, for every $\lambda \in \mathbb{R}$,

$$K'(\lambda) \leq \frac{1}{\lambda^2}\, \psi(\lambda^2)$$

where $\psi(\lambda) = \log \mathrm{E}(\mathrm{e}^{\lambda \Gamma(f)})$. Now, ψ is non-negative, non-decreasing and convex, and $\psi(0) = 0$. Therefore $\psi(\lambda)/\lambda$ is non-decreasing in $\lambda \geq 0$. Recalling that $K(0) = \mathrm{E}(F)$, it follows that

$$K(1) \leq K(0) + \int_0^1 \frac{1}{\lambda^2}\, \psi(\lambda^2)\, d\lambda \leq \mathrm{E}(F) + \psi(1)$$

which is the result. Theorem 2.7 is established. □

2.4 Entropy-energy inequalities and non-Gaussian tails

The preceding basic argument admits a number of variations, some of which will be developed in the next chapters. We investigate first the case of defective logarithmic Sobolev inequality.

A defective logarithmic Sobolev inequality is of the type

$$\mathrm{Ent}_\mu(f^2) \leq a \mathrm{E}_\mu(f^2) + 2\mathcal{E}(f), \quad f \in \mathcal{A} \tag{2.26}$$

where $a \geq 0$. Of course, if $a = 0$, this is just a classical logarithmic Sobolev inequality. We would like to know if the preceding concentration inequalities of Gaussian type still hold under such a defective inequality, and whether the latter again determines the best exponential integrability in (2.23). According to the discussion in the preceding section, it will be enough to deal with the setting of Proposition 2.3.

Proposition 2.8. *In the framework of Proposition 2.3, assume that for some $a > 0$ and for every $f \in \mathcal{A}$,*

$$\mathrm{Ent}(\mathrm{e}^f) \leq a\mathrm{E}(\mathrm{e}^f) + \frac{1}{2}\, N(f)^2\, \mathrm{E}(\mathrm{e}^f).$$

Then, whenever $N(F) \leq 1$,

$$\mathrm{E}\big(\mathrm{e}^{\alpha F^2}\big) < \infty$$

for every $\alpha < \frac{1}{2}$.

Proof. Working first with a sequence $(F_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ such that $F_n \to F$, we may and do assume that $F \in \mathcal{A}$. Apply the defective logarithmic Sobolev inequality to λF for every $\lambda \in \mathbb{R}$. Letting as before $H(\lambda) = \mathrm{E}(\mathrm{e}^{\lambda F})$, we get

$$\lambda H'(\lambda) - H(\lambda)\log H(\lambda) \leq \Big(a + \frac{\lambda^2}{2}\Big) H(\lambda).$$

If $K(\lambda) = \frac{1}{\lambda}\log H(\lambda)$, we see that, for every $\lambda > 0$,

$$K'(\lambda) \leq \frac{a}{\lambda^2} + \frac{1}{2}\,.$$

Hence, for every $\lambda \geq 1$,

$$K(\lambda) = K(1) + \int_1^\lambda K'(u)du \leq K(1) + a + \frac{\lambda}{2}\,.$$

It follows that, for $\lambda \geq 1$,

$$\mathrm{E}(\mathrm{e}^{\lambda F}) \leq \big(\mathrm{E}(\mathrm{e}^F)\big)^\lambda \mathrm{e}^{a\lambda + \lambda^2/2}. \tag{2.27}$$

Let us choose first $\lambda = 2$. Then $\mathrm{E}(\mathrm{e}^{2F}) \leq A\mathrm{E}(\mathrm{e}^F)^2$ with $A = \mathrm{e}^{2(a+1)}$. Let m be large enough so that $\mu(|F| \geq m) \leq 1/4A$. Then $\mu(\mathrm{e}^F \geq \mathrm{e}^m) < 1/4A$ and

$$\begin{aligned}
\mathrm{E}(\mathrm{e}^F) &\leq \mathrm{e}^m + \mu(\mathrm{e}^F \geq \mathrm{e}^m)^{1/2}\big(\mathrm{E}(\mathrm{e}^{2F})\big)^{1/2} \\
&\leq \mathrm{e}^m + \sqrt{A}\,\mu(\mathrm{e}^F \geq \mathrm{e}^m)^{1/2}\mathrm{E}(\mathrm{e}^F) \\
&\leq 2\,\mathrm{e}^m.
\end{aligned}$$

Coming back to (2.27), for every $\lambda \geq 1$,

$$\mathrm{E}(\mathrm{e}^{\lambda F}) \leq 2^\lambda \mathrm{e}^{(m+a)\lambda + \lambda^2/2} = \mathrm{e}^{B\lambda + \lambda^2/2}$$

where $B = m + a + \log 2$. By Chebyshev's inequality,

$$\mu(F \geq r) \leq \mathrm{e}^{Br - r^2/2}$$

for every $r \geq A + 1$. Together with the same inequality for $-F$, the conclusion follows from the proof of Proposition 1.2. Proposition 2.8 is therefore established. $\square$

Inequality (2.25) actually fits into the more general framework of inequalities between entropy and energy introduced in [Ba1]. Given a non-negative function Ψ on $\mathbb{R}_+$, let us say that we have an entropy-energy inequality whenever for all f in $\mathcal{A}$ with $\mathrm{E}_\mu(f^2) = 1$,

$$\mathrm{Ent}_\mu(f^2) \leq \Psi\big(\mathcal{E}(f)\big). \tag{2.28}$$

By homogeneity, logarithmic Sobolev inequalities correspond to linear functions Ψ whereas defective logarithmic Sobolev inequalities correspond to affine Ψ's. Assume Ψ to be concave. Then (2.28) is equivalent to a family of defective logarithmic Sobolev inequalities

$$\mathrm{Ent}_\mu(f^2) \leq \varepsilon \mathrm{E}_\mu(f^2) + C(\varepsilon)\mathcal{E}(f), \quad \varepsilon \geq 0. \tag{2.29}$$

It is plain that, in the various settings studied above, the Laplace transform approach may be adapted to such an entropy-energy function. Depending upon to the rate at which Ψ increases to infinity, or, equivalently upon the behavior of $C(\varepsilon)$ as $\varepsilon \to 0$, various integrability results on Lipschitz functions may be obtained. It may even happen that Lipschitz functions are bounded if Ψ does not increase too quickly.

On the pattern of Proposition 2.3, we describe a general result that yields a variety of Laplace transform and tail inequalities for Lipschitz functions under some entropy-energy inequality. An alternate description of the next statement is presented in the paper [G-R] on the basis of (2.29). As will be studied in Parts 4 and 5, the form of the entropy-energy inequalities of Proposition 2.9 below is adapted to the concept of modified logarithmic Sobolev inequalities which often arise when the chain rule formula for the energy fails.

Let $\mathcal{A}$ be a class of functions in $\mathrm{L}^1(\mu)$. For every f in $\mathcal{A}$, let $N(f) \geq 0$. According to the argument developed for Proposition 2.3, the proof of the following statement is straighforward.

Proposition 2.9. *Let $\mathcal{A}$ be such that, for every $f \in \mathcal{A}$ and $\lambda \in \mathbb{R}$, $\lambda f \in \mathcal{A}$, $E(\mathrm{e}^{\lambda f}) < \infty$ and $N(\lambda f) = |\lambda| N(f)$. Assume there is a measurable function $B(\lambda) \geq 0$ on $\mathbb{R}_+$ such that for every $f \in \mathcal{A}$ with $N(f) \leq \lambda$,*

$$\mathrm{Ent}(\mathrm{e}^f) \leq B(\lambda)\,\mathrm{E}(\mathrm{e}^f).$$

Then, for every F in $\mathcal{A}$ such that $N(F) \leq 1$,

$$\mathrm{E}(\mathrm{e}^{\lambda F}) \leq \exp\bigg(\lambda \mathrm{E}(F) + \lambda \int_0^\lambda \frac{B(s)}{s^2}\, ds\bigg)$$

for every $\lambda \in \mathbb{R}$.

By homogeneity of N, Proposition 2.3 corresponds to the choice of $B(\lambda) = \lambda^2/2$, $\lambda \geq 0$.

The various examples discussed on the basis of Proposition 2.3 may also be reconsidered in this context. Suppose, for example, that, for some generalized modulus of gradient $|\nabla f|$, the entropy-energy inequality (2.28) holds. Then, by the change of variable formula, for every f with $\|\nabla f\|_\infty \leq \lambda$,

$$\mathrm{Ent}(\mathrm{e}^f) \leq \Psi\Big(\frac{\lambda^2}{4}\Big)\, \mathrm{E}(\mathrm{e}^f).$$

Now, depending upon how $B(\lambda)$ grows as λ goes to infinity, Proposition 2.9 will describe various tail estimates of Lipschitz functions. Rather than to discuss this in detail, let us briefly examine three specific behaviors of $B(\lambda)$.

Corollary 2.10. *In the setting of Proposition 2.9, if*

$$\int^\infty \frac{B(\lambda)}{\lambda^2}\, d\lambda < \infty, \tag{2.30}$$

then there exists $C > 0$ such that $\|F\|_\infty \leq C$ for every F such that $N(F) \leq 1$.

Proof. It is an easy matter to see from (2.30) and Proposition 2.9, that

$$\mathrm{E}(\mathrm{e}^{\lambda|F|}) \leq \mathrm{e}^{C\lambda}$$

for some $C > 0$ and all $\lambda \geq 0$ large enough. By Chebyshev's inequality, this implies that

$$\mu\big(|F| \geq 2C\big) \leq \mathrm{e}^{-C\lambda} \to 0$$

as $\lambda \to \infty$. Corollary 2.10 is proved. Actually, if N is the Lipschitz norm on a metric space (X, d), the diameter of X will be finite (less than or equal to $2C$), see [Le2]. □

In the second example, we consider a Gaussian behavior only for the small values of λ. The statement describes the typical tail of the exponential distribution (cf. Section 4.1).

Corollary 2.11. *In the setting of Proposition 2.9, assume that for some $c > 0$ and $\lambda_0 > 0$,*

$$B(\lambda) \leq c\lambda^2 \tag{2.31}$$

for every $0 \leq \lambda \leq \lambda_0$. Then, if F is such that $N(F) \leq 1$, we have $\mathrm{E}(|F|) < \infty$ and, for every $r \geq 0$,

$$\mu\big(F \geq \mathrm{E}(F) + r\big) \leq \exp\bigg(-\min\Big(\frac{\lambda_0 r}{2}, \frac{r^2}{4c}\Big)\bigg).$$

Proof. Arguing as next to Proposition 2.3, we may assume that $F \in \mathcal{A}$. With the notation of the proof of Proposition 2.3, for every $0 \leq \lambda \leq \lambda_0$,

$$K'(\lambda) \leq c\lambda^2.$$

Therefore $K(\lambda) \le K(0) + c\lambda$ so that

$$\mathrm{E}(\mathrm{e}^{\lambda F}) \le \mathrm{e}^{\lambda \mathrm{E}(F) + c\lambda^2}$$

for every $0 \le \lambda \le \lambda_0$ thus. By Chebyshev's inequality,

$$\mu\big(F \ge \mathrm{E}(F) + r\big) \le \mathrm{e}^{-\lambda r + c\lambda^2}.$$

If $r \le 2c\lambda_0$, choose $\lambda = \frac{r}{2c}$ while if $r \le 2c\lambda_0$, we simply take $\lambda = \lambda_0$. The conclusion easily follows. □

A third example of interest concerns Poisson tails on which we will come back in Part 5.

Corollary 2.12. *In the setting of Proposition 2.9, assume that for some $c, d > 0$,*

$$B(\lambda) \le c\lambda^2\, \mathrm{e}^{d\lambda} \tag{2.32}$$

for every $\lambda \ge 0$. Then, if F is such that $N(F) \le 1$, we have $\mathrm{E}(|F|) < \infty$ and, for every $r \ge 0$,

$$\mu\big(F \ge \mathrm{E}(F) + r\big) \le \exp\bigg(-\frac{r}{4d}\, \log\Big(1 + \frac{dr}{2c}\Big)\bigg).$$

In particular, $\mathrm{E}(\mathrm{e}^{\alpha |F| \log_+ |F|}) < \infty$ for sufficiently small $\alpha > 0$.

Proof. It is similar to the preceding ones. We have

$$K'(\lambda) \le c\, \mathrm{e}^{d\lambda}, \quad \lambda \ge 0.$$

Hence, $K(\lambda) \le K(0) + \frac{c}{d}(\mathrm{e}^{d\lambda} - 1)$, that is

$$\mathrm{E}(\mathrm{e}^{\lambda F}) \le \mathrm{e}^{\lambda \mathrm{E}(F) + \frac{c\lambda}{d}(\mathrm{e}^{d\lambda} - 1)}, \quad \lambda \ge 0.$$

By Chebyschev's inequality, for every $r \ge 0$ and $\lambda \ge 0$,

$$\mu\big(F \ge \mathrm{E}(F) + r\big) \le \mathrm{e}^{-\lambda r + \frac{c\lambda}{d}(\mathrm{e}^{d\lambda} - 1)}.$$

When $r \le \frac{4c}{d}$ (the constants are not sharp), choose $\lambda = \frac{r}{4c}$ so that

$$\mathrm{e}^{-\lambda r + \frac{c\lambda}{d}(\mathrm{e}^{d\lambda} - 1)} \le \mathrm{e}^{-\lambda r + 2c\lambda^2} = \mathrm{e}^{-\frac{r^2}{8c}},$$

while, when $r \ge \frac{4c}{d}$, choose $\lambda = \frac{1}{d}\log(\frac{dr}{2c})$ for which

$$\mathrm{e}^{-\lambda r + \frac{c\lambda}{d}(\mathrm{e}^{d\lambda} - 1)} \le \mathrm{e}^{-\frac{r}{2d}\log(\frac{dr}{2c})}.$$

These two estimates together yield the inequality of Corollary 2.11. The proof is complete. □

The inequality of Corollary 2.12 describes the classic Gaussian tail behavior for the small values of r and the Poisson behavior for the large values of r (with respect to the ratio $\frac{c}{d}$). The constants have no reason to be sharp.

We refer to the recent work [G-R] for further examples in this line of investigation.

2.5 Poincaré inequalities and concentration

In the last section, we apply the preceding functional approach in case of a spectral gap inequality. As we have seen (Proposition 2.1), spectral gap inequalities are usually weaker than logarithmic Sobolev inequalities, and, as a result, they only imply exponential integrability of Lipschitz functions. The result goes back to M. Gromov and V. Milman [G-M] (on a compact Riemannian manifold but with an argument that works similarly in a more general setting; see also [Br]). It has been investigated recently in [A-M-S] and [A-S] using moment bounds, and in [Sc] using a differential inequality on Laplace transforms similar to Herbst's argument. We follow here the approach of S. Aida and D. Stroock [A-S].

Assume that for some energy function $\mathcal{E}$ on a class $\mathcal{A}$,

$$\mathrm{Var}(f) \le C\mathcal{E}(f).$$

Apply this inequality to $\mathrm{e}^{f/2}$. If $\mathcal{E}$ is the Dirichlet form associated to a symmetric Markov semigroup (2.7), we can apply (2.8) to get

$$\mathrm{E}(\mathrm{e}^f) - \mathrm{E}(\mathrm{e}^{f/2})^2 \le \frac{C}{2}\, |||f|||_\infty^2\, \mathrm{E}(\mathrm{e}^f).$$

In case $\mathcal{E}$ is the energy of a local gradient satisfying the chain rule formula, the constant $\frac{1}{2}$ is improved to $\frac{1}{4}$. The following statement thus summarizes the various instances which may be considered.

Let again $\mathcal{A}$ be a subset of $\mathrm{L}^1(\mu)$. For every $f \in \mathcal{A}$, let $N(f) \ge 0$. We agree that $N(f) \le 1$ for some function f on X if f is the limit of a sequence of functions $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with $N(f_n) \le 1$ for every n.

Proposition 2.13. *Let $\mathcal{A}$ be such that, for every $f \in \mathcal{A}$ and every $\lambda \in \mathbb{R}$, $\mathrm{E}(\mathrm{e}^{\lambda f}) < \infty$ and $N(\lambda f) = |\lambda| N(f)$. Assume that for some $C > 0$ and every $f \in \mathcal{A}$,*

$$\mathrm{E}(\mathrm{e}^f) - \mathrm{E}(\mathrm{e}^{f/2})^2 \le C N(f)^2 \mathrm{E}(\mathrm{e}^f). \tag{2.33}$$

Then, for every F such that $N(F) \le 1$, $\mathrm{E}(|F|) < \infty$ and

$$\mathrm{E}\big(\mathrm{e}^{\lambda(F-\mathrm{E}(F))}\big) \le \prod_{k=0}^{\infty} \left(\frac{1}{1-\frac{C\lambda^2}{4^k}}\right)^{2^k} \tag{2.34}$$

for all $|\lambda| < 1/\sqrt{C}$. In particular,

$$\mathrm{E}(\mathrm{e}^{\alpha|F|}) < \infty$$

for every $\alpha < 1/\sqrt{C}$.

Proof. Assume first $F \in \mathcal{A}$ with $N(F) \leq 1$. Set $H(\lambda) = \mathrm{E}(\mathrm{e}^{\lambda F})$, $\lambda \geq 0$. Applying (2.33) to λF yields

$$H(\lambda) - H\Big(\frac{\lambda}{2}\Big)^2 \leq C\lambda^2 H(\lambda).$$

Hence, for every $\lambda < 1/\sqrt{C}$,

$$H(\lambda) \leq \frac{1}{1 - C\lambda^2}\, H\Big(\frac{\lambda}{2}\Big)^2.$$

Applying the same inequality for $\lambda/2$ and iterating, yields, after n steps,

$$H(\lambda) \leq \prod_{k=0}^{n-1}\bigg(\frac{1}{1 - \frac{C\lambda^2}{4^k}}\bigg)^{2^k} H\Big(\frac{\lambda}{2^n}\Big)^{2^n}.$$

Now $H(\lambda/\alpha)^\alpha \to \mathrm{e}^{\lambda \mathrm{E}(F)}$ as $\alpha \to \infty$. Hence, (2.34) is satisfied for this F which we assumed in $\mathcal{A}$. In particular, if $0 < \lambda_0 < 1/\sqrt{C}$, and if

$$K_0 = K_0(\lambda_0) = \prod_{k=0}^{\infty}\bigg(\frac{1}{1 - \frac{C\lambda_0^2}{4^k}}\bigg)^{2^k} < \infty,$$

then

$$\mu\big(\big|F - \mathrm{E}(F)\big| \geq r\big) \leq 2K_0\, \mathrm{e}^{-\lambda_0 r} \tag{2.35}$$

for every $r \geq 0$. Applying (2.35) to a sequence $(F_n)_{n\in\mathbb{N}}$ converging to F with $N(F_n) \leq 1$, and arguing as next to Proposition 2.3 immediately yields the full conclusion of the Proposition. The proof is thus complete. $\square$

The infinite product (2.34) has been estimated in [B-L1] by

$$\frac{1+\sqrt{C}}{1-\sqrt{C}}\,.$$

The example of the exponential measure investigated in Section 4.1 below shows that the condition $|\lambda| < 1/\sqrt{C}$ in Proposition 2.13 is optimal. Namely, let ν be the measure with density $\frac{1}{2}\,\mathrm{e}^{-|x|}$ with respect to Lebesgue measure on $\mathbb{R}$. Then, by Lemma 4.1,

$$\mathrm{Var}_\nu(f) \leq 4\,\mathrm{E}_\nu\big({f'}^2\big)$$

for every smooth f. Therefore, if $N(f) = \|f'\|_\infty$, (2.33) holds with $C = 1$, which is optimal as shown by the case $f(x) = x$.

Proposition 2.13 actually strengthens the early observation by R. Brooks [Br]. Namely, if M is a complete Riemannian manifold with finite volume $V(M)$, and if $V(x,r)$ is the volume of the ball $B(x,r)$ with center x and radius $r \geq 0$, then M has spectral gap zero as soon as

$$\liminf_{r\to\infty} \frac{-1}{r}\, \log\big[V(M) - V(x,r)\big] = 0 \tag{2.36}$$

for some (all) x in M.

3. DEVIATION INEQUALITIES FOR PRODUCT MEASURES

In the recent years, M. Talagrand has developed striking new methods for investigating the concentration of measure phenomenon for product measures. These ideas led to significant progress in an number of various areas such as probability in Banach spaces, empirical processes, geometric probability, statistical mechanics... The interested reader will find in the important contribution [Ta6] a complete account of these methods and results (see also [Ta7]). In this chapter, we indicate an alternate approach to some of Talagrand's inequalities based on logarithmic Sobolev inequalities and the methods of Chapter 2. The main point is that while concentration inequalities do not necessarily tensorize, the results follow from stronger logarithmic Sobolev inequalities which, as we know, do tensorize. In particular, we emphasize dimension free results.

The main deviation inequalities for convex functions form the core of Section 3.2, introduced by the discrete concentration property with respect to the Hamming metric in 3.1. Applications to sharp bounds on empirical processes conclude the chapter.

While it is uncertain whether this approach could recover Talagrand's abstract principles, the deviation inequalities themselves follow rather easily from it. On the abstract inequalities themselves, let us mention here the recent alternate approach by K. Marton [Mar1], [Mar2] and A. Dembo [De] (see also [D-Z]) based on information inequalities and coupling in which the concept of entropy also plays a crucial role. Hypercontraction methods were already used in [Kw-S] to study integrability of norms of sums of independent vector valued random variables. The work by K. Marton also involves Markov chains. Her arguments have been brought into relation recently with the logarithmic Sobolev inequality approach, and her results have been extended to larger classes of Markov chains, by P.-M. Samson [Sa]. We review some of these ideas in Section 3.3.

3.1 Concentration with respect to the Hamming metric

A first result on concentration in product spaces is the following. Let $(X_i, \mathcal{B}_i, \mu_i)$, $i = 1, \ldots, n$, be are arbitrary probability space, and let $P = \mu_1 \otimes \cdots \otimes \mu_n$ be a product measure on the product space $X = X_1 \times \cdots \times X_n$. A generic point in X is denoted by $x = (x_1, \ldots, x_n)$. Then, for every F on X such that $|F(x) - F(y)| \leq 1$

whenever $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ only differ by one coordinate,

$$P\big(F \geq \mathrm{E}_P(F) + r\big) \leq \mathrm{e}^{-r^2/2n}. \tag{3.1}$$

This inequality can be established by rather elementary martingale arguments [Mau1], [M-S], and was important in the early developments of concentration in product spaces (cf. [Ta6]). Our first aim will be to realize that it is also an elementary consequence of the logarithmic Sobolev approach developed in Section 2.3. We owe this observation to S. Kwapień.

Let f on the product space X. Recall we define f_i on X_i, $i = 1, \ldots, n$, by $f_i(x_i) = f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$ with $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ fixed.

Proposition 3.1. For every f on the product space X,

$$\mathrm{Ent}_P(\mathrm{e}^f) \leq \frac{1}{2} \sum_{i=1}^n \mathrm{E}_P\bigg(\int\!\!\int \big(f_i(x_i) - f_i(y_i)\big)^2 \mathrm{e}^{f_i(x_i)} d\mu_i(x_i) d\mu_i(y_i)\bigg).$$

Proof. The proof is elementary. We may assume f bounded. By the product property of entropy, it is enough to deal with the case $n = 1$. By Jensen's inequality,

$$\mathrm{Ent}_P(\mathrm{e}^f) \leq \mathrm{E}_P(f\mathrm{e}^f) - \mathrm{E}_P(\mathrm{e}^f)\mathrm{E}_P(f).$$

The right-hand-side of the latter may then be rewritten as

$$\frac{1}{2} \int\!\!\int \big(f(x) - f(y)\big)\big(\mathrm{e}^{f(x)} - \mathrm{e}^{f(y)}\big) dP(x) dP(y).$$

Since

$$(u - v)(\mathrm{e}^u - \mathrm{e}^v) \leq \tfrac{1}{2}\,(u - v)^2(\mathrm{e}^u + \mathrm{e}^v), \quad u, v \in \mathbb{R},$$

the conclusion easily follows. □

As a consequence of Proposition 3.1, if

$$N(f) = \sup_{x \in X} \bigg(\int \sum_{i=1}^n \big(f_i(x_i) - f_i(y_i)\big)^2 d\mu_i(y_i)\bigg)^{1/2},$$

then

$$\mathrm{Ent}_P(\mathrm{e}^f) \leq \frac{1}{2}\, N(f)^2 \mathrm{E}_P(\mathrm{e}^f).$$

Therefore, applying Proposition 2.3, if F is a Lipschitz function on X such that

$$|F(x) - F(y)| \leq \mathrm{Card}\{1 \leq i \leq n; x_i \neq y_i\},$$

then $N(F) \leq \sqrt{n}$ from which (3.1) follows.

This basic example actually indicates the route we will follow next, in particular with convex functions. Before turning to this case, let us mention that Proposition 3.1 has a clear analogue for variance that states that

$$\mathrm{Var}_P(f) \leq \frac{1}{2} \sum_{i=1}^n \mathrm{E}_P\bigg(\int\!\!\int \big(f_i(x_i) - f_i(y_i)\big)^2 d\mu_i(x_i) d\mu_i(y_i)\bigg). \tag{3.2}$$

3.2 Deviation inequalities for convex functions

One of the first important results underlying M. Talagrand's developments is the following inequality for arbitrary product measures [Ta1], [J-S] (see also [Mau2]). Let F be a convex Lipschitz function on $\mathbb{R}^n$ with $\|F\|_{\mathrm{Lip}} \le 1$. Let μ_i, $i = 1, \ldots, n$, be probability measures on $[0,1]$ and denote by P the product probability measure $\mu_1 \otimes \cdots \otimes \mu_n$. Then, for every $t \ge 0$,

$$P(F \ge m + t) \le 2\,\mathrm{e}^{-t^2/4} \tag{3.3}$$

where m is a median of F for P. As in the Gaussian case (1.20), this bound is dimension free, a feature of fundamental importance in this investigation. However, contrary to the Gaussian case, it is known that the convexity assumption on F is essential (cf. [L-T], p. 25). The proof of (3.3) [in the preceding references] is based on the inequality

$$\int \mathrm{e}^{\frac{1}{4}\, d(\cdot, \mathrm{Conv}(A))^2} dP \le \frac{1}{P(A)}$$

(where d is the Euclidean distance) which is established by geometric arguments and a simple induction on the number of coordinates. It has since been embedded in an abstract framework which M. Talagrand calls convex hull approximation (cf. [Ta6], [Ta7]). M. Talagrand also introduced the concept of approximation by a finite number of points [Ta2], [Ta6], [Ta7]. These powerful abstract tools have been used in particular to study sharp deviations inequalities for large classes of functions (cf. Section 3.4).

The aim of this section is to provide a simple proof of inequality (3.3) based on the functional inequalities presented in Part 2. The point is that while the deviation inequality (3.3) has no reason to be tensorizable, it is actually a consequence of a logarithmic Sobolev inequality, which only needs to be proved in dimension one. The main result in this direction is the following statement. Let thus $\mu_1, \ldots, \mu_n$ be arbitrary probability measures on $[0,1]$ and let P be the product probability $P = \mu_1 \otimes \cdots \otimes \mu_n$. We say that a function f on $\mathbb{R}^n$ is separately convex if it is convex in each coordinate. Recall that a convex function on $\mathbb{R}$ is continuous and almost everywhere differentiable. We denote by ∇f the usual gradient of f on $\mathbb{R}^n$ and by $|\nabla f|$ its Euclidean length.

Theorem 3.2. *Let f be a function on $\mathbb{R}^n$ such that $\log f^2$ is separately convex ($f^2 > 0$). Then, for any product probability P on $[0,1]^n$,*

$$\mathrm{Ent}_P(f^2) \le 4\,\mathrm{E}_P\big(|\nabla f|^2\big)$$

Notice that Theorem 3.2 amounts to saying that for every separately convex function f on $\mathbb{R}^n$,

$$\mathrm{Ent}_P(\mathrm{e}^f) \le \mathrm{E}_P\big(|\nabla f|^2 \mathrm{e}^f\big). \tag{3.4}$$

Proof. By a simple approximation, it is enough to deal with sufficiently smooth functions. We establish a somewhat stronger result, namely that for any product probability P on $\mathbb{R}^n$, and any smooth separately convex function f,

$$\mathrm{Ent}_P(\mathrm{e}^f) \leq \iint \sum_{i=1}^{n} (x_i - y_i)^2 (\partial_i f)^2(x)\,\mathrm{e}^{f(x)} dP(x)dP(y). \tag{3.5}$$

By Proposition 3.1, it is enough to show that for every $i = 1, \ldots, n$,

$$\begin{aligned}\frac{1}{2}\iint \big(f_i(x_i) - f_i(y_i)\big)^2 \mathrm{e}^{f_i(x_i)} d\mu_i(x_i)d\mu_i(y_i) \\ \leq \iint (x_i - y_i)^2 {f_i}'(x_i)^2\,\mathrm{e}^{f_i(x_i)} d\mu_i(x_i)d\mu_i(y_i).\end{aligned}$$

We may thus assume that $n = 1$. Now,

$$\begin{aligned}\frac{1}{2}\iint \big(f_(x) - f(y)\big)^2 \mathrm{e}^{f(x)} d\mu(x)d\mu(y) \\ \leq \iint_{\{f(x)\geq f(y)\}} \big(f(x) - f(y)\big)^2 \mathrm{e}^{f(x)} d\mu(x)d\mu(y).\end{aligned}$$

Since f is convex, for all $x, y \in \mathbb{R}$,

$$f(x) - f(y) \leq (x - y)f'(x).$$

The proof is easily completed. Theorem 3.2 is established. □

It should be emphasized that inequality (3.5), established in the preceding proof for arbitrary product measures on $\mathbb{R}^n$ is actually a stronger version of Theorem 3.2 which is particulary used for norms of sums of independent random vectors (Section 3.4). This inequality puts forward the generalized gradient (in dimension one)

$$\big|\nabla f(x)\big| = \left(\int (x - y)^2 f'(y)^2 d\mu(y)\right)^{1/2}$$

of statistical interest.

With a little more effort, the constant of the logarithmic Sobolev inequality of Theorem 3.2 may be improved to 2 (which is probably the optimal constant). We need simply improve the estimate of the entropy in dimension one. To this end, recall the variational caracterization of entropy ([H-S]) as

$$\mathrm{Ent}(\mathrm{e}^f) = \inf_{c>0} \mathrm{E}_P\big(f\mathrm{e}^f - (\log c + 1)\,\mathrm{e}^f + c\big). \tag{3.6}$$

Let P be a probability measure concentrated on $[0, 1]$. Let f be (smooth and) convex on $\mathbb{R}$. Let then $y \in [0, 1]$ be a point at which f is minimum and take $c = \mathrm{e}^{f(y)}$ (in (3.6)). For every $x \in [0, 1]$,

$$\begin{aligned}f(x)\,\mathrm{e}^{f(x)} - (\log c + 1)\,\mathrm{e}^{f(x)} + c &= \big[f(x) - f(y)\big]\mathrm{e}^{f(x)} - \big[\mathrm{e}^{f(x)} - \mathrm{e}^{f(y)}\big] \\ &= \big[\big(f(x) - f(y)\big) - 1 + \mathrm{e}^{-(f(x)-f(y))}\big]\mathrm{e}^{f(x)} \\ &\leq \frac{1}{2}\big[f(x) - f(y)\big]^2 \mathrm{e}^{f(x)}\end{aligned}$$

since $u - 1 + \mathrm{e}^{-u} \le \frac{u^2}{2}$ for every $u \ge 0$. Hence, by convexity, and since $x, y \in [0,1]$,

$$f(x)\,\mathrm{e}^{f(x)} - (\log c + 1)\,\mathrm{e}^{f(x)} + c \le \frac{1}{2}\,{f'(x)}^2 \mathrm{e}^{f(x)}$$

from which we deduce, together with (3.6), that

$$\mathrm{Ent}_P(\mathrm{e}^f) \le \frac{1}{2}\,\mathrm{E}_P\big({f'}^2\mathrm{e}^f\big).$$

We may now apply the results of Section 2.3 to get Gaussian deviation inequalities for convex Lipschitz functions with respect to product measures. (On the discrete cube, see also [Bob1]).

Corollary 3.3. *Let F be a separately convex Lipschitz function on $\mathbb{R}^n$ with Lipschitz constant $\|F\|_{\mathrm{Lip}} \le 1$. Then, for every product probability P on $[0,1]^n$, and every $r \ge 0$,*

$$P\big(F \ge \mathrm{E}_P(F) + r\big) \le \mathrm{e}^{-r^2/2}.$$

This inequality is the analogue of (3.3) with the mean instead of the (a) median m and the improved bound $\mathrm{e}^{-t^2/2}$.

The proof of Corollary 3.3 is a direct application of Corollary 2.5. Only some regularization procedure has to be made precise. Replacing F by a convolution with a Gaussian kernel, we may actually suppose that $|\nabla F| \le 1$ everywhere. Then, the argument is entirely similar to the one detailed, for example in the Gaussian case (after Corollary 2.6). The result follows by approximation.

M. Talagrand [Ta1] (see also [J-S], [Mau2], [Ta6], [Ta7]) actually showed deviation inequalities under the level m, that is an inequality for $-F$ (F convex). It yields a concentration result of the type

$$P\big(|F - m| \ge r\big) \le 4\,\mathrm{e}^{-r^2/4}, \quad r \ge 0. \tag{3.7}$$

It does not seem that such a deviation inequality for $-F$, F convex, follows from the preceding approach (since e^{-F} need not be convex). At a weak level though, we may use Poincaré inequalities. Indeed, we may first state the analogue of Theorem 3.2 for variance, whose proof is similar. This result was first mentioned in [Bob2].

Proposition 3.4. *Let f be a separately convex function on $\mathbb{R}^n$. Then, for any product probability P on $[0,1]^n$,*

$$\mathrm{Var}_P(f) \le \mathrm{E}_P\big(|\nabla f|^2\big).$$

Therefore, for any separately convex function F with $\|F\|_{\mathrm{Lip}} \le 1$,

$$P\big(\big|F - \ E_P(F)\big| \ge r\big) \le \frac{1}{r^2}$$

for every $r \ge 0$. As seems to be indicated by the results in the next section, the convexity in each coordinate might not be enough to ensure deviation under the mean or the median. Using alternate methods, we will see indeed that sharp deviation inequalities do hold for concave functions, even under less stringent assumptions than Lipschitz. Although deviation inequalities above the mean or the median are the useful inequalities in probability and its applications, concentration inequalities are sometimes important issues (e.g. in geometry of Banach spaces [M-S], percolation, spin glasses... [Ta6]).

Corollary 3.3 of course extends to probability measures μ_i supported on $[a_i, b_i]$, $i = 1, \dots, n$, (following for example (3.5) of the proof of Theorem 3.2, or by scaling). In particular, if P is a product measure on $[a,b]^n$ and if F is separately convex on $\mathbb{R}^n$ with Lipschitz constant less than or equal to 1, for every $r \geq 0$,

$$P\big(F \geq \mathrm{E}_P(F) + r\big) \leq \mathrm{e}^{-r^2/2(b-a)^2}.$$

Let us also recall one typical application of these deviation inequalities to norms of random series. Let η_i, $i = 1, \dots, n$, be independent random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $|\eta_i| \leq 1$ almost surely. Let v_i, $i = 1, \dots, n$, be vectors in some arbitrary Banach space E with norm $\|\cdot\|$. Then, for every $r \geq 0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \eta_i v_i\right\| \geq \mathbb{E}\left\|\sum_{i=1}^n \eta_i v_i\right\| + r\right) \leq \mathrm{e}^{-r^2/8\sigma^2}$$

where

$$\sigma^2 = \sup_{\|\xi\| \leq 1} \sum_{i=1}^n \langle \xi, v_i \rangle^2.$$

This inequality is the analogue of the Gaussian deviation inequalities (1.8) and (1.24). For the proof, simply consider F on $\mathbb{R}^n$ defined by

$$F(x) = \left\|\sum_{i=1}^n x_i v_i\right\|, \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Then, by duality, for $x, y \in \mathbb{R}^n$,

$$\big|F(x) - F(y)\big| \leq \left\|\sum_{i=1}^n (x_i - y_i) v_i\right\| = \sup_{\|\xi\| \leq 1} \sum_{i=1}^n (x_i - y_i) \langle \xi, v_i \rangle \leq \sigma |x - y|,$$

where the last step is obtained from the Cauchy-Schwarz inequality.

3.3 Information inequalities and concentration

Recently, K. Marton [Mar1], [Mar2] (see also [Mar3]) studied the preceding concentration inequalities in the context of contracting Markov chains. Her approach is based on information inequalities and coupling ideas. Specifically, she is using convexity of entropy together with Pinsker's inequality [Pi]

$$\|\mu - \nu\|_{\mathrm{T.V.}} \leq \sqrt{\frac{1}{2} \operatorname{Ent}_\mu\left(\frac{d\nu}{d\mu}\right)} \tag{3.8}$$

where the probability measure ν is assumed to be absolutely continuous with respect to μ with density $\frac{d\nu}{d\mu}$. That such an inequality entails concentration properties may be shown in the following way. Given a separable metric space (X,d) and two Borel probability measures μ and ν on X, set

$$W_1(\mu,\nu) = \inf \iint d(x,y) d\pi(x,y)$$

where the infimum runs over all probability measures π on the product space $X \times X$ with marginals μ and ν. Consider now the inequality

$$W_1(\mu,\nu) \le \sqrt{2C\,\mathrm{Ent}_\mu\Big(\frac{d\nu}{d\mu}\Big)} \tag{3.9}$$

for some $C > 0$. By the coupling characterization of the total variation distance, Pinsker's inequality corresponds to the trivial distance on X (and to $C = \frac{1}{4}$). Let then A and B with $\mu(A), \mu(B) > 0$, and consider the conditional probabilities $\mu_A = \mu(\cdot|A)$ and $\mu_B = \mu(\cdot|B)$. By the triangle inequality and (3.9),

$$\begin{aligned} W_1(\mu_A,\mu_B) &\le W_1(\mu,\mu_A) + W_1(\mu,\mu_B) \\ &\le \sqrt{2C\,\mathrm{Ent}_\mu\Big(\frac{d\mu_A}{d\mu}\Big)} + \sqrt{2C\,\mathrm{Ent}_\mu\Big(\frac{d\mu_B}{d\mu}\Big)} \\ &= \sqrt{2C\log\frac{1}{\mu(A)}} + \sqrt{2C\log\frac{1}{\mu(B)}}\,. \end{aligned} \tag{3.10}$$

Now, all measures with marginals μ_A and μ_B must be supported on $A \times B$, so that, by the definition of W_1,

$$W_1(\mu_A,\mu_B) \ge d(A,B) = \inf\big\{d(x,y); x \in A, y \in B\big\}.$$

Then (3.10) implies a concentration inequality. Fix A with, say, $\mu(A) \ge \frac{1}{2}$ and take B the complement of A_r for $r \ge 0$. Then $d(A,B) \ge r$ so that

$$r \le \sqrt{2C\log\frac{1}{\mu(A)}} + \sqrt{2C\log\frac{1}{1-\mu(A_r)}} \le \sqrt{2C\log 2} + \sqrt{2C\log\frac{1}{1-\mu(A_r)}}\,.$$

Hence, whenever $r \ge 2\sqrt{2C\log 2}$ for example,

$$1 - \mu(A_r) \le \mathrm{e}^{-r^2/8C}.$$

Now, the product property of entropy allows us to tensorize Pinsker-type inequalities to produce concentration in product spaces. For example, this simple scheme may be used to recover, even with sharp constants, the concentration (3.1) with respect to the Hamming metric. Indeed, if we let d be the Hamming metric on the product space $X = X_1 \times \cdots \times X_n$, starting with (3.8), convexity of entropy shows that for any probability measure Q on X absolutely continuous with respect to the product measure $P = \mu_1 \otimes \cdots \otimes \mu_n$,

$$W_1(P, Q) \le \sqrt{\frac{n}{2}\,\mathrm{Ent}_P\Big(\frac{dQ}{dP}\Big)}$$

from which (3.1) follows according to the preceding argument.

It might be worthwhile noting that S. Bobkov and F. Götze [B-G] recently proved that an inequality such as (3.9) holding for all measures ν absolutely continous with respect to μ is actually equivalent to the Gaussian bound

$$\mathrm{E}_\mu(\mathrm{e}^{\lambda F}) \le \mathrm{e}^{\lambda \mathrm{E}_\mu(F) + C\lambda^2/2}$$

on the Lipschitz functions F on (X, d) with $\|F\|_{\mathrm{Lip}} \le 1$. This observation connects the information theory approach to the logarithmic Sobolev approach emphasized in this work. It also shows that a logarithmic Sobolev inequality in this case is a stronger statement than a Pinsker-type inequality.

For the Gaussian measure γ on $\mathbb{R}^n$ equipped with the Euclidean distance d_2, M. Talagrand [Ta9] proved that, not only (3.9) holds but

$$W_2(\gamma, \nu) \le \sqrt{2\,\mathrm{Ent}_\gamma\Big(\frac{d\nu}{d\gamma}\Big)} \tag{3.11}$$

where now

$$W_2(\gamma, \nu) = \inf\bigg(\int\!\!\int d_2(x, y)^2 d\pi(x, y)\bigg)^{1/2}.$$

He further investigated in this paper the case of the exponential distribution to recover its concentration properties (cf. Section 4.1). Recently, it was proved in [O-V] that (3.11) may be shown to follow from the Gaussian logarithmic Sobolev inequality (2.15).

In order to cover with these methods the inequalities for convex functions of Section 3.2, K. Marton [Mar2] introduced another metric on measures in the form of

$$d_2(\mu, \nu) = \mathrm{E}_\mu\bigg(\Big(1 - \frac{d\nu}{d\mu}\Big)_+^2\bigg)^{1/2}.$$

This distance is analogous to the variational distance and one can actually show that

$$d_2(\mu, \nu) = \inf\bigg(\int \mathbb{P}\big(\xi \ne y \,|\, \zeta = y\big)^2 d\nu(y)\bigg)^{1/2}$$

where the infimum is over all couples of random variables (ξ, ζ) such that ξ has distribution μ and ζ distribution ν. Note that $d_2(\mu, \nu)$ is not symmetric in μ, ν. Together with the appropriate information inequality on d_2 and convexity of relative entropy, she proved in this way the concentration inequalities for convex functions of the preceding section. Her arguments has been further developed by A. Dembo [De] to recover in this way most of M. Talagrand's abstract inequalities.

But K. Marton's approach was initially devoted to some non-product Markov chains for which it appears to be a powerful tool. More precisely, let P be a Markov chain on $[0,1]^n$ with transition kernels K_i, $i = 1, \ldots, n$, that is

$$dP(x_1, \ldots, x_n) = K_n(x_{n-1}, dx_n) \cdots K_2(x_1, dx_2) K_1(dx_1).$$

Assume that, for some $0 \le a < 1$, for every $i = 1, \ldots, n$, and every $x, y \in [0,1]$,

$$\big\| K_i(x, \cdot) - K_i(y, \cdot) \big\|_{\mathrm{T.V.}} \le a. \tag{3.12}$$

The case $a = 0$ of course corresponds to independent kernels K_i. The main result of [Mar2] (expressed on functions) is the following.

Theorem 3.5. *Let P be a Markov chain on $[0,1]^n$ satisfying (3.12) for some $0 \le a < 1$. For every convex Lipschitz map F on $\mathbb{R}^n$ with $\|F\|_{\mathrm{Lip}} \le 1$,*

$$P\big(F \ge \mathrm{E}_P(F) + r\big) \le \mathrm{e}^{-(1-\sqrt{a})^2 r^2/4}$$

for every $r \ge 0$ and similarly for $-F$.

This result has been extended in [Mar4] and, independently in [Sa], to larger classes of dependent processes. Moreover, in [Sa], P.-M. Samson brings into relation the information approach with the logarithmic Sobolev approach. Let P and Q be probability measures on $\mathbb{R}^n$. Following the one-dimensional definition of d_2, set

$$d_2(P, Q) = \inf \sup_{\alpha} \iint \sum_{i=1}^n \alpha_i(y) \mathrm{I}_{x_i \neq y_i} d\pi(x, y)$$

where the infimum is over all probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals P and Q and the supremum runs over all $\alpha = (\alpha_1, \ldots, \alpha_n)$ where the α_i's are non-negative functions on $\mathbb{R}^n$ such that

$$\int \sum_{i=1}^n \alpha_i^2(y) dQ(y) \le 1.$$

As shown by K. Marton, we have similarly a coupling description as

$$d_2(P, Q) = \inf \bigg(\int \sum_{i=1}^n \mathbb{P}\big(\xi_i \neq y_i \,|\, \zeta_i = y_i\big)^2 dQ(y) \bigg)^{1/2}$$

where the infimum runs over all random variables $\xi = (\xi_1, \ldots, \xi_n)$ and $\zeta = (\zeta_1, \ldots, \zeta_n)$ such that ξ has distribution P and ζ distribution Q.

Let now P denote the distribution of a sample $X_1, \ldots, X_n$ of real random variables. Following Marton's techniques, for any Q absolutely continuous with respect to P,

$$\max\big(d_2(P, Q), d_2(Q, P)\big) \le \sqrt{2\|M\| \,\mathrm{Ent}_P\Big(\frac{dQ}{dP}\Big)} \tag{3.13}$$

where $\|M\|$ is the operator norm of a certain mixing matrix M that measures the L^2-dependence of the variables $X_1, \ldots, X_n$. Now, $\|M\|$ may be shown to be bounded independently of the dimension in a number of interesting cases, including Doeblin recurrent Markov chains and Φ-mixing processes (cf. [Mar4], [Sa]). (3.13) then yields concentration inequalities for new classes of measures and processes.

Moreover, it is shown in [Sa] how (3.13) may be considered as a kind of dual version of the logarithmic Sobolev inequalities for convex (and concave) functions of Section 3.2 above. Let f be (smooth and) convex on $[0,1]^n$. By Jensen's inequality,

$$\frac{\mathrm{Ent}_P(\mathrm{e}^f)}{\mathrm{E}_P(\mathrm{e}^f)} \le \int f(x) \frac{\mathrm{e}^{f(x)}}{\mathrm{E}_P(\mathrm{e}^f)}\, dP(x) - \int f(y) dP(y).$$

Let P^f be the probability measure on $[0,1]^n$ whose density with respect to P is $\mathrm{e}^f / \mathrm{E}_P(\mathrm{e}^f)$. Let π be a probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals P and P^f. Then,

$$\frac{\mathrm{Ent}_P(\mathrm{e}^f)}{\mathrm{E}_P(\mathrm{e}^f)} \le \iint \big[f(y) - f(x)\big]\, d\pi(x,y).$$

Since f is convex, for every $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in [0,1]^n$,

$$f(x) - f(y) \le \sum_{i=1}^n |x_i - y_i| \big|\partial_i f(x)\big| \le \sum_{i=1}^n \big|\partial_i f(x)\big| \mathrm{I}_{x_i \ne y_i}\,.$$

As a consequence, for all probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals P and P^f,

$$\frac{\mathrm{Ent}_P(\mathrm{e}^f)}{\mathrm{E}_P(\mathrm{e}^f)} \le \iint \sum_{i=1}^n \big|\partial_i f(x)\big| \mathrm{I}_{x_i \ne y_i} d\pi(x,y).$$

According to the definition of $d_2(P, P^f)$, and by the Cauchy-Schwarz inequality,

$$\frac{\mathrm{Ent}_P(\mathrm{e}^f)}{\mathrm{E}_P(\mathrm{e}^f)} \le d_2(P, P^f) \bigg(\sum_{i=1}^n \int \big|\partial_i f(x)\big|^2 dP^f(x) \bigg)^{1/2}. \tag{3.14}$$

Since

$$\frac{dP^f}{dP} = \frac{\mathrm{e}^f}{\mathrm{E}_P(\mathrm{e}^f)}$$

we get from (3.13) and (3.14) that

$$\frac{\mathrm{Ent}_P(\mathrm{e}^f)}{\mathrm{E}_P(\mathrm{e}^f)} \le \|M\|^{1/2} \bigg(2\, \frac{\mathrm{Ent}_P(\mathrm{e}^f)}{\mathrm{E}_P(\mathrm{e}^f)} \bigg)^{1/2} \bigg(\int |\nabla f|^2 \frac{\mathrm{e}^f}{\mathrm{E}_P(\mathrm{e}^f)} dP \bigg)^{1/2}.$$

It follows that for every (smooth) convex function on $[0,1]^n$,

$$\mathrm{Ent}_P(\mathrm{e}^f) \le 2\|M\| \mathrm{E}_P\big(|\nabla f|^2 \mathrm{e}^f\big) \tag{3.15}$$

which amounts to the inequality of Theorem 3.2.

It is worthwhile noting that the same proof for a concave function f yields instead of (3.15)

$$\mathrm{Ent}_P(\mathrm{e}^f) \leq 2\|M\| \mathrm{E}_P\big(|\nabla f|^2\big)\mathrm{E}_P(\mathrm{e}^f). \tag{3.16}$$

These observations clarify the discussion on separately convex or concave functions in Section 3.2. In contrast to Theorem 3.2, the proof of these results fully uses the convexity or concavity assumptions on f rather than only convexity in each coordinate. Together with Herbst's argument, these inequalities imply the conclusions of Theorem 3.5. On the other hand, deviation inequalities under the mean for concave functions F only require $\mathrm{E}_P(|\nabla F|^2) \leq 1$.

3.4 Applications to bounds on empirical processes

Sums of independent random variables are a natural application of the preceding deviation inequalities for product measures. In this section, we survey some of these applications, with a particular emphasis on bounds for empirical processes.

Tail probabilities for sums of independent random variables have been extensively studied in classical Probability theory. One finished result is the so-called Bennett inequality (after contributions by Bernstein, Kolmogorov, Prohorov, Hoeffding etc). Let $X_1, \ldots, X_n$ be independent mean-zero real-valued random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $|X_i| \leq C$, $i = 1, \ldots, n$, and $\sum_{i=1}^n \mathbb{E}(X_i^2) \leq \sigma^2$. Set $S_n = X_1 + \cdots + X_n$. Then, for every $r \geq 0$,

$$\mathbb{P}(S_n \geq r) \leq \exp\bigg(-\frac{r}{2C}\,\log\bigg(1 + \frac{Cr}{\sigma^2}\bigg)\bigg). \tag{3.17}$$

Such an inequality describes the Gaussian tail for the values of r which are small with respect to σ^2, and the Poisson behavior for the large values (think, for example, of a sample of independent Bernoulli variables, with probability of success either $\frac{1}{2}$ or on the order of $\frac{1}{n}$.)

Now, in statistical applications, one is interested in such a bound uniformly over classes of functions, and importance of such inequalities has been emphasized recently in the statistical treatment of selection of models by L. Birgé and P. Massart [B-M1], [B-M2], [B-B-M]. More precisely, let $X_1, X_2, \ldots, X_n, \ldots$ be independent random variables with values in some measurable space $(S, \mathcal{S})$ with identical distribution $\mathcal{P}$, and let, for $n \geq 1$,

$$\mathcal{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$$

be the empirical measures (on $\mathcal{P}$). A class $\mathcal{F}$ of real measurable functions on S is said to be a Glivenko-Cantelli class if $\sup_{f \in \mathcal{F}} |\mathcal{P}_n(f) - \mathcal{P}(f)|$ converges almost surely to 0. It is a Donsker class if, in a sense to be made precise, $\sqrt{n}(\mathcal{P}_n(f) - \mathcal{P}(f))$, $f \in \mathcal{F}$, converges in distribution toward a centered Gaussian process with covariance function $\mathcal{P}(fg) - \mathcal{P}(f)\mathcal{P}(g)$, $f, g \in \mathcal{F}$. These definitions naturally extend the classic example of the class of all indicator functions of intervals $(-\infty, t]$, $t \in \mathbb{R}$ (studied precisely by Glivenko-Cantelli and Donsker). These asymptotic properties however often need to be turned into tail inequalities at fixed n on classes $\mathcal{F}$ which are as rich as possible (to determine accurate approximation by empirical models). In particular, these bounds aim to be as close as possible to the one-dimensional inequality (3.17) (corresponding to a class $\mathcal{F}$ reduced to only one function).

Sharp bounds for empirical processes have been obtained by M. Talagrand [Ta5], [Ta8] as a consequence of his abstract inequalities for product measures. We observe here that the functional approach based on logarithmic Sobolev inequalities developed in the preceding sections may be use to produce similar bounds. The key idea is to exploit the logarithmic Sobolev inequality (3.5) emphasized in the proof of Theorem 3.2 and to apply it to norm of sums of independent vector valued random variables. The convexity properties of the norm of a sum allow us to easily estimate the gradient on the right-hand side of (3.5). It yields to the following result for which we refer to [Le4] for further details. Let as before X_i, $i = 1, \ldots, n$, be independent random variables with values in some space S, and let $\mathcal{F}$ be a countable class of measurable functions on S. Set

$$Z = \sup_{f \in \mathcal{F}} \biggl| \sum_{i=1}^{n} f(X_i) \biggr|.$$

Theorem 3.6. *If $|f| \leq C$ for every f in $\mathcal{F}$, and if $\mathbb{E}f(X_i) = 0$ for every $f \in \mathcal{F}$ and $i = 1, \ldots, n$, then, for all $r \geq 0$,*

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + r\big) \leq 3 \exp\biggl(-\frac{r}{KC} \log\biggl(1 + \frac{Cr}{\sigma^2 + C\mathbb{E}(Z)}\biggr)\biggr)$$

where $\sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \mathbb{E}f^2(X_i)$ and $K > 0$ is a numerical constant.

This statement is as close as possible to (3.17). With respect to this inequality, the main feature is the deviation property with respect to the mean $\mathbb{E}(Z)$. Such an inequality of course belongs to the concentration phenomenon, with the two parameters $\mathbb{E}(Z)$ and σ^2 which are similar to the Gaussian case (1.24). Bounds on $\mathbb{E}(Z)$ require different tools (chaining, entropy, majorizing measures cf. [L-T]). The proof of Theorem 3.6 is a rather easy consequence of (3.5) for the Gaussian tail. It is a little bit more difficult for the Poissonian part. It is based on the integration of the following differential inequality, consequence of a logarithmic Sobolev inequality for convex functionals,

$$\lambda H'(\lambda) - H(\lambda) \log H(\lambda) \leq \lambda^2 \mathbb{E}\biggl(\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \big(f(X_i) - f(Y_i)\big)^2 \mathrm{e}^{\lambda Z}\biggr) \tag{3.18}$$

for $\lambda \geq 0$, where, as usual, $H(\lambda) = \mathbb{E}(\mathrm{e}^{\lambda Z})$, and where $(Y_i)_{1 \leq i \leq n}$ is an independent copy of the sequence $(X_i)_{1 \leq i \leq n}$ (cf. [Le4]). Integration of this inequality is performed in an improved way in [Mas] yielding sharper numerical constants, that are even optimal in the case of a class consisting of non-negative functions.

Deviations under the mean (i.e. bounds for $\mathbb{P}\{Z \leq \mathbb{E}(Z) - r\}$) may be deduced similarly from the logarithmic Sobolev approach. This was overlooked in [Le4] and we are grateful to P.-M. Samson for pointing out that the argument in [Le4] actually also yields such a conclusion. Namely, since the functions in $\mathcal{F}$ are assumed to be (uniformly) bounded (by C), an elementary inspection of the arguments of [Le4] shows that (3.18) for $\lambda \leq 0$ holds with λ^2 replaced by $\lambda^2 \mathrm{e}^{-2C\lambda}$ in front of the right-hand term. Since the Gaussian bounds (where deviation above or under the mean is really sensible) only require (3.18) for the small values of λ, the same argument is actually enough to conclude to a deviation under the mean. In particular, the bound of Theorem 3.6 also controls $\mathbb{P}\{|Z - \mathbb{E}(Z)| \geq r\}$ (up to numerical constants).

4. MODIFIED LOGARITHMIC SOBOLEV INEQUALITIES FOR LOCAL GRADIENTS

M. Talagrand discovered a few years ago [Ta3] that products of the usual exponential distribution somewhat surprisingly satisfy a concentration property which, in some respect, is stronger than Gaussian concentration. Our first aim here will be to show, following [B-L1], that this result can be seen as a consequence of some appropriate logarithmic Sobolev inequality which we call modified. Modified logarithmic Sobolev inequalities actually appear in various contexts and further examples will be presented, for discrete gradients, in the next chapter. Their main interest is that they tensorize with two parameters on the gradient, one on its supremum norm, and one on the usual quadratic norm. This feature is the appropriate explanation for the concentration property of the exponential measure.

The first paragraph is devoted to the modified logarithmic Sobolev inequality for the exponential measure. We then describe the product properties of modified logarithmic Sobolev inequalities. In the last section, we show, in a general setting, that all measures with a spectral gap (with respect to a local gradient) do satisfy the same modified inequality as the exponential distribution. Most of the results presented here are taken from the joint paper [B-L1] with S. Bobkov.

4.1 The exponential measure

In the paper [Ta3], M. Talagrand proved an isoperimetric inequality for the product measure of the exponential distribution which implies the following concentration property. Let ν^n be the product measure on $\mathbb{R}^n$ when each factor is endowed with the measure ν of density $\frac{1}{2}e^{-|x|}$ with respect to Lebesgue measure. Then, for every Borel set A with $\nu^n(A) \geq \frac{1}{2}$ and every $r \geq 0$,

$$\nu^n\big(A + \sqrt{r}B_2 + rB_1\big) \geq 1 - e^{-r/K} \tag{4.1}$$

for some numerical constant $K > 0$ where B_2 is the Euclidean unit ball and B_1 is the ℓ^1 unit ball in $\mathbb{R}^n$, i.e.

$$B_1 = \bigg\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n; \sum_{i=1}^n |x_i| < 1\bigg\}.$$

A striking feature of (4.1) is that it may be used to improve some aspects of the Gaussian concentration (1.10) especially for cubes [Ta3], [Ta4]. Consider indeed the increasing map $\psi : \mathbb{R} \to \mathbb{R}$ that transform ν into the one-dimensional canonical Gaussian measure γ. It is a simple matter to check that

$$\big|\psi(x) - \psi(y)\big| \leq C \min\big(|x-y|, |x-y|^{1/2}\big), \quad x, y \in \mathbb{R}, \tag{4.2}$$

for some numerical constant $C > 0$. The map $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\Psi(x) = (\psi(x_i))_{1 \leq i \leq n}$ transforms ν^n into γ^n. Consider now Borel a set A of $\mathbb{R}^n$ such that $\gamma^n(A) \geq \frac{1}{2}$. Then

$$\gamma^n\big(\Psi\big(\Psi^{-1}(A) + \sqrt{r}B_2 + rB_1\big)\big) = \nu^n\big(\Psi^{-1}(A) + \sqrt{r}B_2 + rB_1\big) \geq 1 - \mathrm{e}^{-r/K}.$$

However, it follows from (4.2) that

$$\Psi\big(\Psi^{-1}(A) + \sqrt{r}B_2 + rB_1\big) \subset A + C'\sqrt{r}B_2.$$

Thus (4.1) improves upon (1.6). To illustrate the improvement, let

$$A = \big\{x \in \mathbb{R}^n ; \max_{1 \leq i \leq n} |x_i| \leq m\big\}$$

where $m = m(n)$ is chosen so that $\gamma^n(A) \geq \frac{1}{2}$ (and hence $m(n)$ is of order $\sqrt{\log n}$). Then, when $r \geq 1$ is very small compared to $\log n$, it is easily seen that actually

$$\begin{aligned} \Psi\big(\Psi^{-1}(A) + \sqrt{r}B_2 + rB_1\big) &\subset A + C_1\bigg(\frac{\sqrt{r}}{\sqrt{\log n}}\, B_2 + \frac{r}{\sqrt{\log n}}\, B_1\bigg) \\ &\subset A + C_2\sqrt{\frac{r}{\log n}}\, \sqrt{r}B_2. \end{aligned}$$

As for Gaussian concentration, inequality (4.1) may be translated equivalently on functions in the following way (see the end of the section for details). For every real-valued function F on $\mathbb{R}^n$ such that $\|F\|_{\mathrm{Lip}} \leq \alpha$ and

$$\big|F(x) - F(y)\big| \leq \beta \sum_{i=1}^n |x_i - y_i|, \quad x, y \in \mathbb{R}^n,$$

for every $r \geq 0$,

$$\nu^n(F \geq m + r) \leq \exp\bigg(-\frac{1}{K} \min\bigg(\frac{r}{\beta}, \frac{r^2}{\alpha^2}\bigg)\bigg) \tag{4.3}$$

for some numerical constant $K > 0$ where m is either the mean or a median of F for ν_n. Again, this inequality extends in the appropriate sense the case of linear functions F. By Rademacher's theorem, the hypotheses on F are equivalent to saying that F is almost everywhere differentiable with

$$\sum_{i=1}^n |\partial_i F|^2 \leq \alpha^2 \quad \text{and} \quad \max_{1 \leq i \leq n} |\partial_i F| \leq \beta \quad \text{a.e..}$$

Our first aim here will be to present an elementary proof of (4.3) (and thus (4.1)) based on logarithmic Sobolev inequalities. An alternate proof, however close to Talagrand's ideas, has already been given by B. Maurey using inf-convolution [Mau2] (see also [Ta6]). M. Talagrand himself obtained recently another proof as a consequence of a stronger transportation cost inequality [Ta9] (cf. Section 3.3). Our approach is simpler even than the transportation method and is based on the results of Section 2.3. Following the procedure there in case of the exponential distribution would require to determine the appropriate logarithmic Sobolev inequality satisfied by ν^n. We cannot hope for an inequality such as the Gaussian logarithmic Sobolev inequality (2.15) to hold simply because it would imply that linear functions have a Gaussian tail for ν^n. To investigate logarithmic Sobolev inequalities for ν^n, it is enough, by the fundamental product property of entropy, to deal with the dimension one. One first inequality may be deduced from the Gaussian logarithmic Sobolev inequality. Given a smooth function f on $\mathbb{R}$, apply (2.15) in dimension 2 to $g(x,y) = f(\frac{x^2+y^2}{2})$. Let $\tilde{\nu}$ denote the one-sided exponential distribution with density e^{-x} with respect to Lebesgue measure on $\mathbb{R}_+$, and let $\tilde{\nu}^n$ denote the product measure on $\mathbb{R}_+^n$. Then

$$\mathrm{Ent}_{\tilde{\nu}}(f^2) \leq 4 \int x f'(x)^2 d\tilde{\nu}(x).$$

Hence, for every smooth f on $\mathbb{R}_+^n$,

$$\mathrm{Ent}_{\tilde{\nu}^n}(f^2) \leq 4 \int \sum_{i=1}^n x_i \big|\partial_i f(x)\big|^2 d\tilde{\nu}^n(x). \tag{4.4}$$

It does not seem however that this logarithmic Sobolev inequality (4.4) can yield the concentration property (4.3) via the Laplace transform approach of Section 2.3. In a sense, this negative observation is compatible with the fact that (4.3) improves upon some aspects of the Gaussian concentration. We thus have to look for some other version of the logarithmic Sobolev inequality for the exponential distribution. To this aim, let us observe that, at the level of Poincaré inequalities, there are two distinct inequalities. For simplicity, let us deal again only with $n = 1$. The first one, in the spirit of (4.4), is

$$\mathrm{Var}_{\tilde{\nu}}(f) \leq \int x f'(x)^2 d\tilde{\nu}(x). \tag{4.5}$$

This may be shown, either from the Gaussian Poincaré inequality as before, with however a worse constant, or by noting that the first eigenvalue of the Laguerre generator with invariant measure $\tilde{\nu}$ is 1 (cf. [K-S]. By the way, that 4 is the best constant in (4.4) is an easy consequence of our arguments. Namely, if (4.4) holds with a constant $C < 4$, a function f, on $\mathbb{R}_+$ for simplicity, such that $xf'(x)^2 \leq 1$ almost everywhere would be such that $\int \mathrm{e}^{f^2/4} d\tilde{\nu}_1 < \infty$ by Corollary 2.6. But the example of $f(x) = 2\sqrt{x}$ contradicts this consequence. We thus recover in this simple way the main result of [K-S].) The second Poincaré inequality appeared in the work by M. Talagrand [Ta3], actually going back to [Kl], and states that

$$\mathrm{Var}_{\tilde{\nu}}(f) \leq 4\, \mathrm{E}_{\tilde{\nu}}\big({f'}^2\big). \tag{4.6}$$

These two inequalities are not comparable and, in a sense, we are looking for an analogue of (4.6) for entropy.

To introduce this result, let us first recall the proof of (4.6). We will work with the double exponential distribution ν. It is plain that all the results hold, with the obvious modifications, for the one-sided exponential distribution $\tilde{\nu}$. Denote by $\mathcal{L}^n$ the space of all continuous almost everywhere differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that $\int |f| d\nu^n < \infty$, $\int |\nabla f| d\nu^n < \infty$ and $\lim_{x_i \to \pm\infty} \mathrm{e}^{-|x_i|} f(x_1, \ldots, x_i, \ldots x_n) = 0$ for every $i = 1, \ldots, n$ and $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \in \mathbb{R}$. The main argument of the proof is the following simple observation. If $\varphi \in \mathcal{L}^1$, by the integration by parts formula,

$$\int \varphi d\nu = \varphi(0) + \int \mathrm{sgn}(x) \varphi'(x) d\nu(x). \tag{4.7}$$

Lemma 4.1. *For every $f \in \mathcal{L}^1$,*

$$\mathrm{Var}_\nu(f) \leq 4\, \mathrm{E}_\nu\big(f'^2\big).$$

Proof. Set $g(x) = f(x) - f(0)$. Then, by (4.7) and the Cauchy-Schwarz inequality,

$$\mathrm{E}_\nu(g^2) = 2 \int \mathrm{sgn}(x) g'(x) g(x) d\nu(x) \leq 2 \big(\mathrm{E}_\nu(g'^2)\big)^{1/2} \big(\mathrm{E}_\nu(g^2)\big)^{1/2}.$$

Since $\mathrm{Var}_\nu(f) = \mathrm{Var}_\nu(g) \leq \mathrm{E}_\nu(g^2)$, and $g' = f'$, the lemma follows. $\square$

We turn to the corresponding inequality for entropy and the main result of this section.

Theorem 4.2. *For every $0 < c < 1$ and every Lipschitz function f on $\mathbb{R}$ such that $|f'| \leq c < 1$ almost everywhere,*

$$\mathrm{Ent}_\nu\big(\mathrm{e}^f\big) \leq \frac{2}{1-c}\, \mathrm{E}_\nu\big(f'^2 \mathrm{e}^f\big).$$

Note that Theorem 4.2, when applied to functions εf as $\varepsilon \to 0$, implies Lemma 4.1. Theorem 4.2 is the first example of what we will call a modified logarithmic Sobolev inequality. We only use Theorem 4.2 for some fixed valued of c, for example $c = \frac{1}{2}$.

Proof. Changing f into $f + \mathrm{const}$ we may assume that $f(0) = 0$. Since

$$u \log u \geq u - 1, \quad u \geq 0,$$

we have

$$\mathrm{Ent}_\nu\big(\mathrm{e}^f\big) \leq \mathrm{E}_\nu(f\mathrm{e}^f - \mathrm{e}^f + 1).$$

Since $|f'| \leq \lambda < 1$ almost everywhere, the functions e^f, $f\mathrm{e}^f$ and $f^2\mathrm{e}^f$ all belong to $\mathcal{L}^1$. Therefore, by repeated use of (4.7),

$$\mathrm{E}_\nu(f\mathrm{e}^f - \mathrm{e}^f + 1) = \int \mathrm{sgn}(x) f'(x) f(x) \mathrm{e}^{f(x)} d\nu(x)$$

and

$$\mathrm{E}_\nu(f^2\mathrm{e}^f) = 2\int \mathrm{sgn}(x)f'(x)f(x)\mathrm{e}^{f(x)}d\nu(x) + \int \mathrm{sgn}(x)f'(x)f(x)^2\mathrm{e}^{f(x)}d\nu(x).$$

By the Cauchy-Schwarz inequality and the assumption on f',

$$\mathrm{E}_\nu(f^2\mathrm{e}^f) \le 2\big(\mathrm{E}_\nu\big({f'}^2\mathrm{e}^f\big)\big)^{1/2}\big(\mathrm{E}_\nu(f^2\mathrm{e}^f)\big)^{1/2} + c\,\mathrm{E}_\nu(f^2\mathrm{e}^f)$$

so that

$$\mathrm{E}_\nu(f^2\mathrm{e}^f) \le \left(\frac{2}{1-c}\right)^2 \mathrm{E}_\nu\big({f'}^2\mathrm{e}^f\big).$$

Now, by the Cauchy-Schwarz inequality again,

$$\begin{aligned}\mathrm{Ent}_{\nu_1}\big(\mathrm{e}^f\big) &\le \int \mathrm{sgn}(x)f'(x)f(x)\mathrm{e}^{f(x)}d\nu_1(x)\\ &\le \big(\mathrm{E}_\nu\big({f'}^2\mathrm{e}^f\big)\big)^{1/2}\big(\mathrm{E}_\nu(f^2\mathrm{e}^f)\big)^{1/2} \le \frac{2}{1-c}\,\mathrm{E}_\nu\big({f'}^2\mathrm{e}^f\big)\end{aligned}$$

which is the result. Theorem 4.2 is established. □

We are now ready to describe the application to Talagrand's concentration inequality (4.3). As a consequence of Theorem 4.2 and of the product property of entropy (Proposition 2.2), for every smooth enough function F on $\mathbb{R}^n$ such that $\max_{1\le i\le n}|\partial_i F| \le 1$ almost everywhere and every λ, $|\lambda| \le c < 1$,

$$\mathrm{Ent}_{\nu^n}\big(\mathrm{e}^{\lambda F}\big) \le \frac{2\lambda^2}{1-c}\,\mathrm{E}_{\nu^n}\bigg(\sum_{i=1}^n(\partial_i F)^2\mathrm{e}^{\lambda F}\bigg). \tag{4.8}$$

Let us take for simplicity $c = \frac{1}{2}$ (although $c < 1$ might improve some numerical constants below). Assume now moreover that $\sum_{i=1}^n(\partial_i F)^2 \le \alpha^2$ almost everywhere. Then, by (4.8),

$$\mathrm{Ent}_{\nu^n}(\mathrm{e}^{\lambda F}) \le 4\alpha^2\lambda^2\mathrm{E}_{\nu^n}(\mathrm{e}^{\lambda F})$$

for every $|\lambda| \le \frac{1}{2}$. As a consequence of Corollary 2.11, we get that

$$\nu^n\big(F \ge \mathrm{E}_{\nu^n}(F) + r\big) \le \exp\bigg(-\frac{1}{4}\,\min\Big(r, \frac{r^2}{4\alpha^2}\Big)\bigg) \tag{4.9}$$

for every $r \ge 0$. By homogeneity, this inequality amounts to (4.3) (with $K = 16$) and our claim is proved. As already mentioned, we have a similar result for the one-sided exponential measure.

To complete this section, let us sketch the equivalence between (4.1) and (4.3). (Although we present the argument for ν^n only, it extends to more general situations, as will be used in the next section.) To see that (4.1) implies (4.3), simply apply (4.1) to $A = \{F \le m\}$ where m is a median of F for ν^n and note that

$$A + \sqrt{r}B_2 + rB_1 \subset \big\{F \le m + \alpha\sqrt{r} + \beta r\big\}.$$

Using a routine argument (cf. the end of Section 1.3), the deviation inequality (4.3) from either the median or the mean are equivalent up to numerical constants (with possibly a further constant in front of the exponential function). Now starting from (4.3) with m the mean for example, consider, for $A \subset \mathbb{R}^n$ and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$F_A(x) = \inf_{a \in A} \sum_{i=1}^{n} \min\big(|x_i - a_i|, |x_i - a_i|^2\big).$$

For $r > 0$, set then $F = \min(F_A, r)$. We have $\sum_{i=1}^n |\partial_i F|^2 \leq 4r$ and $\max_{1 \leq i \leq n} |\partial_i F| \leq 2$ almost everywhere. Indeed, it is enough to prove this result for $G = \min(G_a, r)$ for every fixed a where

$$G_a(x) = \sum_{i=1}^{n} \min\big(|x_i - a_i|, |x_i - a_i|^2\big).$$

Now, almost everywhere, and for every $i = 1, \ldots, n$, $|\partial_i G_a(x)| \leq 2|x_i - a_i|$ if $|x_i - a_i| \leq 1$ whereas $|\partial_i G_a(x)| \leq 1$ if $|x_i - a_i| > 1$. Therefore, $\max_{1 \leq i \leq n} |\partial_i G_a(x)| \leq 2$ and

$$\sum_{i=1}^{n} \big|\partial_i G_a(x)\big|^2 \leq 4 \sum_{i=1}^{n} \min\big(|x_i - a_i|, |x_i - a_i|^2\big) = 4 G_a(x)$$

which yields the announced claim. Now, if $\nu^n(A) \geq \frac{1}{2}$,

$$\mathrm{E}_{\nu^n}(F) \leq r\big(1 - \nu^n(A)\big) \leq \frac{r}{2}\,.$$

It then follows from (4.3) that

$$\nu^n(F_A \geq r) = \nu^n(F \geq r) \leq \nu^n\big(F \geq \mathrm{E}_{\nu^n}(F) + \tfrac{r}{2}\big) \leq \mathrm{e}^{-r/16K}.$$

Since $\{F_A \leq r\} \subset A + \sqrt{r} B_2 + r B_1$, the result follows.

4.2 Modified logarithmic Sobolev inequalities

The inequality put forward in Theorem 4.2 for the exponential measure is a first example of what we call modified logarithmic Sobolev inequalities. In order to describe this notion in some generality, we take again the general setting of Part 2. Let thus $(X, \mathcal{B}, \mu)$ be a probability space, and let $\mathcal{A}$ be a subset of $\mathrm{L}^1(\mu)$. Consider a "gradient" operator Γ on $\mathcal{A}$ such that $\Gamma(f) \geq 0$ and $\Gamma(\lambda f) = \lambda^2 \Gamma(f)$ for every $f \in \mathcal{A}$ and $\lambda \in \mathbb{R}$. Examples are $\Gamma(f) = |\nabla f|^2$ for a generalized modulus of gradient (2.3), or $\Gamma(f) = |Df|^2$ for a discrete gradient (such as (2.9)).

Definition 4.3. *We say that μ satisfies a modified logarithmic Sobolev inequality with respect to Γ (on $\mathcal{A}$) if there is a function $B(\lambda) \geq 0$ on $\mathbb{R}_+$ such that, whenever $\|\Gamma(f)\|_\infty^{1/2} \leq \lambda$,*

$$\mathrm{Ent}_\mu(\mathrm{e}^f) \leq B(\lambda)\, \mathrm{E}_\mu\big(\Gamma(f)\, \mathrm{e}^f\big)$$

for all f in $\mathcal{A}$ such that $\mathrm{E}_\mu(\mathrm{e}^f) < \infty$.

According to Theorem 4.2, the exponential measure ν on the line satisfies a modified logarithmic Sobolev inequality with respect to the usual gradient with $B(\lambda)$ bounded for the small values of λ. On the other hand, the Gaussian measure γ satisfies a modified logarithmic Sobolev inequality with $B(\lambda) = \frac{1}{2}$, $\lambda \geq 0$.

Definition 4.3 might appear very similar to the inequalities investigated via Proposition 2.9. Actually, Definition 4.3 implies that

$$\mathrm{Ent}_\mu(\mathrm{e}^f) \leq \lambda^2 B(\lambda)\, \mathrm{E}_\mu(\mathrm{e}^f)$$

for every f with $\|\Gamma(f)\|_\infty \leq \lambda$. In particular, if $B(\lambda)$ is bounded for the small values of λ, Lipschitz functions will have an exponential tail according to Corollary 2.11.

The main new feature here is that the modified logarithmic Sobolev inequality of Definition 4.3 tensorizes in terms of two parameters rather than only the Lipschitz bound. This property is summarized in the next proposition which is an elementary consequence of the product property of entropy (Proposition 2.2).

Let $(X_i, \mathcal{B}_i, \mu_i)$, $i = 1, \ldots, n$, be probability spaces and denote by $P = \mu_1 \otimes \cdots \otimes \mu_n$ on the product space $X = X_1 \times \cdots \times X_n$. Consider operators Γ_i on classes $\mathcal{A}_i$, $i = 1, \ldots, n$. If f is a function on the product space, for each i, f_i is the function f depending on the i-th variable with the other coordinates fixed.

Proposition 4.4. *Assume that for every f on $(X_i, \mathcal{B}_i, \mu_i)$ such that $\|\Gamma_i(f)\|_\infty^{1/2} \leq \lambda$,*

$$\mathrm{Ent}_{\mu_i}(\mathrm{e}^f) \leq B(\lambda)\, \mathrm{E}_{\mu_i}\big(\Gamma_i(f)\, \mathrm{e}^f\big),$$

$i = 1, \ldots, n$. Then, for every f on the product space such that $\max_{1 \leq i \leq n} \|\Gamma_i(f_i)\|_\infty^{1/2} \leq \lambda$,

$$\mathrm{Ent}_P(\mathrm{e}^f) \leq B(\lambda)\, \mathrm{E}_P\bigg(\sum_{i=1}^n \Gamma_i(f_i)\, \mathrm{e}^f\bigg).$$

According to the behavior of $B(\lambda)$, this proposition yields concentration properties in terms of two parameters,

$$\max_{1 \leq i \leq n} \|\Gamma_i(f_i)\|_\infty^{1/2} \quad \text{and} \quad \bigg\|\sum_{i=1}^n \Gamma_i(f_i)\bigg\|_\infty.$$

For example, if $B(\lambda) \leq c$ for $0 \leq \lambda \leq \lambda_0$, following the proof of (4.9), the product measure P will satisfy the same concentration inequality as the one for the exponential measure (4.3). In the next chapter, we investigate cases such as $B(\lambda) \leq c\mathrm{e}^{d\lambda}$, $\lambda \geq 0$, related to the Poisson measure. Rather than to discuss some further abstract result according to the behavior of $B(\lambda)$ (in the spirit of Corollaries 2.11 and 2.12), we refer to Corollary 4.6 and Theorem 5.5 for examples of applications.

4.3 Poincaré inequalities and modified logarithmic Sobolev inequalities

In this section, we show that the concentration properties of the exponential measure described in Section 4.1 is actually shared by all measures satisfying a Poincaré inequality (with respect to a local gradient). More precisely, we show, following [B-L1], that every such measure satisfies the modified logarithmic Sobolev inequality of Theorem 4.2.

Let thus $|\nabla f|$ be a generalized modulus of gradient on a metric space (X, d), satisfying thus the chain rule formula (2.5). Throughout this paragraph, we assume that μ is a probability measure on X equipped with the Borel σ-field $\mathcal{B}$ such that for some $C > 0$ and all f in $\mathrm{L}^2(\mu)$,

$$\mathrm{Var}_\mu(f) \leq C\,\mathrm{E}_\mu\big(|\nabla f|^2\big). \tag{4.10}$$

We already know from Proposition 2.13 that such a spectral gap inequality implies exponential integrability of Lipschitz functions. We actually show that it also implies a modified logarithmic Sobolev inequality which yields concentration properties for the product measures μ^n.

Theorem 4.5. *For any function f on X such that $\|\nabla f\|_\infty \leq \lambda < 2/\sqrt{C}$,*

$$\mathrm{Ent}_\mu\big(\mathrm{e}^f\big) \leq B(\lambda)\,\mathrm{E}_\mu\big(|\nabla f|^2 \mathrm{e}^f\big)$$

where

$$B(\lambda) = \frac{C}{2}\left(\frac{2 + \lambda\sqrt{C}}{2 - \lambda\sqrt{C}}\right)^2 \mathrm{e}^{\sqrt{5C}\lambda}.$$

We refer to the paper [B-L1] for the proof of Theorem 4.5.

Now, $B(\lambda)$ is uniformly bounded for the small values of λ, for example $B(\lambda) \leq 3\mathrm{e}^5 C/2$ when $\lambda \leq 1/\sqrt{C}$. As a corollary, we obtain, following the proof of (4.9) and the discussion on Proposition 4.4, a concentration inequality of Talagrand's type for the product measure μ^n of μ on X^n. If f is a function on the product space X^n, denote by $|\nabla_i f|$ the length of the gradient with respect to the i-th coordinate.

Corollary 4.6. *Denote by μ^n the product of μ on X^n. Then, for every function F on X^n such that*

$$\sum_{i=1}^n |\nabla_i F|^2 \leq \alpha^2 \quad \text{and} \quad \max_{1 \leq i \leq n} |\nabla_i F| \leq \beta$$

μ-almost everywhere, $\mathrm{E}_{\mu^n}(|F|) < \infty$ and

$$\mu^n\big(F \geq \mathrm{E}_{\mu^n}(F) + r\big) \leq \exp\left(-\frac{1}{K}\min\left(\frac{r}{\beta}, \frac{r^2}{\alpha^2}\right)\right)$$

where $K > 0$ only depends on the constant C in the Poincaré inequality (4.10).

One may obtain a similar statement for products of possibly different measures μ with a uniform lower bound on the constants in the Poincaré inequalities (4.10).

Following the argument at the end of Section 4.1, Corollary 4.6 may be turned into an inequality on sets such as (4.1). More precisely, if $\mu^n(A) \geq \frac{1}{2}$, for every $r \geq 0$ and some numerical constant $K > 0$,

$$\mu^n\big(F_A^h \geq r\big) \leq \mathrm{e}^{-r/K}$$

where $h(x,y) = \min(d(x,y), d(x,y)^2)$, $x, y \in X$, and, for $x = (x_1, \dots, x_n) \in X^n$ and $A \subset X^n$,

$$F_A^h(x) = \inf_{a \in A} \sum_{i=1}^n h(x_i, a_i).$$

Using analogues of the norm $|||\cdot|||_\infty$, Theorem 4.5 and Corollary 4.6 have been recently extended in [H-T] to the example of the invariant measure of a reversible Markov chain on a finite state space. The main idea consists in showing that the various uses of the chain rule formula in the proof of Theorem 4.5 may be properly extended to this case (see also [A-S] for extensions of the chain rule formula).

Let us observe that for the case of the exponential measure ν, $C = 4$ by Lemma 4.1 so that, for $\lambda < 1$,

$$B(\lambda) = 2\left(\frac{1+\lambda}{1-\lambda}\right)^2 \mathrm{e}^{2\sqrt{5}\lambda}$$

which is somewhat worse than the constant given by Theorem 4.2.

In any case, an important feature of the constant $B(\lambda)$ of Theorem 4.5 is that $B(\lambda) \to C/2$ as $\lambda \to 0$. In particular (and as in Theorem 4.2), the modified logarithmic Sobolev inequality of Theorem 4.5 implies the Poincaré inequality (4.10) by applying it to functions εf with $\varepsilon \to 0$. Poincaré inequality and the modified logarithmic Sobolev inequality of Theorem 4.5 are thus equivalent.

On the other hand, let us consider the case of the canonical Gaussian measure γ on the real line for which, by (2.16),

$$\mathrm{Var}_\gamma(f) \le \mathrm{E}_\gamma\big(f'^2\big).$$

Let φ be a smooth function on $\mathbb{R}$, for example C^2 with bounded derivatives. Apply the multidimensional analogue (Proposition 4.4) of Theorem 4.5 to the functions

$$f(x) = \varphi\Big(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\Big), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

for which $\max_{1\le i\le n} |\partial_i f| \le \|\varphi\|_{\mathrm{Lip}}/\sqrt{n} = \beta_n < 2$ for n large enough. By the rotational invariance of Gaussian measures, and since $\beta_n \to 0$, we get in the limit

$$\mathrm{Ent}_\gamma(\mathrm{e}^\varphi) \le \frac{1}{2}\,\mathrm{E}_\gamma\big(\varphi'^2 \mathrm{e}^\varphi\big)$$

that is (after the change of functions $\mathrm{e}^\varphi = g^2$) Gross's logarithmic Sobolev inequality (2.15) with optimal constant. Therefore, for the Gaussian measure, Poincaré and logarithmic Sobolev inequalities are in a sense equivalent.

5. MODIFIED LOGARITHMIC SOBOLEV INEQUALITIES IN DISCRETE SETTINGS

We investigate here concentration and logarithmic Sobolev inequalities for discrete gradients which typically do not satisfy a chain rule formula. One such example considered here is $Df(x) = f(x+1) - f(x)$, $x \in \mathbb{N}$. With respect to such gradient, natural measures such as Poisson measures do not satisfy a logarithmic Sobolev inequality in its classical formulation, but rather some modified inequality. Following the recent works [B-L2] and [G-R] on the subject, we study mainly here Poisson type logarithmic Sobolev inequalities and their related concentration properties. The results of this part are taken from the paper [B-L2] with S. Bobkov.

5.1 Logarithmic Sobolev inequality for Bernoulli and Poisson measures

As was presented in Section 2.2, in his seminal 1975 paper, L. Gross [Gr1] proved a logarithmic Sobolev inequality on the two-point space. Namely, let μ be the uniform measure on $\{0,1\}$. Then, for any f on $\{0,1\}$,

$$\mathrm{Ent}_\mu(f^2) \le \frac{1}{2}\,\mathrm{E}_\mu\big(|Df|^2\big) \tag{5.1}$$

where

$$\big|Df(x)\big| = \big|f(1) - f(0)\big| = \big|f(x+1) - f(x)\big|$$

(x modulo 2). It is easily seen that the constant $\frac{1}{2}$ is optimal.

The question of the best constant in the previous logarithmic Sobolev inequality for non-symmetric Bernoulli measure was settled seemingly only quite recently. Let μ_p be the Bernoulli measure on $\{0,1\}$ with $\mu_p(\{1\}) = p$ and $\mu_p(\{0\}) = q = 1-p$. Then, for any f on $\{0,1\}$,

$$\mathrm{Ent}_{\mu_p}(f^2) \le pq\,\frac{\log p - \log q}{p-q}\,\mathrm{E}_{\mu_p}\big(|Df|^2\big). \tag{5.2}$$

The constant is optimal, and is equal to $\frac{1}{2}$ when $p = q = \frac{1}{2}$. This result is mentioned in [H-Y] without proof, and worked out in [D-SC]. A simple proof, due S. Bobkov, is presented in the notes [SC2]. O. Rothaus mentioned to the authors of [D-SC] that he computed this constant several years back from now. The main feature of this constant is that, when $p \neq q$, it significantly differs from the spectral gap given by the inequality

$$\mathrm{Var}_{\mu_p}(f) \leq pq\, \mathrm{E}_{\mu_p}\big(|Df|^2\big). \tag{5.3}$$

Although inequality (5.2) is optimal, it presents a number of weak points. First of all, the product property of entropy which allows us, together with the central limit theorem, to deduce the logarithmic Sobolev inequality for Gaussian measures from the one for Bernoulli is optimal in the symmetric case. As soon as $p \neq q$, the central limit theorem on the basis of (5.2) only yields the Gaussian logarithmic Sobolev inequality (2.15) with a worse constant. A second limit theorem of interest is of course the Poisson limit. However, after tensorization, (5.2) cannot yield a logarithmic Sobolev inequality for Poisson measures. (Although the constant in (5.2) is bounded as $p \to 0$, we would need it to be of the order of p for $p \to 0$.) There is of course a good reason at that, namely that Poisson measures do not satisfy logarithmic Sobolev inequalities! This is well known to a number of people but let us briefly convince ourselves of this claim. Denote thus by π_θ the Poisson measure on $\mathbb{N}$ with parameter $\theta > 0$ and let us assume that, for some constant $C > 0$, and all f, say bounded, on $\mathbb{N}$,

$$\mathrm{Ent}_{\pi_\theta}(f^2) \leq C\, \mathrm{E}_{\pi_\theta}\big(|Df|^2\big) \tag{5.4}$$

where here $Df(x) = f(x+1) - f(x)$, $x \in \mathbb{N}$. Apply (5.4) to the indicator function of the interval $[k+1, \infty)$ for each $k \in \mathbb{N}$. We get

$$-\pi_\theta\big([k+1,\infty)\big) \log \pi_\theta\big([k+1,\infty)\big) \leq C\pi_\theta\big(\{k\}\big)$$

which is clearly impossible as k goes to infinity. Similarly, (5.4) cannot hold with the addition of an extra $C\mathrm{E}_\theta(f^2)$ on the right-hand side. It is important for the further developments to notice, according to [G-R], that the exponential integrability results of Section 2.3 with the norm $|||\cdot|||_\infty$ cannot be used at this point to rule out (5.4). Indeed, (5.4) implies via (2.8) that

$$\mathrm{Ent}_{\pi_\theta}(\mathrm{e}^f) \leq \frac{C}{2}\, |||f|||_\infty^2 \mathrm{E}_{\pi_\theta}(\mathrm{e}^f).$$

By (2.11),

$$|||f|||_\infty^2 = \sup_{x \in \mathbb{N}} \Big(Df(x)^2 + \frac{x}{\theta}\, Df(x-1)^2 \Big).$$

As an application of Corollary 2.4, if F on $\mathbb{N}$ is such that $|||F|||_\infty \leq 1$, we would conclude from the logarithmic Sobolev inequality (5.4) that $\mathrm{E}_{\pi_\theta}(\mathrm{e}^{\alpha F^2}) < \infty$ for some $\alpha > 0$. But now, if $|||F|||_\infty \leq 1$, then $DF(x) \leq \sqrt{\frac{\theta}{x+1}}$ for every x. This directly implies that $\mathrm{E}_{\pi_\theta}(\mathrm{e}^{\alpha F^2}) < \infty$ for every α which thus would not contradict Corollary 2.4. The norm $|||F|||_\infty$ is therefore not well adapted to our purposes here, and we will rather consider $\sup_{x \in \mathbb{N}} |DF(x)|$ under which we will describe exponential integrability of Poisson type.

One may therefore be led to consider some variations of inequality (5.2) that could behave better under the preceding limits, in particular one could think of modified logarithmic Sobolev inequalities. However, we follow a somewhat different route and turn to an alternate variation of possible own interest.

An equivalent formulation of the Gaussian logarithmic Sobolev inequality (2.15), on the line for simplicity, is that, for any smooth f on $\mathbb{R}$ with strictly positive values,

$$\mathrm{Ent}_\gamma(f) \le \frac{1}{2}\,\mathrm{E}_\gamma\Big(\frac{1}{f}\,f'^2\Big). \tag{5.5}$$

That (5.5) is equivalent to (2.15) simply follows from a change of functions together with the chain rule formula for the usual gradient on $\mathbb{R}$. Of course, such a change may not be performed equivalently on discrete gradients, so that there is some interest to study an inequality such as

$$\mathrm{Ent}_{\mu_p}(f) \le C\,\mathrm{E}_{\mu_p}\Big(\frac{1}{f}\,|Df|^2\Big) \tag{5.6}$$

on $\{0,1\}$ for the Bernoulli measure μ_p and to ask for the best constant C as a function of p. Our first result will be to show that the best constant C in (5.6) is pq. The behavior in p is thus much better than in (5.2) as $p \to 0$ or 1, and will allow us to derive a modified logarithmic Sobolev inequality for Poisson measure in the limit. The following is taken from the recent work [B-L2]. An alternate proof of Theorem 5.1 and Corollary 5.3 below using the Γ_2 calculus of [Ba1], [Ba2] may be found in [A-L].

For any $n \ge 1$, we denote by μ_p^n the product measure of μ_p on $\{0,1\}^n$. If f is a function on $\{0,1\}^n$, and $x = (x_1, \dots, x_n) \in \{0,1\}^n$, set

$$|Df|^2(x) = \sum_{i=1}^n \big|f(x+e_i) - f(x)\big|^2$$

where $(e_1, \dots, e_n)$ is the canonical basis of $\mathbb{R}^n$ and the addition is modulo 2. p is arbitrary in $[0,1]$, and $q = 1-p$.

Theorem 5.1. *For any positive function f on $\{0,1\}^n$,*

$$\mathrm{Ent}_{\mu_p^n}(f) \le pq\,\mathrm{E}_{\mu_p^n}\Big(\frac{1}{f}\,|Df|^2\Big).$$

Proof. By the product property of entropy, it is enough to deal with the case $n=1$. The proof is based on the following calculus lemma.

Lemma 5.2 *Consider a function*

$$U(p) = \mathrm{Ent}_{\mu_p}(f) - pq\,\mathrm{E}_{\mu_p}(g), \quad 0 \le p \le 1,$$

where f and g are arbitrary non-negative functions on $\{0,1\}$. Then $U(p) \le 0$ for every p if and only if

$$U'(0) \le 0 \le U'(1). \tag{5.7}$$

If, additionally, $f(0) \geq f(1)$ and $g(0) \geq g(1)$ (respectively $f(0) \leq f(1)$ and $g(0) \leq g(1)$), then the condition (5.7) may be weakened into $U'(0) \leq 0$ (respectively $U'(1) \geq 0$).

Proof. Set $a = f(1)$, $b = f(0)$, $\alpha = g(1)$, $\beta = g(0)$, so that

$$U(p) = (pa \log a + qb \log b) - (pa + qb) \log(pa + qb) - pq(p\alpha + q\beta).$$

Since $U(0) = U(1) = 0$, the condition (5.7) is necessary for U to be non-positive. Now, assume (5.7) is fulfilled. Differentiating in p, we have

$$\begin{aligned} U'(p) &= (a \log a - b \log b) - (a - b)\big(\log(pa + qb) + 1\big) \\ &\qquad\qquad\qquad + (p - q)(p\alpha + q\beta) - pq(\alpha - \beta), \\ U''(p) &= -(a - b)^2 (pa + qb)^{-1} + 2(p\alpha + q\beta) + 2(p - q)(\alpha - \beta), \\ U'''(p) &= (a - b)^3 (pa + qb)^{-2} + 6(\alpha - \beta), \\ U''''(p) &= -2(a - b)^4 (pa + qb)^{-3}. \end{aligned}$$

Since $U'''' \leq 0$, U'' is concave. Hence, formally three situations are possible.

1) $U'' \geq 0$ on [0,1]. In this case, U is convex and thus $U \leq 0$ on [0,1] in view of $U(0) = U(1) = 0$.

2) $U'' \leq 0$ on [0,1]. By (5.7), this case is not possible unless U is identically 0.

3) For some $0 \leq p_0 < p_1 \leq 1$, $U'' \leq 0$ on $[0, p_0]$, $U'' \geq 0$ on $[p_0, p_1]$, and $U'' \leq 0$ on $[p_1, 1]$. In this case, U is concave on $[0, p_0]$, and, due to the assumption $U'(0) \leq 0$, one may conclude that U is non-increasing on $[0, p_0]$. In particular, $U \leq 0$ on $[0, p_0]$. It is then necessary that $U(p_1) \leq 0$. Indeed, U is concave on $[p_1, 1]$, hence the assumption $U(p_1) > 0$ together with $U(1) = 0$ would imply $U'(1) < 0$ which contradicts (5.7). As a result, by convexity of U on $[p_0, p_1]$, we get $U \leq 0$ on $[p_0, p_1]$. At last, $U \leq 0$ on $[p_1, 1]$, since U is concave on $[p_1, 1]$, $U(p_1) \leq 0$ and $U'(1) \geq 0$ (in particular, U is non-decreasing on this interval). The first part of Lemma 2 is thus proved.

We turn to the second part. Again, since $U(0) = U(1) = 0$, any of the conditions $U'(0) \leq 0$ or $U'(1) \geq 0$ is necessary for U to be non-positive on $[0, 1]$. Now, assume that $a \geq b$, $\alpha \geq \beta$, and $U'(0) \leq 0$ (the other case is similar). Then $U''' \geq 0$, and hence U'' is non-decreasing on [0,1]. Again three cases are formally possible.

1) $U'' \geq 0$ on [0,1]. In this case, U is convex, and thus $U \leq 0$ on [0,1] in view of $U(0) = U(1) = 0$.

2) $U'' \leq 0$ on [0,1]. This can only occur if $U \equiv 0$.

3) For some $0 \leq p_0 \leq 1$, $U'' \leq 0$ on $[0, p_0]$ and $U'' \geq 0$ on $[p_0, 1]$. In this case, U is concave on $[0, p_0]$, and, due to the fact that $U'(0) \leq 0$, one may conclude that U is non-increasing on $[0, p_0]$. In particular $U \leq 0$ on $[0, p_0]$. At last, $U \leq 0$ on $[p_0, 1]$ since U is convex on this interval and $U(p_0) \leq 0$ and $U(1) = 0$. Lemma 2 is established. □

We turn to the proof of Theorem 5.1. Note first the following. In the notation of the proof of Lemma 5.2, set

$$R(a, b) = a \log \tfrac{a}{b} - (a - b).$$

Clearly, $R(a,b) \geq 0$ for all $a, b > 0$. Then,

$$U'(0) \leq 0 \quad \text{if and only if} \quad \beta \geq R(a,b) \tag{5.8}$$

while

$$U'(1) \geq 0 \quad \text{if and only if} \quad \alpha \geq R(b,a) \tag{5.9}$$

Fix f with stricly positive values on $\{0,1\}$. Apply then Lemma 5.2 to $g = \delta/f$, $\delta > 0$. According to (5.8) and (5.9), the optimal value of $\delta > 0$ in the inequality

$$\mathrm{Ent}_{\mu_p}(f) \leq \delta pq \, \mathrm{E}_{\mu_p}\Big(\frac{1}{f}\Big) \tag{5.10}$$

provided $p \in [0,1]$ is arbitrary is given by

$$\delta = \max\big\{bR(a,b), aR(b,a)\big\},$$

where $a = f(1)$, $b = f(0)$. By symmetry, one may assume that $a \geq b \geq 0$. Then, $bR(a,b) \leq aR(b,a)$. Indeed, for fixed $b > 0$, the function $\rho(a) = aR(b,a) - bR(a,b)$ has derivative $\rho'(a) = 2R(b,a) \geq 0$. Hence, $\rho(a) \geq \rho(b) = 0$. Thus, $\delta = aR(b,a)$, $a > b > 0$. Now, fixing $b > 0$, consider

$$u(a) = aR(b,a) = a\big(b \log \tfrac{b}{a} - (b-a)\big), \quad a > b.$$

We have $u'(a) = b \log \frac{b}{a} - 2(b-a)$, thus $u(b) = u'(b) = 0$ and, for every $a > 0$,

$$u''(a) = 2 - \frac{b}{a} \leq 2.$$

Hence, by a Taylor expansion, denoting by a_0 some middle point between a and b, we get

$$\delta = u(a) = u(b) + u'(b)(a-b) + \frac{1}{2}\, u''(a_0)(a-b)^2 \leq \Big(1 - \frac{b}{2a}\Big)(a-b)^2.$$

Therefore, $\delta \leq (a-b)^2 = |f(1) - f(0)|^2$ in (5.10) which is the result. Theorem 5.1 is established. $\square$

Observe that in the process of the proof of Theorem 5.1, we actually proved a somewhat better inequality. Namely, for any positive function f on $\{0,1\}$,

$$\mathrm{Ent}_{\mu_p}(f) \leq pq\Big(1 - \frac{1}{2M(f)}\Big)\, \mathrm{E}_{\mu_p}\Big(\frac{1}{f}\, |Df|^2\Big)$$

where

$$M(f) = \max\bigg\{\frac{f(1)}{f(0)}\, , \, \frac{f(0)}{f(1)}\bigg\}.$$

By the product property of entropy, for any f with strictly positive values on $\{0,1\}^n$,

$$\mathrm{Ent}_{\mu_p^n}(f) \leq pq\Big(1 - \frac{1}{2M(f)}\Big)\, \mathrm{E}_{\mu_p^n}\Big(\frac{1}{f}\, |Df|^2\Big) \tag{5.11}$$

where

$$M(f) = \max_{x \in \{0,1\}^n} \max_{1 \leq i \leq n} \frac{f(x+e_i)}{f(x)}\,.$$

As announced, the logarithmic Sobolev inequality of Theorem 5.1 may be used in the limit to yield a logarithmic Sobolev inequality for Poisson measure. Take namely φ on $\mathbb{N}$ such that $0 < c \leq \varphi \leq C < \infty$ and apply Theorem 1 to

$$f(x) = f(x_1, \ldots, x_n) = \varphi(x_1 + \cdots + x_n), \quad x = (x_1, \ldots, x_n) \in \{0,1\}^n,$$

with this time $p = \frac{\theta}{n}$, $\theta > 0$ (for every n large enough). Then, setting $S_n = x_1 + \cdots + x_n$,

$$|Df|^2(x) = (n - S_n)\big[\varphi(S_n + 1) - \varphi(S_n)\big]^2 + S_n\big[\varphi(S_n) - \varphi(S_n - 1)\big]^2.$$

Therefore,

$$\begin{aligned} \mathrm{Ent}_{\mu_p^n}\big(\varphi(S_n)\big) \leq \frac{\theta}{n}\Big(1 - \frac{\theta}{n}\Big)\mathrm{E}_{\mu_p^n}\bigg(\frac{1}{\varphi(S_n)}\Big((n{-}S_n)\big[\varphi(S_n + 1) - \varphi(S_n)\big]^2 \\ + S_n\big[\varphi(S_n) - \varphi(S_n - 1)\big]^2\Big)\bigg). \end{aligned}$$

The distribution of S_n under $\mu_{\theta/n}^n$ converges to π_θ. Using that $0 < c \leq \varphi \leq C < \infty$ and that $\frac{1}{n}\mathrm{E}_{\mu_p^n}(S_n) \to 0$, we immediately obtains the following corollary.

Corollary 5.3. *For any f on $\mathbb{N}$ with strictly positive values,*

$$\mathrm{Ent}_{\pi_\theta}(f) \leq \theta\, \mathrm{E}_{\pi_\theta}\Big(\frac{1}{f}\,|Df|^2\Big)$$

where we recall that here $Df(x) = f(x+1) - f(x)$, $x \in \mathbb{N}$.

The example of $f(x) = \mathrm{e}^{-cx}$, $x \in \mathbb{N}$, as $c \to \infty$ shows that one cannot expect a better factor of θ in the preceding corollary.

Theorem 5.1 may also be used to imply the Gaussian logarithmic Sobolev inequality up to a constant 2. Actually, using the refined inequality (5.11), we can reach the optimal constant. Let indeed $\varphi > 0$ be smooth enough on $\mathbb{R}$, for example C^2 with bounded derivatives, and apply (5.11) to

$$f(x_1, \ldots, x_n) = \varphi\Big(\frac{x_1 + \cdots + x_n - np}{\sqrt{npq}}\Big)$$

for fixed p, $0 < p < 1$. Under the smoothness properties on φ, it is easily seen that $M(f) \to 1$ as $n \to \infty$. Therefore, by the Gaussian central limit theorem, we deduce in the classical way inequality (5.5) for φ. Changing φ into φ^2, and using a standard approximation procedure, we get Gross's logarithmic Sobolev inequality (2.15) with its best constant. Another consequences of this sharp form are the spectral gap inequalities for μ_p^n and π_θ. Applying (5.11) to $1 + \varepsilon f$ and letting ε go to 0, we get, since $M(1 + \varepsilon f) \to 1$,

$$\mathrm{Var}_{\mu_p^n}(f) \leq pq\, \mathrm{E}_{\mu_p^n}\big(|Df|^2\big) \tag{5.12}$$

and

$$\mathrm{Var}_{\pi_\theta}(f) \leq \theta\, \mathrm{E}_{\pi_\theta}\big(|Df|^2\big). \tag{5.13}$$

5.2 Modified logarithmic Sobolev inequalities and Poisson tails

In analogy with the Gaussian concentration properties of Section 2.3, the logarithmic Sobolev inequalities of the type of those of Theorem 5.1 and Corollary 5.3 entail some information on the Poisson behavior of Lipschitz functions. For simplicity, we only deal with the case of measures on $\mathbb{N}$. According to the preceding section, the results below apply in particular to the Poisson measure π_θ.

Let μ be a probability measure on $\mathbb{N}$ such that, for some constant $C > 0$,

$$\mathrm{Ent}_\mu(f) \leq C\, \mathrm{E}_\mu\Big(\frac{1}{f}\,|Df|^2\Big) \tag{5.14}$$

for all functions f on $\mathbb{N}$ with positive values, where $Df(x) = f(x+1) - f(x)$, $x \in \mathbb{N}$. As usual, we would like to apply (5.14) to e^f. In this discrete setting, $|D(\mathrm{e}^f)| \leq |Df|\mathrm{e}^f$ is obviously false in general. However,

$$\big|D(\mathrm{e}^f)\big| \leq |Df|\,\mathrm{e}^{|Df|}\mathrm{e}^f. \tag{5.15}$$

Indeed, for every $x \in \mathbb{N}$,

$$\big|D(\mathrm{e}^f)(x)\big| = \big|\mathrm{e}^{f(x+1)} - \mathrm{e}^{f(x)}\big| = \big|Df(x)\big|\mathrm{e}^\tau$$

for some $\tau \in]f(x), f(x+1)[$ or $]f(x+1), f(x)[$. Since $\tau \leq f(x) + |Df(x)|$, the claims follows. Let now f on $\mathbb{N}$ be such that $\sup_{x\in\mathbb{N}} |Df(x)| \leq \lambda$. It follows from (5.14) and (5.15) that

$$\mathrm{Ent}_\mu(\mathrm{e}^f) \leq C\mathrm{e}^{2\lambda}\mathrm{E}_\mu\big(|Df|^2\mathrm{e}^f\big). \tag{5.16}$$

In particular,

$$\mathrm{Ent}_\mu(\mathrm{e}^f) \leq C\lambda^2\mathrm{e}^{2\lambda}\mathrm{E}_\mu(\mathrm{e}^f). \tag{5.17}$$

As a consequence of Corollary 2.12, we obtain a first result on Poisson tails of Lipschitz functions.

Proposition 5.4. *Let μ be a probability measure on $\mathbb{N}$ such that, for some constant $C > 0$,*

$$\mathrm{Ent}_\mu(f) \leq C\, \mathrm{E}_\mu\Big(\frac{1}{f}\,|Df|^2\Big)$$

for all functions f on $\mathbb{N}$ with positive values, where $Df(x) = f(x+1) - f(x)$, $x \in \mathbb{N}$. Then, for any F such that $\sup_{x\in\mathbb{N}} |DF(x)| \leq 1$, we have $\mathrm{E}_\mu(|F|) < \infty$ and, for all $r \geq 0$,

$$\mu\big(F \geq \mathrm{E}_\mu(F) + r\big) \leq \exp\Big(-\frac{r}{8}\,\log\Big(1 + \frac{r}{C}\Big)\Big).$$

In particular, $\mathrm{E}_\mu(\mathrm{e}^{\alpha|F|\log_+|F|}) < \infty$ for sufficiently small $\alpha > 0$.

The inequality of Proposition 5.4 describes the classical Gaussian tail behavior for the small values of r and the Poisson behavior for the large values of r (with respect to C). The constants have no reason to be sharp.

Of course, inequality (5.16) is part of the family of modified logarithmic Sobolev inequalities investigated in Section 4.2, with a function $B(\lambda)$ of the order of $\mathrm{e}^{2\lambda}$, $\lambda \geq 0$. According to Proposition 4.4, it may be tensorized in terms of two distinct norms on the gradients. The following statement is then an easy consequence of this observation.

Theorem 5.5. *Let μ be some measure on $\mathbb{N}$. Assume that for every f on $\mathbb{N}$ with $\sup_{x \in \mathbb{N}} |Df(x)| \leq \lambda$,*

$$\mathrm{Ent}_\mu(\mathrm{e}^f) \leq B(\lambda)\, \mathrm{E}_\mu\big(|Df|^2 \mathrm{e}^f\big) \tag{5.18}$$

where, as function of $\lambda \geq 0$,

$$B(\lambda) \leq c\, \mathrm{e}^{d\lambda}$$

for some $c, d > 0$. Denote by μ^n the product measure on $\mathbb{N}^n$. Let F be a function on $\mathbb{N}^n$ such that, for every $x \in \mathbb{N}^n$,

$$\sum_{i=1}^n \big|F(x+e_i) - F(x)\big|^2 \leq \alpha^2 \quad \text{and} \quad \max_{1 \leq i \leq n} \big|F(x+e_i) - F(x)\big| \leq \beta.$$

Then $\mathrm{E}_{\mu^n}(|F|) < \infty$ and, for every $r \geq 0$,

$$\mu^n\big(F \geq \mathrm{E}_{\mu^n}(F) + r\big) \leq \exp\bigg(-\frac{r}{2d\beta}\, \log\Big(1 + \frac{\beta d r}{4c\alpha^2}\Big)\bigg).$$

Proof. We tensorize (5.18) according to Proposition 4.4 to get that for every f on $\mathbb{N}^n$ such that $\max_{1 \leq i \leq n} |f(x+e_i) - f(x)| \leq \lambda$ for every $x \in \mathbb{N}^n$,

$$\mathrm{Ent}_{\mu^n}(\mathrm{e}^f) \leq B(\lambda)\, \mathrm{E}_{\mu^n}\bigg(\sum_{i=1}^n |D_i f|^2 \mathrm{e}^f\bigg) \tag{5.19}$$

where $D_i f(x) = f(x+e_i) - f(x)$, $i = 1, \ldots, n$. We then proceed exactly as in Corollary 2.12. Fix F on $\mathbb{N}^n$ satisfying the hypotheses of the statement. We may assume, by homogeneity, that $\beta = 1$. Furthermore, arguing as in Section 2.3, we may assume throughout the argument that F is bounded. Apply (5.19) to λF for every $\lambda \in \mathbb{R}$. Setting $H(\lambda) = \mathrm{E}_{\mu^n}(\mathrm{e}^{\lambda F})$, we get

$$\lambda H'(\lambda) - H(\lambda) \log H(\lambda) \leq \alpha^2 \lambda^2 B(\lambda) H(\lambda).$$

Therefore, with, as usual, $K(\lambda) = \frac{1}{\lambda} \log H(\lambda)$,

$$K'(\lambda) \leq \alpha^2 B(\lambda) \leq \alpha^2 c\, \mathrm{e}^{d\lambda}.$$

It follows that, for every $\lambda \geq 0$,

$$K(\lambda) \leq K(0) + \alpha^2\, \frac{c}{d}\, (\mathrm{e}^{d\lambda} - 1).$$

In other words,

$$\mathrm{E}_{\mu^n}(\mathrm{e}^{\lambda F}) \leq \mathrm{e}^{\lambda \mathrm{E}_{\mu^n}(F) + c\alpha^2 \lambda (\mathrm{e}^{d\tau} - 1)/d} \tag{5.20}$$

which holds for every $\lambda \in \mathbb{R}$ (changing F into $-F$.) We conclude with Chebyshev's exponential inequality. For every λ,

$$\mu^n\big(F \geq \mathrm{E}_{\mu^n}(F) + r\big) \leq \mathrm{e}^{-\lambda r + c\alpha^2 \lambda (\mathrm{e}^{d\lambda} - 1)/d}.$$

If $dr \leq 4c\alpha^2$ (for example), choose $\lambda = r/4c\alpha^2$ whereas when $dr \geq 4c\alpha^2$, take

$$\lambda = \frac{1}{d} \log\Big(\frac{dr}{2c\alpha^2}\Big).$$

The proof is easily completed. □

5.3 Sharp bounds

To conclude this work, we study the sharp form of the modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. As in Section 5.1, we start with the Bernoulli measure. The following statement will be our basic result.

Theorem 5.6. *For any function f on $\{0,1\}^n$,*

$$\mathrm{Ent}_{\mu_p^n}(\mathrm{e}^f) \leq pq\, \mathrm{E}_{\mu_p^n}\Big(\big(|Df|\mathrm{e}^{|Df|} - \mathrm{e}^{|Df|} + 1\big)\, \mathrm{e}^f\Big).$$

Proof. It is similar to the proof of Theorem 5.1 and relies on the next lemma.

Lemma 5.7. *The optimal constant $\delta > 0$ in the inequality*

$$\mathrm{Ent}_{\mu_p}(\mathrm{e}^f) \leq \delta pq\, \mathrm{E}_{\mu_p}(\mathrm{e}^f)$$

provided p is arbitrary in $[0,1]$ and $f : \{0,1\} \to \mathbb{R}$ is fixed is given by

$$\delta = a\,\mathrm{e}^a - \mathrm{e}^a + 1$$

where $a = |f(1) - f(0)|$.

Proof One may assume that $f(0) = 0$ and $f(1) = a$. The inequality we want to optimize becomes

$$p(1+x)\log(1+x) - (1+px)\log(1+px) \leq \delta pq(1+px) \tag{5.21}$$

where $x = \mathrm{e}^a - 1 \geq 0$. Consider the function $U = U(p)$ which is the difference between the left-hand-side and the right-hand side of (5.21). Then $U(0) = U(1) = 0$ and $U''' \geq 0$. As in the proof of Lemma 5.2, to find the best constant δ amounts to show the inequality $U'(0) \leq 0$. But

$$U'(0) = (1+x)\log(1+x) - x - \delta = a\,\mathrm{e}^a - \mathrm{e}^a + 1 - \delta$$

which is the result. $\square$

According to Lemma 5.7, the theorem is proved in dimension one. We now simply observe that the inequality may be tensorized. By the product property of entropy, we get namely, for every f on $\{0,1\}^n$,

$$\begin{aligned}&\mathrm{Ent}_{\mu_p^n}(\mathrm{e}^f)\\ &\leq pq\int\sum_{i=1}^n\Big(|f(x+e_i)-f(x)|\mathrm{e}^{|f(x+e_i)-f(x)|}-\mathrm{e}^{|f(x+e_i)-f(x)|}+1\Big)\mathrm{e}^{f(x)}d\mu_p^n(x)\end{aligned} \tag{5.22}$$

where we recall that $(e_1,\dots,e_n)$ is the canonical basis of $\mathbb{R}^n$ and that $x+e_i$ is understood here modulo 2. The function $Q(v)=\sqrt{v}\mathrm{e}^{\sqrt{v}}-\mathrm{e}^{\sqrt{v}}+1$, $v\geq 0$, is increasing and convex on $[0,\infty)$ with $Q(0)=0$. Hence, setting $a_i=|f(x+e_i)-f(x)|$, $i=1,\dots,n$,

$$\sum_{i=1}^n Q(a_i^2)\leq Q\bigg(\sum_{i=1}^n a_i^2\bigg)=Q\big(|Df(x)|^2\big)=|Df(x)|\mathrm{e}^{|Df(x)|}-\mathrm{e}^{|Df(x)|}+1.$$

Theorem 5.6 is therefore established. $\square$

As for Corollary 5.3, the Poisson limit theorem on (5.22) yields the following consequence for π_θ.

Corollary 5.8. *For any function f on $\mathbb{N}$,*

$$\mathrm{Ent}_{\pi_\theta}(\mathrm{e}^f)\leq\theta\,\mathrm{E}_{\pi_\theta}\Big(\big(|Df|\mathrm{e}^{|Df|}-\mathrm{e}^{|Df|}+1\big)\,\mathrm{e}^f\Big).$$

Corollary 5.8 is sharp in many respect. It becomes an equality for linear functions of the type $f(x)=cx+d$, $c\geq 0$. Furthermore, applying Theorem 5.6 and Corollary 5.8 to εf with $\varepsilon\to 0$ yields the Poincaré inequalities (5.12) and (5.13) for μ_p^n and π_θ respectively. This is easily verified using the fact that $a\mathrm{e}^a-\mathrm{e}^a+1$ behaves like $\frac{1}{2}a^2$ for small a.

As announced, the preceding statements actually describe sharp forms of modified logarithmic Sobolev inequalities in this context. As a consequence of Theorem 5.6 and Corollary 5.8, we namely get

Corollary 5.9. *For any function F on $\{0,1\}^n$ with $\max_{1\leq i\leq n}|f(x+e_i)-f(x)|\leq\lambda$ for every x in $\{0,1\}^n$,*

$$\mathrm{Ent}_{\mu_p^n}(\mathrm{e}^f)\leq pq\,\frac{\lambda\,\mathrm{e}^\lambda-\mathrm{e}^\lambda+1}{\lambda^2}\,\mathrm{E}_{\mu_p^n}\big(|Df|^2\mathrm{e}^f\big).$$

The case $n=1$ is just Lemma 5.7 together with the fact that $\lambda^{-2}[\lambda\,\mathrm{e}^\lambda-\mathrm{e}^\lambda+1]$ is non-decreasing in $\lambda\geq 0$. The corollary follows by tensorization. Similarly,

Corollary 5.10. *For any function f on $\mathbb{N}$ with $\sup_{x\in\mathbb{N}}|Df(x)|\leq\lambda$,*

$$\mathrm{Ent}_{\pi_\theta}(\mathrm{e}^f)\leq\theta\,\frac{\lambda\,\mathrm{e}^\lambda-\mathrm{e}^\lambda+1}{\lambda^2}\,\mathrm{E}_{\pi_\theta}\big(|Df|^2\mathrm{e}^f\big).$$

Again, via the central limit theorem, both Corollary 5.9 and Corollary 5.10 contain the Gaussian logarithmic Sobolev inequality. Let indeed φ be smooth enough on $\mathbb{R}$ and apply Corollary 5.9 to

$$f(x_1, \dots, x_n) = \varphi\Big(\frac{x_1 + \cdots + x_n - np}{\sqrt{npq}}\Big).$$

Then,

$$\max_{1 \le i \le n} |f(x + e_i) - f(x)| \le \frac{1}{\sqrt{npq}} \, \|\varphi\|_{\mathrm{Lip}} \to 0$$

as $n \to \infty$ and the result follows since $a\mathrm{e}^a - \mathrm{e}^a + 1 \sim \frac{1}{2}a^2$ as $a \to 0$. The same argument may be developed on the product form of Corollary 5.10 together with the central limit theorem for sums of independent Poisson random variables.

Due to the sharp constant in Corollary 5.9, the tail estimate of Theorem 5.5 may be improved. We namely get instead of (5.20) in the proof of Theorem 5.5

$$\mathrm{E}_{\mu_p^n}(\mathrm{e}^{\lambda F}) \le \mathrm{e}^{\lambda\, \mathrm{E}_{\mu_p^n}(F) + \lambda\alpha^2(\mathrm{e}^{\lambda} - 1 - \lambda)}.$$

The same holds for π_θ^n and this bound is sharp since, when $n = 1$ for example, it becomes an equality for $F(x) = x$, $x \in \mathbb{N}$. Together with Chebyshev's inequality and a straightforward minimization procedure, we get, for F on $\{0,1\}^n$ say, such that, for every $x = (x_1, \dots, x_n) \in \{0,1\}^n$,

$$\sum_{i=1}^n \big|F\big(x + e_i\big) - F(x)\big|^2 \le \alpha^2 \quad \text{and} \quad \max_{1 \le i \le n} \big|F\big(x + e_i\big) - F(x)\big| \le \beta$$

where $(e_1, \dots, e_n)$ is the canonical basis of $\mathbb{R}^n$, then, for every $r \ge 0$,

$$\mu_p^n\big(F \ge \mathrm{E}_{\mu_p^n}(F) + r\big) \le \exp\Big(-\Big(\frac{r}{\beta} + \frac{pq\alpha^2}{\beta^2}\Big) \log\Big(1 + \frac{\beta r}{pq\alpha^2}\Big) + \frac{r}{\beta}\Big). \tag{5.23}$$

A similar inequality thus holds for π_θ^n changing pq into θ. Such an inequality may be considered as an extension of the classical exponential inequalities for sums of independent random variables with parameters the size and the variance of the variables, and describing a Gaussian tail for the small values of r and a Poisson tail for its large values (cf. (3.17). It belongs to the family of concentration inequalities for product measures deeply investigated by M. Talagrand [Ta6]. With respect to [Ta6], the study presented here develops some new aspects related to concentration for Bernoulli measures and penalties [Ta6, §2].

6. SOME APPLICATIONS TO LARGE DEVIATIONS AND TO BROWNIAN MOTION ON A MANIFOLD

In this part, we present some applications of the ideas developed around concentration and logarithmic Sobolev inequalities to large deviations and to Brownian motion on a manifold. We show indeed how logarithmic Sobolev inequalities and exponential integrability can be reformulated as a large deviation upper bound. We then discuss some recent logarithmic Sobolev inequality for Wiener measure on the paths of a Riemannian manifold. We apply it to give a large deviation bound for the uniform distance of Brownian motion from its starting point on a manifold with non-negative Ricci curvature.

6.1 Logarithmic Sobolev inequalities and large deviation upper bounds

On some measurable space $(X, \mathcal{B})$, let $(\mu_n)_{n \in \mathbb{N}}$ be a family of probability measures. Consider some generalized square gradient Γ on a class $\mathcal{A}$ of functions on X such that, for every $f \in \mathcal{A}$ and $\lambda \in \mathbb{R}$, $\Gamma(\lambda f) = \lambda^2 \Gamma(f) \geq 0$. Γ could be either the square of a generalized modulus of gradient (2.3), or of some discrete one (2.9). Assume now that, for each $n \in \mathbb{N}$, there exists $c_n > 0$ such that, for every f in $\mathcal{A}$,

$$\mathrm{Ent}_{\mu_n}(\mathrm{e}^f) \leq \frac{c_n}{2}\, \mathrm{E}_{\mu_n}\big(\Gamma(f)\, \mathrm{e}^f\big). \tag{6.1}$$

Given two measurable sets A and B in X, set

$$d(A, B) = \inf_{x \in A, y \in B} \sup_{\|\Gamma(f)\|_\infty \leq 1} \big|f(x) - f(y)\big|. \tag{6.2}$$

Define $\mathcal{V}$ the class of all those $V \in \mathcal{B}$ such that $\lim_{n \to \infty} \mu_n(V) = 1$, and for every $A \in \mathcal{B}$, set

$$r(A) = \sup\big\{r \geq 0; \text{there exists}\, V \in \mathcal{V}\, \text{such that}\, d(A, V) \geq r\big\}.$$

Theorem 6.1. *Under (6.1), for every $A \in \mathcal{B}$,*

$$\limsup_{n \to \infty} \frac{c_n}{2} \log \mu_n(A) \leq -r(A)^2.$$

Proof. It is straighforward. Let $0 < r < r(A)$. Then, for some V in $\mathcal{V}$, and every n,

$$\mu_n(A) \le \mu_n(d_V \ge r).$$

Denote by d_V the distance to the set V, and let $F_V = \min(d_V, r)$. As a consequence of Corollary 2.5,

$$\mu_n\big(F_V \ge \mathrm{E}_{\mu_n}(F_V) + r\big) \le \mathrm{e}^{-r^2/2c_n}.$$

Repeating the argument leading to (1.28),

$$\mu_n(A) \le \mathrm{e}^{-\mu_n(V)^2 r^2/2c_n}$$

for every n. Since $\mu_n(V) \to 1$ as $n \to \infty$, the conclusion follows. Theorem 6.1 is established. □

Theorem 6.1 was used in [BA-L] to describe large deviations without topology for Gaussian measures. The operator Γ in this case relates to the Gross-Malliavin derivative and distance has to be understood with respect to the reproducing kernel Hilbert space (cf. [Le3]). In this case, the set functional $r(\cdot)$ is easily seen to connect with the classical rate functional in abstract Wiener spaces and, provided a topology has been fixed, coincide with this functional on the closure of A. Let more precisely μ be a Gaussian measure on the Borel sets $\mathcal{B}$ of a Banach space X with reproducing kernel Hilbert space $\mathcal{H}$. Denote by $\mathcal{K}$ the unit ball of $\mathcal{H}$. For every $\varepsilon > 0$, set $\mu_\varepsilon(\cdot) = \mu(\varepsilon^{-1}\cdot)$ and define the class $\mathcal{V}$ as those elements $V \in \mathcal{B}$ such that $\lim_{\varepsilon\to 0}\mu_\varepsilon(V) = 1$. In this case,

$$r(A) = \sup\big\{r \ge 0; \text{there exists}\, V \in \mathcal{V}\, \text{such that}\, (V + r\mathcal{K}) \cap A = \emptyset\big\}.$$

Then, for any Borel set A,

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log \mu_\varepsilon(A) \le -\frac{1}{2}\, r(A)^2.$$

Similar lower bounds can be described, however as simple consequences of the Cameron-Martin formula.

6.2 Some tail estimate for Brownian motion on a manifold

We have seen in (1.10) that if $(X_t)_{t\in T}$ is a (centered) Gaussian process such that $\sup_{t\in T}|X_t| < \infty$, then

$$\lim_{r\to\infty} \frac{1}{r^2} \log \mathbb{P}\big\{\sup_{t\in T}|X_t| \ge r\big\} = -\frac{1}{2\sigma^2}$$

where $\sigma = \sup_{t\in T}(\mathbb{E}(X_t^2))^{1/2}$. In particular, if $(B_t)_{t\ge 0}$ is Brownian motion in $\mathbb{R}^n$ starting from the origin, for every $T > 0$,

$$\lim_{r\to\infty} \frac{1}{r^2} \log \mathbb{P}\big\{\sup_{0\le t\le T}|B_t| \ge r\big\} = -\frac{1}{2T}\,. \tag{6.3}$$

As illustrated by these notes, this result (6.3) may be seen as a consequence of the logarithmic Sobolev inequality for Gaussian measures (cf. Corollary 2.6 and Section 1.3). Our aim here will be show that the same method may be followed for Brownian motion on a manifold. Theorem 6.2 below is known and follows from the Lyons-Takeda forward and backward martingale method [Ly], [Tak]. We only aim to show here unity of the method, deriving this large deviation estimate from logarithmic Sobolev inequalities for heat kernel measures and Wiener measures on path spaces developed recently by several authors. We actually do not discuss here analysis on path spaces and only use what will be necessary to this bound. We refer for example to [Hs2] for details and references. Once the proper logarithmic Sobolev inequality is released, the proof of the upper bound is straightforward, and inequality (6.6) might be of independent interest. The lower bound requires classical volume estimates in Riemannian geometry.

Let thus M be a complete non-compact Riemannian manifold with dimension n and distance d. Denote by $(B_t)_{t\geq 0}$ Brownian motion on M starting from $x_0 \in M$.

Theorem 6.2. *If M has non-negative Ricci curvature, for every $T > 0$,*

$$\lim_{r\to\infty} \frac{1}{r^2} \log \mathbb{P}\big\{ \sup_{0\leq t\leq T} d(B_t, x_0) \geq r \big\} = -\frac{1}{2T} \,.$$

As a consequence,

$$\mathbb{E}\Big(\exp\big(\alpha(\sup_{0\leq t\leq T} d(B_t, x_0))^2\big)\Big) < \infty$$

for every $\alpha < \frac{1}{2T}$.

Proof. . We first establish the upper bound in the preceding limit. Let $p_t(x,y)$ be the heat kernel on M, fundamental solution of the heat equation $\frac{\partial}{\partial t} = \frac{1}{2}\Delta$ where Δ is the Laplace-Beltrami operator on M. For fixed $t \geq 0$ and $x \in M$, let $\nu_t = \nu_t(x)$ be the heat kernel measure $p_t(x,y)dy$. The following is the logarithmic Sobolev inequality for the heat kernel measure on a Riemannian manifold with Ricci curvature bounded below [Ba2].

Lemma 6.3. *Assume that* $\mathrm{Ric} \geq -K$, $K \in \mathbb{R}$. *For every $t \geq 0$ and $x \in M$, and every smooth function f on M,*

$$\mathrm{Ent}_{\nu_t}(f^2) \leq 2C(t)\, \mathrm{E}_{\nu_t}\big(|\nabla f|^2\big)$$

where

$$C(t) = C_K(t) = \frac{\mathrm{e}^{Kt} - 1}{K} \quad (= t \text{ si } K = 0).$$

We now perform a Markov tensorization on entropy to describe, according to [Hs1], the logarithmic Sobolev inequality for cylindrical functions on the path space over M. Denote by $W_{x_0}(M)$ the space of continuous functions $x : \mathbb{R}_+ \mapsto M$ with $x(0) = x_0$, and by ν the Wiener measure on $W_{x_0}(M)$. A function f is called cylindrical on $W_{x_0}(M)$, if, for some φ on M^n and fixed times $0 \leq t_1 < \cdots < t_n$,

$f(x) = \varphi(x_{t_1}, \dots, x_{t_n})$. (It will be called smooth if φ is smooth.) If f is a smooth cylindrical function, we denote with some abuse by $\nabla_i f$ the gradient of φ with respect to the i-th coordinate, $i = 1, \dots, n$. According to [Hs2], let $U = (U_t)_{t \geq 0}$ the horizontal lift of Brownian motion to the tangent bundle $\mathcal{O}(M)$ and let $(\phi_{s,t})_{t \geq s}$ be the Ricci flow (matrix-valued process)

$$\frac{d}{dt}\phi_{s,t} = -\frac{1}{2}\,\mathrm{Ric}_{U_t}\phi_{s,t}, \quad \phi_{s,s} = \mathrm{I}\,.$$

The following is Lemma 4.1 in [Hs1] to which we refer for the proof.

Lemma 6.4. *If* $\mathrm{Ric} \geq -K$, $K \in \mathbb{R}$, *for any smooth* f *on* $W_{x_0}(M)$,

$$\mathrm{Ent}_\nu(f^2) \leq 2\sum_{i=1}^n C(t_i - t_{i-1})\mathrm{E}_\nu\left(\left|\sum_{j=i}^n \phi^*_{t_i,t_j} U_{t_j}^{-1}\nabla^j f\right|^2\right)$$

where ϕ^* *is the transpose of* ϕ.

We can now establish the upper bound in the limit of Theorem 6.2. Let $F(x) = \max_{1 \leq i \leq n} d(x_{t_i}, x_0)$, $0 \leq t_1 < \cdots < t_n \leq T$. Then

$$\left|\sum_{j=i}^n \phi^*_{t_i,t_j} U_{t_j}^{-1}\nabla^j F\right|^2 \leq C(T). \tag{6.4}$$

Indeed, for some appropriate partition $(A_j)_{1 \leq j \leq n}$ of $W_{x_0}(M)$, $|\nabla^j F| \leq I_{A_j}$ for every j. On the other hand, since $\mathrm{Ric} \geq -K$, $|\phi^*_{t_i,t_j}| \leq \mathrm{e}^{K(t_j - t_i)/2}$, $t_i < t_j$. Therefore, the left-hand side in (6.4) is bounded above by

$$\begin{aligned}
\sum_{i=1}^n c(t_i - t_{i-1})\sum_{j=i}^n \mathrm{e}^{K(t_j - t_i)} I_{A_j} &= \frac{1}{K}\sum_{j=1}^n I_{A_j}\sum_{i=1}^j (\mathrm{e}^{K(t_i - t_{i-1})} - 1)\mathrm{e}^{K(t_j - t_i)} \\
&= \frac{1}{K}\sum_{j=1}^n I_{A_j}\mathrm{e}^{Kt_j}\sum_{i=1}^j (\mathrm{e}^{-Kt_{i-1}} - \mathrm{e}^{-Kt_i}) \\
&= \sum_{j=1}^n I_{A_j} C(t_j) \\
&\leq \max_{1 \leq j \leq n} C(t_j) \leq C(T)
\end{aligned}$$

which is the result.

Now, apply the logarithmic Sobolev inequality of Lemma 6.4 to λF for every $\lambda \in \mathbb{R}$. We get

$$\mathrm{Ent}_\nu(\mathrm{e}^{\lambda F}) \leq 2C(T)\lambda^2\mathrm{E}_\nu(\mathrm{e}^{\lambda F}).$$

We then conclude, as in Section 2.3, that for every $r \geq 0$,

$$\nu\big(F \geq \mathrm{E}_\nu(F) + r\big) \leq \mathrm{e}^{-r^2/2C(T)}.$$

In other words,

$$\mathbb{P}\big\{\max_{1\le i\le n} d(B_{t_i},x_0) \ge \mathbb{E}\big(\max_{1\le i\le n} d(B_{t_i},x_0)\big)+r\big\} \le \mathrm{e}^{-r^2/2c(T)}. \tag{6.5}$$

Since $\sup_{0\le t\le T} d(B_t,x_0) < \infty$ almost surely, it follows from (6.5), exactly as for (1.24), that $\mathbb{E}(\sup_{0\le t\le T} d(B_t,x_0)) < \infty$ and, for every $r \ge 0$,

$$\mathbb{P}\big\{\sup_{0\le t\le T} d(B_t,x_0) \ge \mathbb{E}\big(\sup_{0\le t\le T} d(B_t,x_0)\big)+r\big\} \le \mathrm{e}^{-r^2/2C(T)}. \tag{6.6}$$

When $K=0$, it immediately yields that

$$\limsup_{r\to\infty} \frac{1}{r^2}\log \mathbb{P}\big\{\sup_{0\le t\le T} d(B_t,x_0)\ge r\big\} \le -\frac{1}{2T}\,.$$

We are left with the lower bound that will follow from known heat kernel minorations. We assume from now on that $\mathrm{Ric}\ge 0$. For every $r\ge 0$,

$$\mathbb{P}\big\{\sup_{0\le t\le T} d(B_t,x_0)\ge r\big\} \ge \mathbb{P}\big\{d(B_T,x_0)\ge r\big\} = \int_{\{x;d(x,x_0)\ge r\}} p_T(x,x_0)dx \tag{6.7}$$

Since $\mathrm{Ric}\ge 0$,

$$p_t(x,y) \ge \frac{1}{(2\pi t)^{n/2}}\,\mathrm{e}^{-d(x,y)^2/2t}$$

for every $x,y\in M$ and $t>0$ [Da, p. 173]. Therefore, for every $\varepsilon>0$,

$$\begin{aligned}\int_{\{x;d(x,x_0)\ge r\}} p_T(x,x_0)dx &\ge \int_{\{x;r+\varepsilon\ge d(x,x_0)\ge r\}} \frac{1}{(2\pi T)^{n/2}}\,\mathrm{e}^{-d(x,x_0)^2/2T}\\ &\ge \frac{1}{(2\pi T)^{n/2}}\,\mathrm{e}^{-(1+\varepsilon)^2r^2/2T}\big[V\big(x_0,(1+\varepsilon)r\big)\big)-V(x_0,r)\big]\end{aligned}$$

where $V(x,s)$, $s\ge 0$ is the Riemannian volume of the (open) geodesic ball $B(x,s)$ with center x and radius s in M. By the Riemannian volume comparison theorem (cf. e.g. [Cha2]), for every x in M and $0<s\le t$,

$$\frac{V(x,t)}{V(x,s)} \le \Big(\frac{t}{s}\Big)^n. \tag{6.8}$$

Let now z on the boundary of $B(x_0,(1+\frac{\varepsilon}{2})r)$. Since

$$B\big(z,\tfrac{\varepsilon}{2}\,r\big)\subset B\big(x_0,(1+\varepsilon)r\big)\setminus B(x_0,r) \quad\text{and}\quad B(x_0,r)\subset B\big(z,(2+\tfrac{\varepsilon}{2})r\big),$$

we get by (6.8),

$$\begin{aligned}V(x_0,r) &\le V\big(z,(2+\tfrac{\varepsilon}{2})r\big)\\ &\le \Big(\frac{4+\varepsilon}{\varepsilon}\Big)^n B\big(z,\tfrac{\varepsilon}{2}\,r\big)\\ &\le \Big(\frac{4+\varepsilon}{\varepsilon}\Big)^n \big[V\big(x_0,(1+\varepsilon)r\big)-V(x_0,r)\big].\end{aligned}$$

Therefore,

$$\big[V\big(x_0,(1+\varepsilon)r\big)-V(x_0,r)\big]\geq\Big(\frac{\varepsilon}{4+\varepsilon}\Big)^nV(x_0,r).$$

Summarizing, for every $r\geq 0$,

$$\int_{\{x;d(x,x_0)\geq r\}}p_T(x,x_0)dx\geq\frac{1}{(2\pi T)^{n/2}}\,\mathrm{e}^{-(1+\varepsilon)^2r^2/2T}\Big(\frac{\varepsilon}{4+\varepsilon}\Big)^nV(x_0,r).$$

It is now a simple matter to conclude from this lower bound and (6.7) that

$$\liminf_{r\to\infty}\frac{1}{r^2}\,\log\mathbb{P}\big\{d(B_T,x_0)\geq r\big\}\geq-\frac{1}{2T}\,.$$

Theorem 6.2 is therefore established. $\square$

7. ON REVERSED HERBST'S INEQUALITIES AND BOUNDS ON THE LOGARITHMIC SOBOLEV CONSTANT

In this chapter, we investigate one instance in which a concentration property, or rather exponential integrability, implies a logarithmic Sobolev inequality. We present the result in the context of the Boltzmann measures already considered in Section 1.2. The argument is based on a recent observation by F.-Y. Wang [Wan] (see also [Ai]). In a more geometric setting, Wang's result also leads to dimension free lower bounds on the logarithmic Sobolev constant in compact manifolds with non-negative Ricci curvature that we review in the second paragraph. In the next two sections, we present a new upper bound on the diameter of a compact Riemannian manifold by the logarithmic Sobolev constant, the dimension and the lower bound on the Ricci curvature. We deduce a sharp upper bound on the logarithmic Sobolev constant in spaces with non-negative Ricci curvature. The last section is due to L. Saloff-Coste. It is shown how the preceding ideas may be developed similarly for discrete models, leading to estimates between the diameter and the logarithmic Sobolev constant.

7.1. Reversed Herbst's inequality

As in Section 1.2, let us consider a C^2 function W on $\mathbb{R}^n$ such that e^{-W} is integrable with respect to Lebesgue measure and let

$$d\mu(x) = Z^{-1}\mathrm{e}^{-W(x)}dx$$

where Z is the normalization factor. μ is the invariant measure of the generator $\mathrm{L} = \frac{1}{2}\Delta - \frac{1}{2}\nabla W \cdot \nabla$. We denote by $W''(x)$ the Hessian of W at the point x.

As we have seen in Theorem 1.1 and (2.17), when, for some $c > 0$, $W''(x) \geq c\,\mathrm{Id}$ for every x, μ satisfies a Gaussian-type isoperimetric inequality as well as a logarithmic Sobolev inequality (with respect to $\mathrm{E}_\mu(|\nabla f|^2)$, and therefore a concentration property. In particular,

$$\int \mathrm{e}^{\alpha|x|^2} d\mu(x) < \infty$$

for every $\alpha < c/2$. The following theorem, due to F.-Y. Wang [Wan] (in a more general setting) is a sort of conserve to this result.

Theorem 7.1. *Assume that for some $c \in \mathbb{R}$, $W''(x) \geq c\,\mathrm{Id}$ for every x and that for some $\varepsilon > 0$,*

$$\iint \mathrm{e}^{(c^- + \varepsilon)|x-y|^2} d\mu(x) d\mu(y) < \infty$$

where $c^- = -\min(c, 0)$. Then μ satisfies the logarithmic Sobolev inequality

$$\mathrm{Ent}_\mu(f^2) \leq C\,\mathrm{E}_\mu\big(|\nabla f|^2\big)$$

for some $C > 0$.

According to (2.17), the theorem is only of interest when $c \leq 0$ (which we assume below). The integrability assumption of the theorem is in particular satisfied when

$$\int \mathrm{e}^{2(c^- + \varepsilon)|x|^2} d\mu(x) < \infty.$$

As a consequence of Section 4 of [Ba-L], we may also conclude under the assumptions of Theorem 7.1 to a Gaussian isoperimetric inequality

$$\mu_s(\partial A) \geq \sqrt{c'}\,\mathcal{U}\big(\mu(A)\big)$$

for some $c' > 0$, in the sense of Section 1. In the recent work [Bob5], the Poincaré inequality for μ is established when $W'' \geq 0$ without any further conditions.

Theorem 7.1 allows us to consider cases when the potential W is not convex. Another instance of this type is provided by the perturbation argument of [H-S]. Assume namely that a Boltzmann measure μ as before satisfies a logarithmic Sobolev inequality with constant C and let $d\nu = T^{-1}\mathrm{e}^{-V}dx$ be such that $\|W - V\|_\infty \leq K$. Then ν satisfies a logarithmic Sobolev inequality with constant $C\mathrm{e}^{4K}$. To prove it, note first that $\mathrm{e}^{-K}T \leq Z \leq \mathrm{e}^{K}T$. As put forward in [H-S], for every $a, b > 0$, $b \log b - b \log a - b + a \geq 0$ and

$$\mathrm{Ent}(f^2) = \inf_{a>0} \mathrm{E}(f^2 \log f^2 - f^2 \log a - f^2 + a).$$

Therefore,

$$\begin{aligned}
\mathrm{Ent}_\nu(f^2) &= \inf_{a>0} \mathrm{E}_\mu\big([f^2 \log f^2 - f^2 \log a - f^2 + a]\,\mathrm{e}^{W-V} Z T^{-1}\big) \\
&\leq \mathrm{e}^{2K} \mathrm{Ent}_\mu(f^2) \\
&\leq C\mathrm{e}^{2K} \mathrm{E}_\mu\big(|\nabla f|^2\big) \\
&\leq C\mathrm{e}^{2K} \mathrm{E}_\nu\big(|\nabla f|^2 \mathrm{e}^{V-W} T Z^{-1}\big) \\
&\leq C\mathrm{e}^{4K} \mathrm{E}_\nu\big(|\nabla f|^2\big).
\end{aligned}$$

(The same argument applies for the variance and Poincaré inequalities.) One odd feature of both Theorem 7.1 and this perturbation argument is that they yield rather poor constants as functions of the dimension (even for simple product measures) and seem therefore of little use in statistical mechanic applications.

Proof of Theorem 7.1. The main ingredient of the proof is the following result of [Wan] which describes a Harnack-type inequality for the Markov semigroup $(P_t)_{t\geq 0}$ with generator $\mathrm{L} = \frac{1}{2}\Delta - \frac{1}{2}\nabla W \cdot \nabla$.

Lemma 7.2. *Under the hypothesis of the theorem, for every bounded measurable function f on $\mathbb{R}^n$, every $x, y \in \mathbb{R}^n$ and every $t > 0$,*

$$P_t f(x)^2 \leq P_t(f^2)(y)\, \mathrm{e}^{c(\mathrm{e}^{ct}-1)^{-1}|x-y|^2}$$

(where we agree that $c(\mathrm{e}^{ct} - 1)^{-1} = t^{-1}$ when $c = 0$).

Proof. We may assume $f > 0$ and smooth. Fix $x, y \in \mathbb{R}^n$ and $t > 0$. Let, for every $0 \leq s \leq t$, $x_s = (s/t)x + (1 - (s/t))y$. Take also a C^1 function h on $[0, t]$ with non-negative values such that $h(0) = 0$ and $h(t) = t$. Set, for $0 \leq s \leq t$,

$$\varphi(s) = P_s\big((P_{t-s}f)^2\big)(x_{h(s)}).$$

Then,

$$\begin{aligned}\frac{d\varphi}{ds} &= P_s\big(|\nabla P_{t-s}f|^2\big)(x_{h(s)}) + t^{-1}h'(s)\big\langle x - y, \nabla P_s\big((P_{t-s}f)^2\big)(x_{h(s)})\big\rangle \\ &\geq P_s\big(|\nabla P_{t-s}f|^2\big)(x_{h(s)}) - t^{-1}\big|h'(s)\big||x - y|\big|\nabla P_s\big((P_{t-s}f)^2\big)(x_{h(s)})\big|.\end{aligned}$$

Now, under the assumption $W'' \geq c$, it is well-known that, for every smooth g and every $u \geq 0$,

$$|\nabla P_u g| \leq \mathrm{e}^{-cu/2} P_u\big(|\nabla g|\big). \tag{7.1}$$

For example, the condition $W'' \geq c$ may be interpreted as a curvature condition and (7.1) then follows e.g. from [Ba2], Proposition 2.3. Therefore,

$$\begin{aligned}\frac{d\varphi}{ds} &\geq P_s\big(|\nabla P_{t-s}f|^2\big)(x_{h(s)}) - t^{-1}\big|h'(s)\big||x - y|\,\mathrm{e}^{-cs/2}P_s\big(\big|\nabla(P_{t-s}f)^2\big|\big)(x_{h(s)}) \\ &\geq P_s\Big(|\nabla P_{t-s}f|^2 - 2t^{-1}\big|h'(s)\big||x - y|\,\mathrm{e}^{-cs/2}P_{t-s}f|\nabla P_{t-s}f|\Big)(x_{h(s)}).\end{aligned}$$

Using that $X^2 - aX \geq -\frac{a^2}{4}$, it follows that

$$\frac{d\varphi}{ds} \geq -t^{-2}|x - y|^2 \mathrm{e}^{-cs} h'(s)^2 \varphi(s).$$

Integrating this differential inequality yields

$$P_t f(x)^2 \leq P_t(f^2)(y) \exp\Big(t^{-2}|x - y|^2 \textstyle\int_0^t \mathrm{e}^{-cs}h'(s)^2 ds\Big).$$

We then simply optimize the choice of h by taking

$$h(s) = t(\mathrm{e}^{ct} - 1)^{-1}(\mathrm{e}^{cs} - 1), \quad 0 \leq s \leq t.$$

The proof of Lemma 7.1 is complete. □

In order to prove Theorem 7.1, we will first show that there is a spectral gap inequality for μ. To this goal, we follow the exposition in [B-L-Q]. Let f be a smooth function on $\mathbb{R}^n$, with $\mathrm{E}_\mu(f) = 0$. By spectral theory, it is easily seen that for every $t \geq 0$,

$$\mathrm{E}_\mu(f^2) \leq \mathrm{E}_\mu\big((P_t f)^2\big) + 2t\,\mathrm{E}_\mu\big(f(-\mathrm{L}f)\big) = \mathrm{E}_\mu\big((P_t f)^2\big) + t\,\mathrm{E}_\mu\big(|\nabla f|^2\big). \tag{7.2}$$

Since $\mathrm{E}_\mu(f) = 0$, for every x,

$$\big|P_t f(x)\big| \leq \int \big|P_t f(x) - P_t f(y)\big| d\mu(y). \tag{7.3}$$

Now,

$$\big|P_t f(x) - P_t f(y))\big| \leq |x-y|\big|\nabla P_t f(z)\big|$$

for some z on the line joining x to y. By (7.1) and by Lemma 7.2 applied to $|\nabla f|$ and to the couple (z,y),

$$\big|\nabla P_t f(z)\big| \leq e^{-ct/2} P_t\big(|\nabla f|\big)(z) \leq e^{-ct/2} P_t\big(|\nabla f|^2\big)(y)^{1/2} \mathrm{e}^{c(\mathrm{e}^{ct}-1)^{-1}|z-y|^2/2}.$$

Therefore by the Cauchy-Schwarz inequality (with respect to the variable y),

$$\bigg(\int \big|P_t f(x) - P_t f(y)\big| d\mu(y)\bigg)^2 \leq e^{-ct}\,\mathrm{E}_\mu\big(|\nabla f|^2\big) \int |x-y|^2 \mathrm{e}^{c(\mathrm{e}^{ct}-1)^{-1}|x-y|^2} d\mu(y).$$

Integrating in $d\mu(x)$, together with (7.2) and (7.3), we get

$$\mathrm{E}_\mu(f^2) \leq e^{-ct}\,\mathrm{E}_\mu\big(|\nabla f|^2\big) \iint |x-y|^2 \mathrm{e}^{c(\mathrm{e}^{ct}-1)^{-1}|x-y|^2} d\mu(x)d\mu(y) + t\,\mathrm{E}_\mu\big(|\nabla f|^2\big).$$

Letting t be sufficiently large, it easily follows from the hypothesis that

$$\mathrm{E}_\mu(f^2) \leq C\,\mathrm{E}_\mu\big(|\nabla f|^2\big) \tag{7.4}$$

for some finite constant C.

It would certainly be possible to prove the logarithmic Sobolev inequality in the same spirit. There is however a simpler route via hypercontractivity which, together with the spectral gap, immediately yields the conclusion. Let us consider again Wang's inequality of Lemma 7.2. Let $1 < \theta < 2$ and write, for every f (bounded to start with) and every $t > 0$,

$$\begin{aligned}\mathrm{E}_\mu\big(|P_t f|^{2\theta}\big) &= \mathrm{E}_\mu\Big(|P_t f|^\theta \big(|P_t f|^2\big)^{\theta/2}\Big)\\ &\leq \iint \big|P_t f(x)\big|^\theta \big(P_t(f^2)(y)\big)^{\theta/2} \mathrm{e}^{\theta c(\mathrm{e}^{ct}-1)^{-1}|x-y|^2/2} d\mu(x)d\mu(y).\end{aligned}$$

By Hölder's inequality, we get that

$$\|P_t f\|_{2\theta} \leq N^{\frac{2-\theta}{4\theta}} \|f\|_2 \tag{7.5}$$

where

$$N = \iint e^{\theta(2-\theta)^{-1}c(e^{ct}-1)^{-1}|x-y|^2} d\mu(x)d\mu(y).$$

Provided θ is sufficiently close to 1 and t large enough, N is finite by the hypothesis. Therefore, P_t satisfies a weak form of hypercontractivity which, as is well-known, is equivalent to a defective logarithmic Sobolev inequality of the type (2.26). We get namely from (7.5) (see [Gr1] or [DeS]),

$$\mathrm{Ent}_\mu(f^2) \le \frac{\theta t}{\theta - 1}\mathrm{E}_\mu\big(|\nabla f|^2\big) + \frac{2-\theta}{2(\theta-1)} \log N \, \mathrm{E}_\mu(f^2) \tag{7.6}$$

for every smooth f. We are left to show that such a defective logarithmic Sobolev inequality may be turned into a true logarithmic Sobolev inequality with the help of the spectral gap (7.4). Again, this is a classical fact that relies on the inequality (cf. [Ro2], [De-S]),

$$\mathrm{Ent}_\mu(f^2) \le \mathrm{Ent}_\mu\big((f - \mathrm{E}_\mu(f))^2\big) + 2\,\mathrm{Var}_\mu(f). \tag{7.7}$$

Inequality (7.6) applied to $f - \mathrm{E}_\mu(f)$ together with (7.7) and (7.4) complete the proof of the theorem. □

Note that N appears in (7.6) in the defective term as $\log N$ whereas in the Poincaré inequality (7.4), it appears as N (or some power of N). This is very sensible for product measures for which usually N is exponential in the dimension.

7.2 Dimension free lower bounds

In this section, we adopt a more geometric point of view and concentrate on lower bounds of the logarithmic Sobolev constant of a (compact) Riemannian manifold M with non-negative Ricci curvature in term of the diameter D of M.

Given some probability measure μ on $(X, \mathcal{B})$, and some energy functional $\mathcal{E}$ on a class $\mathcal{A}$ of functions, we introduced in Section 2.1 the definitions of spectral gap (or Poincaré) and logarithmic Sobolev inequalities. Let us now agree to denote by λ_1 the largest constant $\lambda > 0$ such that for every f in $\mathcal{A}$,

$$\lambda \,\mathrm{Var}_\mu(f) \le \mathcal{E}(f),$$

and by ρ_0 the largest $\rho > 0$ such that for every f in $\mathcal{A}$,

$$\rho \,\mathrm{Ent}_\mu(f^2) \le 2\,\mathcal{E}(f).$$

Although it is usually the case, we cannot always ensure, at this level of generality, that $\lambda_1 \mathrm{Var}(f) \le \mathcal{E}(f)$ and $\rho_0 \mathrm{Ent}(f^2) \le 2\mathcal{E}(f)$ for every $f \in \mathcal{A}$. The estimates we present below are proved using arbitrary $\lambda < \lambda_1$ and $\rho < \rho_0$. This will be mostly understood. By Proposition 2.1, one always has that $\rho_0 \le \lambda_1$. Emphasis has been put in the last years on identifying the logarithmic Sobolev constant and comparing it to the spectral gap.

Let M be a complete connected Riemannian manifold with dimension n and finite volume $V(M)$, and let $d\mu = \frac{dv}{V(M)}$ be the normalized Riemannian measure on M. Compact manifolds are prime examples. Let λ_1 and ρ_0 be respectively the spectral gap and the logarithmic Sobolev constant of μ with respect to Dirichlet form of the Laplace-Beltrami operator Δ (rather than $\frac{1}{2}\Delta$ here) on M, that is

$$\mathcal{E}(f) = \mathrm{E}_\mu\big(f(-\Delta f)\big) = \mathrm{E}_\mu\big(|\nabla f|^2\big)$$

for every smooth enough function f on M. If M is compact, it is known that $0 < \rho_0 \leq \lambda_1$ [Ro1]. Let D be the diameter of M if M is compact.

It is known since [Li], [Z-Y] that when M has non-negative Ricci curvature,

$$\lambda_1 \geq \frac{\pi^2}{D^2}\,. \tag{7.8}$$

Since $\lambda_1 \geq \rho_0$, it has been an open question for some time to prove that a similarly lower bound holds for the logarithmic Sobolev constant ρ_0. This has been proved recently by F.-Y. Wang [Wan] on the basis of his Lemma 7.2. Following [B-L-Q], we present here a simple proof of a somewhat stronger result.

Theorem 7.3. *Let M be a compact Riemannian manifold with diameter D and non-negative Ricci curvature, and denote by λ_1 and ρ_0 the spectral gap and the logarithmic Sobolev constant. Then*

$$\rho_0 \geq \frac{\lambda_1}{1 + 2D\sqrt{\lambda_1}}\,.$$

In particular,

$$\rho_0 \geq \frac{\pi^2}{(1+2\pi)D^2}\,.$$

Proof. We use Lemma 7.2 in this geometric context. Under the curvature assumption Ric ≥ 0, it yields similarly that if $(P_t)_{t\geq 0}$ is the heat semigroup on M (with generator Δ), for every f on M, every $x, y \in M$ and $t > 0$,

$$P_t f(x)^2 \leq P_t f^2(y)\, \mathrm{e}^{d(x,y)^2/2t}$$

where $d(x, y)$ is the geodesic distance from x to y. In particular,

$$\|P_t\|_{2\to\infty} \leq \mathrm{e}^{D^2/4t}.$$

By symmetry,

$$\|P_t\|_{1\to\infty} \leq \|P_{t/2}\|_{1\to 2}\|P_{t/2}\|_{2\to\infty} \leq \mathrm{e}^{D^2/t}. \tag{7.9}$$

To prove the theorem, we then simply follow the usual route based on the heat semigroup as developed in [Ba1], and already described in our proof of the Gaussian logarithmic Sobolev inequality (2.15). Fix $f > 0$ smooth and $t > 0$. We write

$$\begin{aligned}
\mathrm{E}_\mu(f\log f) - \mathrm{E}_\mu(P_t f\log P_t f) &= -\int_0^t \mathrm{E}_\mu\big(\Delta P_s f \log P_s f\big)ds \\
&= \int_0^t \mathrm{E}_\mu\Big(\frac{|\nabla P_s f|^2}{P_s f}\Big)ds.
\end{aligned}$$

Now since $\mathrm{Ric} \geq 0$, $|\nabla P_s f| \leq P_s(|\nabla f|)$ (cf. e.g. [Ba2]). Moreover, by the Cauchy-Schwarz inequality,

$$P_s\big(|\nabla f|\big)^2 \leq P_s\Big(\frac{|\nabla f|^2}{f}\Big) P_s f$$

so that

$$\mathrm{E}_\mu(f \log f) - \mathrm{E}_\mu(P_t f \log P_t f) \leq \int_0^t \mathrm{E}_\mu\bigg(P_s\Big(\frac{|\nabla f|^2}{f}\Big)\bigg) ds = t\, \mathrm{E}_\mu\Big(\frac{|\nabla f|^2}{f}\Big).$$

Now, by (7.9),

$$\mathrm{E}_\mu(P_t f \log P_t f) \leq \mathrm{E}_\mu(f) \log \mathrm{E}_\mu(f) + \frac{D^2}{t}\, \mathrm{E}_\mu(f)$$

(since μ is invariant for P_t). Therefore, for every $t > 0$,

$$\mathrm{Ent}_\mu(f) \leq \frac{D^2}{t}\, \mathrm{E}_\mu(f) + t\, \mathrm{E}_\mu\Big(\frac{|\nabla f|^2}{f}\Big).$$

Changing f into f^2,

$$\mathrm{Ent}_\mu(f^2) \leq \frac{D^2}{t}\, \mathrm{E}_\mu(f^2) + 4t\, \mathrm{E}_\mu\big(|\nabla f|^2\big). \tag{7.10}$$

As we know, this defective logarithmic Sobolev inequality may then be turned into a true logarithmic Sobolev inequality with the help of λ_1 using (7.7). That is, (7.10) applied to $f - \mathrm{E}_\mu(f)$ yields together with (7.7)

$$\begin{aligned}\mathrm{Ent}_\mu(f^2) &\leq \Big(\frac{D^2}{t} + 2\Big) \mathrm{Var}_\mu(f) + 4t\, \mathrm{E}_\mu\big(|\nabla f|^2\big) \\ &\leq \Big(\frac{D^2}{\lambda_1 t} + \frac{2}{\lambda_1} + 4t\Big) \mathrm{E}_\mu\big(|\nabla f|^2\big).\end{aligned}$$

Optimizing in $t > 0$, the first claim of Theorem 7.3 follows. The second claim is then a consequence of (7.8). The proof is complete. □

Similar results may be obtained in manifolds with Ricci curvature bounded below. Formulae are however somewhat more complicated (see [Wan], [B-L-Q]).

7.3 Upper bounds on the logarithmic Sobolev constant

We pursue our brief investigation on the spectral gap and logarithmic Sobolev constant by means of upper bounds. This question has mainly be raised in the framework of a Markov generator with associated Dirichlet form $\mathcal{E}$. It covers in particular Laplace-Beltrami and second-order elliptic operators on manifolds. Let us briefly review a few examples, some of them already alluded to in the previous chapters.

Spectral gaps and logarithmic Sobolev constants coincide for Gaussian measures by (2.15) and (2.16). A first example for which $\rho_0 < \lambda_1$ was brought in light in the paper [K-S] with the Laguerre generator with invariant measure the one-sided exponential distribution. As we have seen indeed in (4.4) and (4.5), $\rho_0 = \frac{1}{2} < 1 = \lambda_1$. On the two-point space $\{0,1\}$ with measure $\mu_p(\{1\}) = p$ and $\mu_p(\{0\}) = q = 1-p$ and energy

$$\mathcal{E}(f) = \mathrm{E}_{\mu_p}\big(|Df|^2\big) = \big|f(1) - f(0)\big|^2,$$

we have seen ((5.2), (5.3)) that $\lambda_1 = 1/pq$ whereas

$$\rho_0 = \frac{2}{pq}\,\frac{p-q}{\log p - \log q}\,.$$

In particular, $\rho_0 = \lambda_1$ only in the symmetric case $p = q = \frac{1}{2}$. Although rather recent, this example clearly indicates that, in general $\rho_0 < \lambda_1$. As discussed in Part 5, Poisson measures may be considered as an extreme case for which λ_1 is strictly positive while $\rho_0 = 0$. On the other hand, by (2.15) and (2.16), $\rho_0 = \lambda_1 = 1$ for the canonical Gaussian measure on $\mathbb{R}^n$.

We turn to another family of examples. Let M be a smooth complete connected Riemannian manifold with dimension n and finite volume $V(M)$, and let $d\mu = \frac{dv}{V(M)}$ be the normalized Riemannian measure on M. Compact manifolds are prime examples. Let λ_1 and ρ_0 be respectively the spectral gap and the logarithmic Sobolev constant of μ with respect to Dirichlet form of the Laplace-Beltrami Δ operator on M. We have seen that when M is compact, $0 < \rho_0 \le \lambda_1$. When $\mathrm{Ric} \ge R > 0$, it goes back to A. Lichnerowicz (cf. [Cha1]) that $\lambda_1 \ge R_n$ where $R_n = \frac{R}{1-\frac{1}{n}}$ with equality if and only if M is a sphere (Obata's theorem). This lower bound has been shown to hold similarly for the logarithmic Sobolev constant by D. Bakry and M. Emery [Ba-E] so that $\lambda_1 \ge \rho_0 \ge R_n$. The case of equality for ρ_0 is a consequence of Obata's theorem due to an improvement of the preceding by O. Rothaus [Ro2] who showed that when M is compact and $\mathrm{Ric} \ge R$ ($R \in \mathbb{R}$),

$$\rho_0 \ge \alpha_n \lambda_1 + (1-\alpha_n) R_n \tag{7.11}$$

where $\alpha_n = 4n/(n+1)^2$. As examples, $\rho_0 = \lambda_1 = n$ on the n-sphere [M-W]. On the n-dimensional torus, $\lambda_1 = \rho_0 = 1$. The question whether $\rho_0 < \lambda_1$ in this setting has been open for some time until the geometric investigation by L. Saloff-Coste [SC1]. He showed that actually the existence of a logarithmic Sobolev inequality in a Riemannian manifold with finite volume and Ricci curvature bounded below forces the manifold to be compact whereas it is known that there exists non-compact manifolds of finite volume with $\lambda_1 > 0$. In particular, there exist compact manifolds of constant negative sectional curvature with spectral gaps uniformly bounded away from zero, and arbitrarily large diameters (cf. [SC1]. This yield examples for which the ratio ρ_0/λ_1 can be made arbitrarily small.

Our first result here is a significant improvement of the quantitative bound of of [SC1].

Theorem 7.4. *Assume that* $\mathrm{Ric} \ge -K$, $K \ge 0$. *If* $\rho_0 > 0$, *then* M *is compact. Furthermore, if* D *is the diameter of* M, *there exists a numerical constant* $C > 0$ *such that*

$$D \le C\sqrt{n}\,\max\Big(\frac{1}{\sqrt{\rho_0}}\,,\,\frac{\sqrt{K}}{\rho_0}\Big).$$

It is known from the theory of hypercontractive semigroups (cf. [De-S]) that conversely there exists $C(n,K,\varepsilon)$ such that

$$\rho_0 \geq \frac{C(n,K,\varepsilon)}{D}$$

when $\lambda_1 \geq \varepsilon$.

The proof of [SC1] uses refined bounds on heat kernel and volume estimates. A somewhat shorter proof is provided in [Le2], still based on heat kernel. We present here a completely elementary argument based on the Riemannian volume comparison theorems and the concentration properties behind logarithmic Sobolev inequalities described in Part 2.

Proof. As a consequence of Corollary 2.6 and (2.25), for every measurable set A in M and every $r \geq 0$,

$$1 - \mu(A_r) \leq \mathrm{e}^{-\rho_0 \mu(A)^2 r^2/2} \tag{7.12}$$

where $A_r = \{x \in M, d(x,A) < r\}$. This is actually the only property that will be used throughout the proof.

We show first that M is compact. We proceed by contradiction and assume that M is not compact. Denote by $B(x,u)$ the geodesic ball in M with center x and radius $u \geq 0$. Choose $A = B(x_0, r_0)$ a geodesic ball such that $\mu(A) \geq \frac{1}{2}$. By non-compactness (and completeness), for every $r \geq 0$, we can take z at distance $r_0 + 2r$ from x_0. In particular, $A \subset B(z, 2(r_0 + r))$. By the Riemannian volume comparison theorem [Cha2], for every $x \in M$ and $0 < s < t$,

$$\frac{V(x,t)}{V(x,s)} \leq \Big(\frac{t}{s}\Big)^n \mathrm{e}^{\sqrt{(n-1)K}t} \tag{7.13}$$

where we recall that $V(x,u)$ is the volume of the ball $B(x,u)$ with center x and radius $u \geq 0$. Therefore,

$$\begin{aligned} V(z,r) &\geq \Big(\frac{r}{2(r_0+r}\Big)^n \mathrm{e}^{-2(r+r_0)\sqrt{(n-1)K}}\, V\big(z, 2(r_0+r)\big) \\ &\geq \frac{1}{2}\Big(\frac{r}{2(r_0+r)}\Big)^n \mathrm{e}^{-2(r_0+r)\sqrt{(n-1)K}}\, V(M). \end{aligned}$$

Since $B(z,r)$ is included in the complement of $A_r = B(x_0, r_0+r)$, we get from (7.12)

$$\frac{1}{2}\Big(\frac{r}{2(r_0+r)}\Big)^n \mathrm{e}^{-2(r_0+r)\sqrt{(n-1)K}} \leq \mathrm{e}^{-\rho_0 r^2/8} \tag{7.14}$$

which is impossible as $r \to \infty$.

Thus M is compact. Denote by D be its diameter. Let $x_0 \in M$ and let $B(x_0, \frac{D}{8})$. We distinguish between two cases. If $\mu(B(x_0, \frac{D}{8})) \geq \frac{1}{2}$, take $A = B(x_0, \frac{D}{8})$. By definition of D, we may choose $r = r_0 = \frac{D}{8}$ in (7.14) to get

$$\frac{1}{2} \cdot \frac{1}{4^n}\, \mathrm{e}^{-\sqrt{(n-1)K}D/2} \leq \mathrm{e}^{-\rho_0 D^2/512}.$$

If $\mu(B(x_0, \frac{D}{8})) < \frac{1}{2}$, apply (7.11) to A the complement of $B(x_0, \frac{D}{8})$. The ball $B(x_0, \frac{D}{16})$ is included in the complement of $A_{D/16}$. Moreover, by (7.13),

$$V\Big(x_0, \frac{D}{16}\Big) \geq \frac{1}{16^n} \, \mathrm{e}^{-\sqrt{(n-1)K}D} \, V(M).$$

Therefore, by (7.12) with $r = \frac{D}{16}$,

$$\frac{1}{16^n} \, \mathrm{e}^{-\sqrt{(n-1)K}D} \leq \mathrm{e}^{-\rho_0 D^2/2048}.$$

In both cases,

$$\rho_0 D^2 - C\sqrt{(n-1)K} \, D - Cn \leq 0$$

for some numerical constant $C > 0$. Hence

$$D \leq \frac{C\sqrt{(n-1)K} + \sqrt{C^2(n-1)K + 4C\rho_0 n}}{2\rho_0}$$

and thus

$$D \leq \frac{C\sqrt{(n-1)K} + \sqrt{C\rho_0 n}}{\rho}$$

which yields the conclusion. The theorem is established. □

Note that the proof shows, under the assumption of Theorem 7.4, that M is compact as soon as

$$\limsup_{r \to \infty} \frac{-1}{r} \, \log\big[1 - \mu\big(B(x,r)\big)\big] = \infty$$

for some (or all) $x \in M$. In particular $\lambda_1 > 0$ under this condition. This observation is a kind of converse to (2.36).

Corollary 7.5. *Let M be a compact Riemannian manifold with dimension n and non-negative Ricci curvature. Then*

$$\rho_0 \leq \frac{Cn}{D^2}$$

for some numerical constant $C > 0$.

Corollary 7.5 has to be compared to Cheng's upper bound on the spectral gap [Che] of compact manifolds with non-negative Ricci curvature

$$\lambda_1 \leq \frac{2n(n+4)}{D^2} \tag{7.15}$$

so that, generically, the difference between the upper bound on λ_1 and ρ_0 seems to be of the order of n. Moreover, it is mentioned in [Che] that there exists examples with $\lambda_1 \approx n^2/D^2$. Although we are not aware of such examples, they indicate perhaps that both Rothaus' lower bound (7.11) and Corollary 7.5 could be sharp.

Note also that (7.11) together with Corollary 7.5 allows us to recover Cheng's upper bound on λ_1 of the same order in n. Actually, the proof of Theorem 7.4 together with the concentration property under the spectral gap (Proposition 2.13) would also yield Cheng's inequality (7.15) up to a numerical constant.

Corollary 7.5 is stated for (compact) manifolds without boundary but it also holds for compact manifolds of non-negative Ricci curvature with convex boundary (and Neuman's conditions). In particular, this result applies to convex bounded domains in $\mathbb{R}^n$ equipped with normalized Lebesgue measure. If we indeed closely inspect the proof of Theorem 7.4 in the latter case for example, we see that what is only required is (7.12), that holds similarly, and the volume comparisons. These are however well-known and easy to establish for bounded convex domains in $\mathbb{R}^n$. In this direction, it might be worthwhile mentioning moreover that the first non-zero Neumann eigenvalue λ_1 of the Laplacian on radial functions on the Euclidean ball B in $\mathbb{R}^n$ behaves as n^2. It may be identified indeed as the square of the first positive zero κ_n of the Bessel function $J_{n/2}$ of order $n/2$ (cf. [Cha1] e.g.). (On a sphere of radius r, there will be a factor r^{-2} by homogeneity.) In particular, standard methods or references [Wat] show that $\kappa_n \approx n$ as n is large. Denoting by ρ_0 the logarithmic Sobolev constant on radial functions on B, a simple adaption of the proof of Theorem 7.4 shows that $\rho_0 \leq Cn$ for some numerical constant $C > 0$. Actually, ρ_0 is of the order of n and this may be shown directly in dimension one by a simple analysis of the measure with density nx^{n-1} on the interval $[0,1]$. We are indebted to S. Bobkov for this observation. One can further measure on this example the difference between the spectral gap and the logarithmic Sobolev constant as the dimension n is large. (On general functions, λ_1 and ρ_0 are both of the order of n, see [Bob5].)

As another application, assume $\mathrm{Ric} \geq R > 0$. As we have seen, by the Bakry-Emery inequality [Ba-E], $\rho_0 \geq R_n$ where $R_n = \frac{R}{1-\frac{1}{n}}$. Therefore, by Corollary 7.5,

$$D \leq C\sqrt{\frac{n-1}{R}}\,.$$

Up to the numerical constant, this is just Myers' theorem on the diameter of a compact manifold $D \leq \pi\sqrt{\frac{n-1}{R}}$ (cf. [Cha2]). This could suggest that the best numerical constant in Corollary 7.5 is π^2.

7.4 Diameter and logarithmic Sobolev constant for Markov chains

As in Section 2.1, let $K(x,y)$ be a Markov chain on a finite state space X with symmetric invariant probability measure μ. As before, let ρ_0 be the logarithmic Sobolev constant of (K,μ) defined as the largest $\rho > 0$ such that

$$\rho\,\mathrm{Ent}_\mu(f^2) \leq 2\mathcal{E}(f,f)$$

for every f on X. Recall that here

$$\mathcal{E}(f,f) = \frac{1}{2}\sum_{x,y\in X}\big(f(x)-f(y)\big)^2 K(x,y)\mu\big(\{x\}\big).$$

Recall also we set

$$|||f|||_\infty^2 = \sup\bigl\{\mathcal{E}(gf,f) - \tfrac{1}{2}\,\mathcal{E}(g,f^2); \|g\|_1 \le 1\bigr\}$$

which, as we have seen, takes here the form

$$|||f|||_\infty^2 = \frac{1}{2}\sup_{x\in X}\sum_{y\in X}\bigl(f(x)-f(y)\bigr)^2 K(x,y).$$

As a consequence of Corollary 2.4, for every F such that $|||F|||_\infty \le 1$,

$$\mu\bigl(F \ge \mathrm{E}_\mu(F) + r\bigr) \le \mathrm{e}^{-\rho_0 r^2/4} \tag{7.16}$$

for every $r \ge 0$. If we then define the distance function associated with $|||\cdot|||_\infty$ as

$$d(x,y) = \sup_{|||f|||_\infty \le 1}\bigl[f(x)-f(y)\bigr],$$

we get immediately from (7.16) and (1.28) that for every set A with $\mu(A) > 0$,

$$\mu(A_r) \ge 1 - \mathrm{e}^{-\rho_0 \mu(A)^2 r^2/4} \tag{7.17}$$

where $A_r = \{x; d(x,A) < r\}$. We are thus exactly in the same conditions as in the proof of Theorem 7.4.

Denote by D the diameter of X for the distance d defined above. We can thus state.

Proposition 7.6. *If μ is nearly constant, that is if there exists C such that, for every x, $\mu(\{x\}) \le C\min_{y\in X}\mu(\{y\})$, then*

$$\rho_0 \le \frac{64\log(C|X|)}{D^2}$$

where $|X|$ is the cardinal of X.

Proof. Consider two points $x, y \in X$ such that $d(x,y) = D$. Let B the ball with center x and radius $D/2$. Let A be the set with the largest measure amongst B and B^c. Then $\mu(A) \ge 1/2$. Observe that either x or y is in the complement $(A_r)^c$ of A_r with $r = D/2$. Indeed, if $A = B$, then $(A_r)^c = \{z; d(x,z) \ge D\}$ and $y \in (A_r)^c$ because $d(x,y) = D$; if $A = B^c$, $x \in (A_r)^c$ because $d(x,A) > D/2$. Hence (7.17) yields

$$\min_{z\in X}\mu\bigl(\{z\}\bigr) \le \mathrm{e}^{-\rho_0 D^2/64}.$$

Since, by the hypothesis on μ, $\min_{z\in X}\mu(\{z\}) \ge (C|X|)^{-1}$, the conclusion follows. □

The distance most often used in the present setting is not d but the combinatoric distance d_c associated with the graph with vertex-set X and edge-set

$\{(x,y) : K(x,y) > 0\}$. This distance can be defined as the minimal number of edges one has to cross to go from x to y. Equivalently,

$$d_c(x,y) = \sup_{\|\nabla f\|_\infty \le 1} \big[f(x) - f(y)\big]$$

where

$$\|\nabla f\|_\infty = \sup\big\{\big|f(x) - f(y)\big|; K(x,y) > 0\big\}.$$

Recall, from Section 2.1, that since $\sum_y K(x,y) = 1$,

$$|||f|||_\infty^2 \le \frac{1}{2}\,\|\nabla f\|_\infty^2.$$

In particular, the combinatoric diameter D_c satisfies $D_c^2 \le D^2/2$.

Let us now survey a number of examples at the light of Proposition 7.6.

Consider the first the hypercube $\{0,1\}^n$ with $K(x,y) = 1/n$ if x, y differ by exactly one coordinate and $K(x,y) = 0$ otherwise. The reversible measure is the uniform distribution and $\rho_0 = 2/n$. Proposition 7.6 tells us that $\rho_0 \le 23/n$.

Consider the Bernoulli-Laplace model of diffusion. This is a Markov chain on the n-sets of an N-set with $n \le N/2$. If the current state is an n-set A, we pick an element x at random in A, an element y at random in A^c and change A to $B = (A \setminus \{x\}) \cup \{y\}$. The kernel K is given by $K(A,B) = 1/[n(N-n)]$ if $|A \cap B| = n-2$ and $K(A,B) = 0$ otherwise. The uniform distribution $\pi(A) = \binom{N}{n}^{-1}$ is the reversible measure. Clearly, $D_c = n$. Hence

$$\rho_0 \le \frac{32 \log \binom{N}{n}}{n^2}\,.$$

In the limit case, $n = N/2$, this yields $\rho_0 \le C/n$ which is the right order of magnitude [L-Y].

Let now the chain random transpositions on the symmetric group S_n. Here, $K(\sigma,\theta) = 2/[n(n-1)]$ if $\theta = \sigma\tau$ for some transposition τ and $K(\sigma,\theta) = 0$ otherwise and $\pi \equiv (n!)^{-1}$, The diameter is $D_c = n-1$ and one knows that ρ_0 is of order $1/n \log n$ [D-SC], [L-Y]. Proposition 7.6 gives

$$\rho_0 (n-1)^2 \le 32 \log(n!).$$

Since we know that $\rho_0 \ge (4n \log n)^{-1}$ [D-SC], we can also conclude from Proposition 7.6 that

$$D \le \sqrt{64 n \log n/\rho_0} \le 16 n \log n.$$

At present writing it is not clear whether or not this bound can be obtained more easily. Note that the upper bound

$$d(x,y) \le \Big(\frac{1}{2} \min_{K(x,y)>0} K(x,y)\Big)^{-1/2} d_c(x,y)$$

only yields $D \leq n^2$, up to a multiplicative constant. It might be worthile observing that in this example, ρ_0 is of order $1/n \log n$ while it has been shown by B. Maurey [Mau1] that concentration (with respect to the combinatoric metric) is satisfied at a rate of the order of $1/n$.

Consider a N-regular graph with N fixed. Let $K(x, y) = 1/N$ if they are neighbors and $K(x, y) = 0$ otherwise. Then $\mu(\{x\}) = 1/|X|$. Assume that the number $N(x, t)$ of elements in the ball $B(x, t)$ with center x and radius t in the combinatoric metric d_c satisfies

$$\forall\, x \in X, \;\; \forall\, t > 0, \quad N(x, 2t) \leq CN(x, t). \tag{7.18}$$

Fix $x, y \in X$ such that $d_c(x, y) = D_c$. Set $A = B(x, D_c/2)$, and let $0 < r < D/4$. Then $B(y, D_c/4)$ is contained in the complement of A_r. Now, by our hypothesis, $N(x, D_c/2) \geq C^{-1}|X|$ and $N(y, D_c/4) \geq C^{-2}|X|$ so that

$$1 - \mu(A_r) \geq C^{-2}, \quad \mu(A) \geq C^{-1}.$$

Reporting in (7.17), we obtain

$$\rho_0 \leq \frac{128 C^2 \log C}{D_c^2}\,.$$

For N and C fixed, this is the right order of magnitude in the class of Cayley graphs of finite groups satisfying the volume doubling condition (7.18). See [D-SC, Theorem 4.1].

As a last example, consider any N-regular graph on a finite set X. Let $K(x, y) = 1/N$ if they are neighbors and $K(x, y) = 0$ otherwise. Then $\mu(\{x\}) = 1/|X|$ and $|X| \leq N^{D_c+1}$ (at least if $N \geq 2$). Thus, we get from Proposition 7.6 that

$$\rho_0 \leq \frac{64(D_c + 1)\log N}{D^2} \leq \frac{92 \log N}{D}\,.$$

Compare with the results of [D-SC] and Section 7.3. This is, in a sense, optimal generically. Indeed, if $|X| \geq 4$, one also have the lower bound [D-SC]

$$\rho_0 \geq \frac{\lambda}{2 D_c \log N}$$

where $1 - \lambda$ is the second largest eigenvalue of K. There are many known families of N-regular graphs (N fixed) such that $|X| \to \infty$ whereas $\lambda \geq \epsilon > 0$ stays bounded away from zero (the so-called expanders graphs). Moreover graphs with this property are “generic” amongst N-regular graphs [Al].

References

[Ai] S. Aida. Uniform positivity improving property, Sobolev inequalities and spectral gaps. J. Funct. Anal. 158, 152–185 (1998).

[A-M-S] S. Aida, T. Masuda, I. Shigekawa. Logarithmic Sobolev inequalities and exponential integrability. J. Funct. Anal. 126, 83–101 (1994).

[A-S] S. Aida, D. Stroock. Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. Math. Res. Lett. 1, 75–86 (1994).

[Al] N. Alon. Eigenvalues and expanders. J. Combin. Theory, Ser. B, 38, 78–88 (1987).

[A-L] C. Ané, M. Ledoux. On logarithmic Sobolev inequalities for continuous time random walks on graphs. Preprint (1998).

[Ba1] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. Ecole d'Eté de Probabilités de St-Flour. Lecture Notes in Math. 1581, 1–114 (1994). Springer-Verlag.

[Ba2] D. Bakry. On Sobolev and logarithmic Sobolev inequalities for Markov semigroups. New trends in Stochastic Analysis. 43–75 (1997). World Scientific.

[Ba-E] D. Bakry, M. Emery. Diffusions hypercontractives. Séminaire de Probabilités XIX. Lecture Notes in Math. 1123, 177–206 (1985). Springer-Verlag.

[Ba-L] D. Bakry, M. Ledoux. Lévy-Gromov's isoperimetric inequality for an infinite dimensional diffusion generator. Invent. math. 123, 259–281 (1996).

[B-L-Q] D. Bakry, M. Ledoux, Z. Qian. Preprint (1997).

[Be] W. Beckner. Personal communication (1998).

[BA-L] G. Ben Arous, M. Ledoux. Schilder's large deviation principle without topology. Asymptotic problems in probability theory: Wiener functionals and asymptotics. Pitman Research Notes in Math. Series 284, 107–121 (1993). Longman.

[B-M1] L. Birgé, P. Massart. From model selection to adaptive estimation. Festschrift for Lucien LeCam: Research papers in Probability and Statistics (D. Pollard, E. Torgersen and G. Yang, eds.) 55–87 (1997). Springer-Verlag.

[B-M2] L. Birgé, P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence (1998). Bernoulli, to appear.

[B-B-M] A. Barron, L. Birgé, P. Massart. Risk bounds for model selection via penalization (1998). Probab. Theory Relat. Fields, to appear.

[Bob1] S. Bobkov. On Gross' and Talagrand's inequalities on the discrete cube. Vestnik of Syktyvkar University, Ser. 1, 1, 12–19 (1995) (in Russian).

[Bob2] S. Bobkov. Some extremal properties of Bernoulli distribution. Probability Theor. Appl. 41, 877–884 (1996).

[Bob3] S. Bobkov. A functional form of the isoperimetric inequality for the Gaussian measure. J. Funct. Anal. 135, 39–49 (1996).

[Bob4] S. Bobkov. An isoperimetric inequality on the discrete cube and an elementary proof of the isoperimetric inequality in Gauss space. Ann. Probability 25, 206–214 (1997).

[Bob5] S. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures (1998). Ann. Probability, to appear.

[B-G] S. Bobkov, F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities (1997). J. Funct. Anal., to appear.

[B-H] S. Bobkov, C. Houdré. Isoperimetric constants for product probability measures. Ann. Probability 25, 184–205 (1997).

[B-L1] S. Bobkov, M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential measure. Probab. Theory Relat. Fields 107, 383–400 (1997).

[B-L2] S. Bobkov, M. Ledoux. On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. J. Funct. Anal. 156, 347–365 (1998).

[Bon] A. Bonami. Etude des coefficients de Fourier des fonctions de $L^p(G)$. Ann. Inst. Fourier 20, 335–402 (1970).

[Bor] C. Borell. The Brunn-Minkowski inequality in Gauss space. Invent. math. 30, 207–216 (1975).

[Br] R. Brooks. On the spectrum of non-compact manifolds with finite volume. Math. Z. 187, 425–437 (1984).

[Cha1] I. Chavel. Eigenvalues in Riemannian geometry. Academic Press (1984).

[Cha2] I. Chavel. Riemannian geometry - A modern introduction. Cambridge Univ. Press (1993).

[Che] S.-Y. Cheng. Eigenvalue comparison theorems and its geometric applications. Math. Z. 143, 289–297 (1975).

[Da] E. B. Davies. Heat kernel and spectral theory. Cambridge Univ. Press (1989).

[D-S] E. B. Davies, B. Simon. Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. J. Funct. Anal. 59, 335–395 (1984).

[De] A. Dembo. Information inequalities and concentration of measure. Ann. Probability 25, 927–939 (1997).

[D-Z] A. Dembo, O. Zeitouni. Transportation approach to some concentration inequalities in product spaces. Elect. Comm. in Probab. 1, 83–90 (1996).

[De-S] J.-D. Deuschel, D. Stroock. Large deviations. Academic Press (1989).

[D-SC] P. Diaconis, L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. Ann. Appl. Prob. 6, 695–750 (1996).

[Eh] A. Ehrhard. Symétrisation dans l'espace de Gauss. Math. Scand. 53, 281–301 (1983).

[G-M] M. Gromov, V. D. Milman. A topological application of the isoperimetric inequality. Amer. J. Math. 105, 843–854 (1983).

[Gr1] L. Gross. Logarithmic Sobolev inequalities. Amer. J. Math. 97, 1061–1083 (1975).

[Gr2] L. Gross. Logarithmic Sobolev inequalities and contractive properties of semigroups. Dirichlet Forms, Varenna 1992. Lect. Notes in Math. 1563, 54–88 (1993). Springer-Verlag.

[G-R] L. Gross, O. Rothaus. Herbst inequalities for supercontractive semigroups. Preprint (1997).

[H-Y] Y. Higuchi, N. Yoshida. Analytic conditions and phase transition for Ising models. Lecture Notes in Japanese (1995).

[H-S] R. Holley, D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. J. Statist. Phys. 46, 1159–1194 (1987).

[H-T] C. Houdré, P. Tetali. Concentration of measure for products of Markov kernels via functional inequalities. Preprint (1997).

[Hs1] E. P. Hsu. Logarithmic Sobolev inequalities on path spaces over Riemannian manifolds. Commun. Math. Phys. 189, 9–16 (1997).

[Hs2] E. P. Hsu. Analysis on Path and Loop Spaces (1996). To appear in IAS/Park City Mathematics Series, Vol. 5, edited by E. P. Hsu and S. R. S. Varadhan, American Mathematical Society and Institute for Advanced Study (1997).

[J-S] W. B. Johnson, G. Schechtman. Remarks on Talagrand's deviation inequality for Rademacher functions. Longhorn Notes, Texas (1987).

[Kl] C. A. J. Klaassen. On an inequality of Chernoff. Ann. Probability 13, 966–974 (1985).

[K-S] A. Korzeniowski, D. Stroock. An example in the theory of hypercontractive semigroups. Proc. Amer. Math. Soc. 94, 87–90 (1985).

[Kw-S] S. Kwapień, J. Szulga. Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces. Ann. Probability 19, 369–379 (1991).

[K-L-O] S. Kwapień, R. Latala, K. Oleszkiewicz. Comparison of moments of sums of independent random variables and differential inequalities. J. Funct. Anal. 136, 258–268 (1996).

[Le1] M. Ledoux. Isopérimétrie et inégalités de Sobolev logarithmiques gaussiennes. C. R. Acad. Sci. Paris, 306, 79–92 (1988).

[Le2] M. Ledoux. Remarks on logarithmic Sobolev constants, exponential integrability and bounds on the diameter. J. Math. Kyoto Univ. 35, 211–220 (1995).

[Le3] M. Ledoux. Isoperimetry and Gaussian Analysis. Ecole d'Eté de Probabilités de St-Flour 1994. Lecture Notes in Math. 1648, 165–294 (1996). Springer-Verlag.

[Le4] M. Ledoux. On Talagrand's deviation inequalities for product measures. ESAIM Prob. & Stat. 1, 63–87 (1996).

[L-T] M. Ledoux, M. Talagrand. Probability in Banach spaces (Isoperimetry and processes). Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag (1991).

[L-Y] T.Y. Lee, H.-T. Yau. Logarithmic Sobolev inequality fo some models of random walks. Preprint (1998).

[Li] P. Li. A lower bound for the first eigenvalue of the Laplacian on a compact manifold. Indiana Univ. Math. J. 28, 1013–1019 (1979).

[Ly] T. Lyons. Random thoughts on reversible potential theory. Summer School in Potentiel Theory, Joensuu 1990. Publications in Sciences 26, 71–114 University of Joensuu.

[MD] C. McDiarmid. On the method of bounded differences. Surveys in Combinatorics. London Math. Soc. Lecture Notes 141, 148–188 (1989). Cambridge Univ. Press.

[Mar1] K. Marton. Bounding $\bar{d}$-distance by information divergence: a method to prove measure concentration. Ann. Probability 24, 857–866 (1996).

[Mar2] K. Marton. A measure concentration inequality for contracting Markov chains. Geometric and Funct. Anal. 6, 556-571 (1997).

[Mar3] K. Marton. Measure concentration for a class of random processes. Probab. Theory Relat. Fields 110, 427–439 (1998).

[Mar4] K. Marton. On a measure concentration of Talagrand for dependent random variables. Preprint (1998).

[Mas] P. Massart. About the constants in Talagrand's deviation inequalities for empirical processes (1998). Ann. Probability, to appear.

[Mau1] B. Maurey. Constructions de suites symétriques. C. R. Acad. Sci. Paris 288, 679–681 (1979).

[Mau2] B. Maurey. Some deviations inequalities. Geometric and Funct. Anal. 1, 188–197 (1991).

[Mi] V. D. Milman. Dvoretzky theorem - Thirty years later. Geometric and Funct. Anal. 2, 455–479 (1992) .

[M-S] V. D. Milman, G. Schechtman. Asymptotic theory of finite dimensional normed spaces. Lecture Notes in Math. 1200 (1986). Springer-Verlag.

[M-W] C. Muller, F. Weissler. Hypercontractivity of the heat semigroup for ultraspherical polynomials and on the n-sphere. J. Funct. Anal. 48, 252–283 (1982).

[O-V] F. Otto, C. Villani. Generalization of an inequality by Talagrand, viewed as a consequence of the logarithmic Sobolev inequality. Preprint (1998).

[Pi] M. S. Pinsker. Information and information stability of random variables and processes. Holden-Day, San Franscico (1964).

[Ro1] O. Rothaus. Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. J. Funct. Anal. 42, 358–367 (1981).

[Ro2] O. Rothaus. Hypercontractivity and the Bakry-Emery criterion for compact Lie groups. J. Funct. Anal. 65, 358–367 (1986).

[Ro3] O. Rothaus. Logarithmic Sobolev inequalities and the growth of L^p norms (1996).

[SC1] L. Saloff-Coste. Convergence to equilibrium and logarithmic Sobolev constant on manifolds with Ricci curvature bounded below. Colloquium Math. 67, 109–121 (1994).

[SC2] L. Saloff-Coste. Lectures on finite Markov chains. Ecole d'Eté de Probabilités de St-Flour 1996. Lecture Notes in Math. 1665, 301–413 (1997). Springer-Verlag.

[Sa] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. Preprint (1998).

[Sc] M. Schmuckenschläger. Martingales, Poincaré type inequalities and deviations inequalities. J. Funct. Anal. 155, 303–323 (1998).

[St] D. Stroock. Logarithmic Sobolev inequalities for Gibbs states. Dirichlet forms, Varenna 1992. Lecture Notes in Math. 1563, 194–228 (1993).

[S-Z] D. Stroock, B. Zegarlinski. The logarithmic Sobolev inequality for continuous spin systems on a lattice. J. Funct. Anal. 104, 299–326 (1992).

[S-T] V. N. Sudakov, B. S. Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. J. Soviet. Math. 9, 9–18 (1978); translated from Zap. Nauch. Sem. L.O.M.I. 41, 14–24 (1974).

[Tak] M. Takeda. On a martingale method for symmetric diffusion process and its applications. Osaka J. Math. 26, 605–623 (1989).

[Ta1] M. Talagrand. An isoperimetric theorem on the cube and the Khintchine-Kahane inequalities. Proc. Amer. Math. Soc. 104, 905–909 (1988).

[Ta2] M. Talagrand. Isoperimetry and integrability of the sum of independent Banach space valued random variables. Ann. Probability 17, 1546–1570 (1989).

[Ta3] M. Talagrand. A new isoperimetric inequality for product measure, and the concentration of measure phenomenon. Israel Seminar (GAFA), Lecture Notes in Math. 1469, 91–124 (1991). Springer-Verlag.

[Ta4] M. Talagrand. Some isoperimetric inequalities and their applications. Proc. of the International Congress of Mathematicians, Kyoto 1990, vol. II, 1011–1024 (1992). Springer-Verlag.

[Ta5] M. Talagrand. Sharper bounds for Gaussian and empirical processes. Ann. Probability 22, 28–76 (1994).

[Ta6] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l'I.H.E.S. 81, 73–205 (1995).

[Ta7] M. Talagrand. A new look at independence. Ann. Probability, 24, 1–34 (1996).

[Ta8] M. Talagrand. New concentration inequalities in product spaces. Invent. math. 126, 505–563 (1996).

[Ta9] M. Talagrand. Transportation cost for Gaussian and other product measures. Geometric and Funct. Anal. 6, 587–600 (1996).

[Wan] F.-Y. Wang. Logarithmic Sobolev inequalities on noncompact Riemannian manifolds. Probab. Theory Relat. Fields 109, 417–424 (1997).

[Wat] G. N. Watson. A treatise on the theory of Bessel functions. Cambridge Univ. Press (1944).

[Z-Y] J. Q. Zhong, H. C. Yang. On the estimate of the first eigenvalue of a compact Riemanian manifold. Sci. Sinica Ser. A 27 (12), 1265–1273 (1984).