École d'Été de Probabilités de Saint-Flour 1994

ISOPERIMETRY AND GAUSSIAN ANALYSIS

Michel Ledoux

Département de Mathématiques,
Laboratoire de Statistique et Probabilités,
Université Paul-Sabatier,
31062 Toulouse (France)

Twenty years after his celebrated St-Flour course
on regularity of Gaussian processes,
I dedicate these notes to my teacher X. Fernique

Table of contents

In memory of A. Ehrhard

The Gaussian isoperimetric inequality, and its related concentration phenomenon, is one of the most important properties of Gaussian measures. These notes aim to present, in a concise and selfcontained form, the fundamental results on Gaussian processes and measures based on the isoperimetric tool. In particular, our exposition will include, from this modern point of view, some of the by now classical aspects such as integrability and tail behavior of Gaussian seminorms, large deviations or regularity of Gaussian sample paths. We will also concentrate on some of the more recent aspects of the theory which deal with small ball probabilities. Actually, the Gaussian concentration inequality will be the opportunity to develop some functional analytic ideas around the concentration of measure phenomenon. In particular, we will see how simple semigroup tools and the geometry of abstract Markov generator may be used to study concentration and isoperimetric inequalities. We investigate in this context some of the deep connections between isoperimetric inequalities and functional inequalities of Sobolev type. We also survey recent work on concentration inequalities in product spaces. Actually, although the main theme is Gaussian isoperimetry and analysis, many ideas and results have a much broader range of applications. We will try to indicate some of the related fields of interest.

The Gaussian isoperimetric and concentration inequalities were developed most vigorously in the study of the functional analytic aspects of probability theory (probability in Banach spaces and its relation to geometry and the local theory of Banach spaces) through the contributions of A. Badrikian, C. Borell, S. Chevet, A. Ehrhard, X. Fernique, H. J. Landau and L. A. Shepp, B. Maurey, V. D. Milman, G. Pisier, V. N. Sudakov and B. S. Tsirel'son, M. Talagrand among others. In particular, the new proof by V. D. Milman of Dvoretzky's theorem on spherical sections of convex bodies started the development of the concentration ideas and of their applications in geometry and probability in Banach spaces. Actually, most of the tools and inspiration come from analysis rather than probability. From this analytical point of view, emphasis is put on inequalities in finite dimension as well as on the fundamental Gaussian measurable structure consisting of the product measure on $\mathbb{R}^{\mathbb{N}}$ when each coordinate is endowed with the standard Gaussian measure. It is no surprise therefore that most of the results, developed in the seventies and eighties, often do not seem familiar to true probabilists, and even analysts on Wiener spaces. The aim of this course is to try to advertise these powerful and useful ideas to the probability community although all the results presented here are known and already appeared elsewhere. In particular, M. Talagrand's ideas and contributions, that strongly influenced the author's comprehension of the subject, take an important part in this exposition.

After a short introduction on isoperimetry, where we present the classical isoperimetric inequality, the isoperimetric inequality on spheres and the Gaussian isoperimetric inequality, our first task, in Chapter 2, will be to develop the concentration of measure phenomenon from a functional analytic point of view based on semigroup theory. In particular we show how the Gaussian concentration inequality may easily be obtained from the commutation property of the Ornstein-Uhlenbeck semigroup. In the last chapter, we further investigate the deep connections between isoperimetric and functional inequalities (Sobolev inequalities, hypercontractivity, heat kernel estimates...). We follow in this matter the ideas of N. Varopoulos in his functional approach to isoperimetric inequalities and heat kernel bounds on groups and manifolds. In Chapter 3, we will survey the remarkable recent isoperimetric and concentration inequalities for product measures of M. Talagrand. This section aims to demonstrate the power of abstract concentration arguments and induction techniques in this setting. These deep ideas appear of potential use in a number of problems in probability and applied probability. In Chapter 4, we present, from the concentration viewpoint, the classical integrability properties and tail behaviors of norms of Gaussian measures or random vectors as well as their large deviations. We also show how the isoperimetric and concentration ideas allow a nontopological approach to large deviations of Gaussian measures. The next chapter deals with the corresponding questions for Wiener chaos as remarkably investigated by C. Borell in the late seventies and early eighties. In Chapter 6, we provide a complete treatment of regularity of Gaussian processes based on the results of R. M. Dudley, X. Fernique, V. N. Sudakov and M. Talagrand. In particular, we present the recent short proof of M. Talagrand, based on concentration, of the necessity of the majorizing measure condition for bounded or continuous Gaussian processes. Chapter 7 is devoted to some of the recent aspects of the study of Gaussian measures, namely small ball probabilities. We also investigate in this chapter some correlation and conditional inequalities for norms of Gaussian measures (which have been applied recently to the support of a diffusion theorem and the Freidlin-Wentzell large deviation principle for stronger topologies on Wiener space). Finally, and as announced, we come back in Chapter 8 to a semigroup approach of the Gaussian isoperimetric inequality based on hypercontractivity. Most chapters are completed with short notes for further reading. We also tried to appropriately complete the list of references although we did not put emphasis on historical details and comments.

I sincerely thank the organizers of the École d'Été de St-Flour for their invitation to present this course. My warmest thanks to Ph. Barbe, M. Capitaine, M. A. Lifshits and W. Stolz for a careful reading of the early version of these notes and to C. Borell and S. Kwapień for several helful comments and indications. Many thanks to P. Baldi, S. Chevet, Ch. Léonard, A. Millet and J. Wellner for their comments, remarks and corrections during the school and to all the participants for their interest in this course.

St-Flour, Toulouse 1994 — Michel Ledoux

1. SOME ISOPERIMETRIC INEQUALITIES

In this first chapter, we present the basic isoperimetric inequalities which form the geometric background of this study. Although we will not directly be concerned with true isoperimetric problems and description of extremal sets later on, these inequalities are at the basis of the concentration inequalities of the next chapter on which most results of these notes will be based. We introduce the isoperimetric ideas with the classical isoperimetric inequality on $\mathbb{R}^n$ but the main result will actually be the isoperimetric property on spheres and its limit version, the Gaussian isoperimetric inequality. More on isoperimetry may be found e.g. in the book [B-Z] as well as in the survey paper [Os] and the references therein.

The classical isoperimetric inequality in $\mathbb{R}^n$ (see e.g. [B-Z], [Ha], [Os]...), which at least in dimension 2 and for convex sets may be considered as one of the oldest mathematical statements (cf. [Os]), asserts that among all compact sets A in $\mathbb{R}^n$ with smooth boundary ∂A and with fixed volume, Euclidean balls are the ones with the minimal surface measure. In other words, whenever $\mathrm{vol}_n(A) = \mathrm{vol}_n(B)$ where B is a ball (and $n > 1$),

$$\mathrm{vol}_{n-1}(\partial A) \geq \mathrm{vol}_{n-1}(\partial B). \tag{1.1}$$

There is an equivalent, although less familiar, formulation of this result in terms of isoperimetric neighborhoods or enlargements which in particular avoids surface measures and boundary considerations; namely, if A_r denotes the (closed) Euclidean neighborhood of A of order $r \geq 0$, and if B is as before a ball with the same volume as A, then, for every $r \geq 0$,

$$\mathrm{vol}_n(A_r) \geq \mathrm{vol}_n(B_r). \tag{1.2}$$

Note that A_r is simply the Minkowski sum $A + B(0,r)$ of A and of the (closed) Euclidean ball $B(0,r)$ with center the origin and radius r. The equivalence between (1.1) and (1.2) follows from the Minkowski content formula

$$\mathrm{vol}_{n-1}(\partial A) = \liminf_{r \to 0} \frac{1}{r} \big[\mathrm{vol}_n(A_r) - \mathrm{vol}_n(A)\big]$$

(whenever the boundary ∂A of A is regular enough). Actually, if we take the latter as the definition of $\mathrm{vol}_{n-1}(\partial A)$, it is not too difficult to see that (1.1) and (1.2) are equivalent for every Borel set A (see Chapter 8 for a related result). The simplest proof of this isoperimetric inequality goes through the Brunn-Minkowski inequality which states that if A and B are two compact sets in $\mathbb{R}^n$, then

$$\mathrm{vol}_n(A+B)^{1/n} \geq \mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n}. \tag{1.3}$$

To deduce the isoperimetric inequality (1.2) from the Brunn-Minkowski inequality (1.3), let $r_0 > 0$ be such that $\mathrm{vol}_n(A) = \mathrm{vol}_n(B(0,r_0))$. Then, by (1.3),

$$\begin{aligned}
\mathrm{vol}_n(A_r)^{1/n} = \mathrm{vol}_n\big(A + B(0,r)\big)^{1/n} &\geq \mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n\big(B(0,r)\big)^{1/n} \\
&= \mathrm{vol}_n\big(B(0,r_0)\big)^{1/n} + \mathrm{vol}_n\big(B(0,r)\big)^{1/n} \\
&= (r_0 + r)\mathrm{vol}_n\big(B(0,1)\big)^{1/n} \\
&= \mathrm{vol}_n\big(B(0,r_0+r)\big)^{1/n} = \mathrm{vol}_n\big(B(0,r_0)_r\big)^{1/n}.
\end{aligned}$$

As an illustration of the methods, let us briefly sketch the proof of the Brunn-Minkowski inequality (1.3) following [Mi-S] (for an alternate simple proof, see [Pi3]). By a simple approximation procedure, we may assume that each of A and B is a union of finitely many disjoint sets, each of which is a product of intervals with edges parallel to the coordinate axes. The proof is by induction on the total number p of these rectangular boxes in A and B. If $p = 2$, that is if A and B are products of intervals with sides of length $(a_i)_{1\leq i\leq n}$ and $(b_i)_{1\leq i\leq n}$ respectively, then

$$\begin{aligned}
\frac{\mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n}}{\mathrm{vol}_n(A+B)^{1/n}} &= \prod_{i=1}^n \left(\frac{a_i}{a_i+b_i}\right)^{1/n} + \prod_{i=1}^n \left(\frac{b_i}{a_i+b_i}\right)^{1/n} \\
&\leq \frac{1}{n}\sum_{i=1}^n \frac{a_i}{a_i+b_i} + \frac{1}{n}\sum_{i=1}^n \frac{b_i}{a_i+b_i} = 1
\end{aligned}$$

where we have used the inequality between geometric and arithmetic means. Now, assume that A and B consist of a total of $p > 2$ products of intervals and that (1.3) holds for all sets A' and B' which are composed of a total of at most $p-1$ rectangular boxes. We may and do assume that the number of rectangular boxes in A is at least 2. Parallel shifts of A and B do not change the volume of A, B or $A+B$. Take then a shift of A with the property that one of the coordinate hyperplanes divides A in such a way that there is at least one rectangular box in A on each side of this hyperplane. Therefore A is the union of A' and A'' where A' and A'' are disjoint unions of a number of rectangular boxes strictly smaller than the number in A. Now shift B parallel to the coordinate axes in such a manner that the same hyperplane divides B into B' and B'' with

$$\frac{\mathrm{vol}_n(B')}{\mathrm{vol}_n(B)} = \frac{\mathrm{vol}_n(A')}{\mathrm{vol}_n(A)} = \lambda.$$

Each of B' and B'' has at most the same number of products of intervals as B has. Now, by the induction hypothesis,

$$\begin{aligned}
&\mathrm{vol}_n(A+B)\\
&\qquad \geq \mathrm{vol}_n(A'+B') + \mathrm{vol}_n(A''+B'')\\
&\qquad \geq \big[\mathrm{vol}_n(A')^{1/n} + \mathrm{vol}_n(B')^{1/n}\big]^n + \big[\mathrm{vol}_n(A'')^{1/n} + \mathrm{vol}_n(B'')^{1/n}\big]^n\\
&\qquad = \lambda\big[\mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n}\big]^n + (1-\lambda)\big[\mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n}\big]^n\\
&\qquad = \big[\mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n}\big]^n
\end{aligned}$$

which is the result. Note that, by concavity, (1.3) implies (is actually equivalent to the fact) that, for every λ in $[0,1]$,

$$\mathrm{vol}_n\big(\lambda A + (1-\lambda)B\big) \geq \big[\lambda \mathrm{vol}_n(A)^{1/n} + (1-\lambda)\mathrm{vol}_n(B)^{1/n}\big]^n \geq \mathrm{vol}_n(A)^{\lambda}\mathrm{vol}_n(B)^{1-\lambda}.$$

In the probabilistic applications, it is the isoperimetric inequality on spheres, rather than the classical isoperimetric inequality, which is of fundamental importance. The use of the isoperimetric inequality on spheres in analysis and probability goes back to the new proof, by V. D. Milman [Mi1], [Mi3], of the famous Dvoretzky theorem on spherical sections of convex bodies [Dv]. Since then, it has been used extensively in the local theory of Banach spaces (see [F-L-M], [Mi-S], [Pi3]...) and in probability theory via its Gaussian version (see below). The purpose of this course is actually to present a complete account on the Gaussian isoperimetric inequality and its probabilistic applications. For the applications to Banach space theory, we refer to [Mi-S], [Pi1], [Pi3].

Very much as (1.1), the isoperimetric inequality on spheres expresses that among all subsets with fixed volume on a sphere, geodesic balls (caps) achieve the minimal surface measure. This inequality has been established independently by E. Schmidt [Sch] and P. Lévy [Lé] in the late forties (but apparently for sets with smooth boundaries). Schmidt's proof is based on the classical isoperimetric rearrangement or symmetrization techniques due to J. Steiner (see [F-L-M] for a complete proof along these lines, perhaps the first in this generality). A nice two-point symmetrization technique may also be used (see [Be2]). Lévy's argument, which applies to more general types of surfaces, uses the modern tools of minimal hypersurfaces and integral currents. His proof has been generalized to Riemannian manifolds with positive Ricci curvature by M. Gromov [Gro], [Mi-S], [G-H-L]. Let M be a compact connected Riemannian manifold of dimension N (≥ 2), and let d be its Riemannian metric and μ its normalized Riemannian measure. Denote by $R(M)$ the infimum of the Ricci tensor $\mathrm{Ric}\,(\cdot,\cdot)$ of M over all unit tangent vectors. Recall that if S^N_ρ is the sphere of radius $\rho > 0$ in $\mathbb{R}^{N+1}$, $R(S^N_\rho) = (N-1)/\rho^2$ (see [G-H-L]). We denote below by σ^N_ρ the normalized rotation invariant measure on S^N_ρ. If A is a subset of M, we let as before $A_r = \{x \in M; d(x,A) \leq r\}$, $r \geq 0$.

Theorem 1.1. *Assume that $R(M) = R > 0$ and let S^N_ρ be the manifold of constant curvature equal to R (i.e. ρ is such that $R(S^N_\rho) = (N-1)/\rho^2 = R$). Let A be* *measurable in M and let B be a geodesic ball, or cap, of S_ρ^N such that $\mu(A) \geq \sigma_\rho^N(B)$. Then, for every $r \geq 0$,*

$$\mu(A_r) \geq \sigma_\rho^N(B_r). \tag{1.4}$$

Theorem 1.1 of course applies to the sphere S_ρ^N itself. Equality in (1.4) occurs only if M is a sphere and A a cap on this sphere. Notice furthermore that Theorem 1.1 applied to sets the diameter of which tends to zero contains the classical isoperimetric inequality in Euclidean space. We refer to [Gro], [Mi-S] or [G-H-L] for the proof of Theorem 1.1.

Theorem 1.1 is of particular interest in probability theory via its limit version which gives rise to the Gaussian isoperimetric inequality, our tool of fundamental importance in this course. The Gaussian isoperimetric inequality may indeed be considered as the limit of the isoperimetric inequality on the spheres S_ρ^N when the dimension N and the radius ρ both tend to infinity in the geometric ($R(S_\rho^N) = (N-1)/\rho^2$) and probabilistic ratio $\rho^2 = N$. It has indeed been known for some time that the measures $\sigma_{\sqrt{N}}^N$ on $S_{\sqrt{N}}^N$, projected on a fixed subspace $\mathbb{R}^n$, converge when N goes to infinity to the canonical Gaussian measure on $\mathbb{R}^n$. To be more precise, denote by $\Pi_{N+1,n}$, $N \geq n$, the projection from $\mathbb{R}^{N+1}$ onto $\mathbb{R}^n$. Let γ_n be the canonical Gaussian measure on $\mathbb{R}^n$ with density $\varphi_n(x) = (2\pi)^{-n/2}\exp(-|x|^2/2)$ with respect to Lebesgue measure (where $|x|$ is the Euclidean norm of $x \in \mathbb{R}^n$).

Lemma 1.2. *For every Borel set A in $\mathbb{R}^n$,*

$$\lim_{N\to\infty} \sigma_{\sqrt{N}}^N\big(\Pi_{N+1,n}^{-1}(A) \cap S_{\sqrt{N}}^N\big) = \gamma_n(A).$$

Lemma 1.2 is commonly known as Poincaré's lemma [MK] although it does not seem to be due to H. Poincaré (cf. [D-F]). The convergence is better than only weak convergence of the sequence of measures $\Pi_{N+1,n}(\sigma_{\sqrt{N}}^N)$ to γ_n. Simple analytic or probabilistic proofs of Lemma 1.2 may be found in the literature ([Eh1], [Gal], [Fe5], [D-F]...). The following elegant proof was kindly communicated to us by J. Rosiński.

Proof. Let $(g_i)_{i\geq 1}$ be a sequence of independent standard normal random variables. For every integer $N \geq 1$, set $R_N^2 = g_1^2 + \cdots + g_N^2$. Now, $(\sqrt{N}/R_{N+1}) \cdot (g_1, \ldots, g_{N+1})$ is equal in distribution to $\sigma_{\sqrt{N}}^N$, and thus $(\sqrt{N}/R_{N+1}) \cdot (g_1, \ldots, g_n)$ is equal in distribution to $\Pi_{N+1,n}(\sigma_{\sqrt{N}}^N)$ ($N \geq n$). Since $R_N^2/N \to 1$ almost surely by the strong law of large numbers, we already get the weak convergence result. Lemma 1.2 is however stronger since convergence is claimed for every Borel set. In order to get the full conclusion, notice that R_n^2, $R_{N+1}^2 - R_n^2$ and $(g_1, \ldots, g_n)/R_n$ are independent. Therefore R_n^2/R_{N+1}^2 is independent of $(g_1, \ldots, g_n)/R_n$ and has beta distribution β with parameters $n/2$, $(N+1-n)/2$. Now,

$$\begin{aligned}\sigma_{\sqrt{N}}^N\big(\Pi_{N+1,n}^{-1}(A) \cap S_{\sqrt{N}}^N\big) &= \mathbb{P}\Big\{\frac{\sqrt{N}}{R_{N+1}}\,(g_1, \ldots, g_n) \in A\Big\} \\ &= \mathbb{P}\Big\{\Big(N\,\frac{R_n^2}{R_{N+1}^2}\Big)^{1/2} \cdot \frac{1}{R_n}\,(g_1, \ldots, g_n) \in A\Big\}.\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sigma^N_{\sqrt{N}}\big(\Pi^{-1}_{N+1,n}(A)\cap S^N_{\sqrt{N}}\big)\\
&\quad=\beta\big(\tfrac{n}{2},\tfrac{N+1-n}{2}\big)^{-1}\int_{S_1^{n-1}}\int_0^1 I_A\big(\sqrt{Nt}x\big)t^{\frac{n}{2}-1}(1-t)^{\frac{N+1-n}{2}-1}d\sigma_1^{n-1}(x)dt\\
&\quad=\beta\big(\tfrac{n}{2},\tfrac{N+1-n}{2}\big)^{-1}\frac{2}{N^{n/2}}\int_{S_1^{n-1}}\int_0^{\sqrt{N}} I_A(ux)u^{n-1}\Big(1-\frac{u^2}{N}\Big)^{\frac{N+1-n}{2}-1}d\sigma_1^{n-1}(x)du
\end{aligned}$$

by the change of variables $u=\sqrt{Nt}$. Letting $N\to\infty$, the last integral converges by the dominated convergence theorem to

$$\frac{2}{2^{n/2}\Gamma(n/2)}\int_{S_1^{n-1}}\int_0^{\infty} I_A(ux)u^{n-1}\mathrm{e}^{-u^2/2}d\sigma_1^{n-1}(x)du$$

which is precisely $\gamma_n(A)$ in polar coordinates. The proof of Lemma 1.2 is thus complete. This proof is easily modified to actually yield uniform convergence of densities on compacts sets ([Eh1], [Gal], [Fe5]) and in the variation metric [D-F]. □

As we have seen, caps are the extremal sets of the isoperimetric problem on spheres. Now, a cap may be regarded as the intersection of a sphere and a half-space, and, by Poincaré's limit, caps will thus converge to half-spaces. There are therefore strong indications that half-spaces will be the extremal sets of the isoperimetric problem for Gaussian measures. A half-space H in $\mathbb{R}^n$ is defined as

$$H=\{x\in\mathbb{R}^n;\langle x,u\rangle\le a\}$$

for some real number a and some unit vector u in $\mathbb{R}^n$. The isoperimetric inequality for the canonical Gaussian measure γ_n in $\mathbb{R}^n$ may then be stated as follows. If A is a set in $\mathbb{R}^n$, A_r denotes below its Euclidean neighborhood of order $r\ge 0$.

Theorem 1.3. *Let A be a Borel set in $\mathbb{R}^n$ and let H be a half-space such that $\gamma_n(A)\ge\gamma_n(H)$. Then, for every $r\ge 0$,*

$$\gamma_n(A_r)\ge\gamma_n(H_r).$$

Since γ_n is both rotation invariant and a product measure, the measure of a half-space is actually computed in dimension one. Denote by Φ the distribution function of γ_1, that is

$$\Phi(t)=\int_{-\infty}^{t}\mathrm{e}^{-x^2/2}\frac{dx}{\sqrt{2\pi}},\quad t\in\mathbb{R}.$$

Then, if $H=\{x\in\mathbb{R}^n;\langle x,u\rangle\le a\}$, $\gamma_n(H)=\Phi(a)$, and Theorem 1.3 expresses equivalently that when $\gamma_n(A)\ge\Phi(a)$, then

$$\gamma_n(A_r)\ge\Phi(a+r)\tag{1.5}$$

for every $r \geq 0$. In other words, if Φ^{-1} is the inverse function of Φ, for every Borel set A in $\mathbb{R}^n$ and every $r \geq 0$,

$$\Phi^{-1}\big(\gamma_n(A_r)\big) \geq \Phi^{-1}\big(\gamma_n(A)\big) + r. \tag{1.6}$$

The Gaussian isoperimetric inequality is thus essentially dimension free, a characteristic feature of the Gaussian setting.

Proof of Theorem 1.3. We prove (1.5) and use the isoperimetric inequality on spheres (Theorem 1.1) and Lemma 1.2. We may assume that $a = \Phi^{-1}(\gamma_n(A)) > -\infty$. Let then $b \in (-\infty, a)$. Since $\gamma_n(A) > \Phi(b) = \gamma_1((-\infty, b])$, by Lemma 1.2, for every N ($\geq n$) large enough,

$$\sigma^N_{\sqrt{N}}\big(\Pi^{-1}_{N+1,n}(A) \cap S^N_{\sqrt{N}}\big) > \sigma^N_{\sqrt{N}}\big(\Pi^{-1}_{N+1,1}(]-\infty, b]) \cap S^N_{\sqrt{N}}\big). \tag{1.7}$$

It is easy to see that $\Pi^{-1}_{N+1,n}(A_r) \cap S^N_{\sqrt{N}} \supset \big(\Pi^{-1}_{N+1,n}(A) \cap S^N_{\sqrt{N}}\big)_r$ where the neighborhood of order r on the right hand side is understood with respect to the geodesic distance on $S^N_{\sqrt{N}}$. Since $\Pi^{-1}_{N+1,1}((-\infty, b]) \cap S^N_{\sqrt{N}}$ is a cap on $S^N_{\sqrt{N}}$, by (1.7) and the isoperimetric inequality on spheres (Theorem 1.1),

$$\begin{aligned}\sigma^N_{\sqrt{N}}\big(\Pi^{-1}_{N+1,n}(A_r) \cap S^N_{\sqrt{N}}\big) &\geq \sigma^N_{\sqrt{N}}\big(\big(\Pi^{-1}_{N+1,n}(A) \cap S^N_{\sqrt{N}}\big)_r\big)\\ &\geq \sigma^N_{\sqrt{N}}\big(\big(\Pi^{-1}_{N+1,1}((-\infty, b]) \cap S^N_{\sqrt{N}}\big)_r\big).\end{aligned}$$

Now, $\big(\Pi^{-1}_{N+1,1}((-\infty, b]) \cap S^N_{\sqrt{N}}\big)_r = \Pi^{-1}_{N+1,1}((-\infty, b + r(N)]) \cap S^N_{\sqrt{N}}$ where (for N large)

$$r(N) = \sqrt{N}\cos\big[\arccos\big(b/\sqrt{N}\big) - r/\sqrt{N}\big] - b.$$

Since $\lim r(N) = r$, by Lemma 1.2 again, $\gamma_n(A_r) \geq \Phi(b+r)$. Since $b < a$ is arbitrary, the conclusion follows. □

Theorem 1.3 is due independently to C. Borell [Bo2] and to V. N. Sudakov and B. S. Tsirel'son [S-T] with the same proof based on the isoperimetric inequality on spheres and Poincaré's limit. A. Ehrhard [Eh2] (see also [Eh3], [Eh5]) gave a different proof using an intrinsic Gaussian symmetrization procedure similar to the Steiner symmetrization used by E. Schmidt in his proof of Theorem 1.1. In any case, Ehrhard's proof or the proof of isoperimetry on spheres are rather delicate, as it is usually the case with isoperimetric inequalities and the description of their extremal sets.

With this same Gaussian symmetrization tool, A. Ehrhard [Eh2] established furthermore a Brunn-Minkowski type inequality for γ_n, however only for convex sets. More precisely, he showed that whenever A and B are convex sets in $\mathbb{R}^n$, for every $\lambda \in [0,1]$,

$$\Phi^{-1}\big(\gamma_n\big(\lambda A + (1-\lambda)B\big)\big) \geq \lambda\Phi^{-1}\big(\gamma_n(A)\big) + (1-\lambda)\Phi^{-1}\big(\gamma_n(B)\big). \tag{1.8}$$

It might be worthwhile noting that if we apply this inequality to B the Euclidean ball with center the origin and radius $r/(1-\lambda)$ and let λ tend to one, we recover inequality (1.6) (for A convex). However, it is still an open problem to know whether (1.8) holds true for every Borel sets A and B and not only convex sets*. It would improve upon the more classical logconcavity of Gaussian measures (cf. [Bo1]) that states that, for every Borel sets A and B, and every $\lambda \in [0,1]$,

$$\gamma_n\big(\lambda A + (1-\lambda)B\big) \geq \gamma_n(A)^\lambda \gamma_n(B)^{1-\lambda}. \tag{1.9}$$

As another inequality of interest, let us note that if A is a Borel set with $\gamma_n(A) = \Phi(a)$, and if $h \in \mathbb{R}^n$,

$$\gamma_n(A+h) \leq \Phi\big(a + |h|\big). \tag{1.10}$$

By rotational invariance, we may assume that $h = re_1$, $r = |h|$, where e_1 is the first unit vector on $\mathbb{R}^n$, changing A into some new set A' with $\gamma_n(A) = \gamma_n(A') = \Phi(a)$. Then, by the translation formula for γ_n,

$$\begin{aligned} \mathrm{e}^{r^2/2}\gamma_n(A'+h) &= \int_{A'} \mathrm{e}^{-rx_1} d\gamma_n(x) \\ &\leq \int_{A' \cap \{x_1 \leq a\}} \mathrm{e}^{-rx_1} d\gamma_n(x) + \mathrm{e}^{-ra}\gamma_n\big(A' \cap \{x_1 > a\}\big). \end{aligned}$$

Since $\gamma_n(A' \cap \{x_1 > a\}) = \gamma_n((A')^c \cap \{x_1 \leq a\})$ where $(A')^c$ is the complement of A',

$$\begin{aligned} \mathrm{e}^{r^2/2}\gamma_n(A'+h) &\leq \int_{A' \cap \{x_1 \leq a\}} \mathrm{e}^{-rx_1} d\gamma_n(x) + \mathrm{e}^{-ra}\gamma_n\big((A')^c \cap \{x_1 \leq a\}\big) \\ &\leq \int_{\{x_1 \leq a\}} \mathrm{e}^{-rx_1} d\gamma_n(x) \\ &= \mathrm{e}^{r^2/2}\gamma_n(x; x_1 \leq a + r) = \mathrm{e}^{r^2/2}\Phi(a+r). \end{aligned}$$

The claim (1.10) follows.

Notes for further reading. In addition to the preceding open problem, the following conjecture is still open. Is it true that for every symmetric closed convex set A in $\mathbb{R}^n$,

$$\gamma_n(\lambda A) \geq \gamma_n(\lambda S) \tag{1.11}$$

for each $\lambda > 1$, where S is a symmetric strip such that $\gamma_n(A) = \gamma_n(S)$? This conjecture has been known since an unpublished preprint by L. Shepp on the existence of strong exponential moments of Gaussian measures (cf. Chapter 4 and [L-S]). Recent work of S. Kwapień and J. Sawa [K-S] shows that the conjecture is true under the additional assumption that A is sufficiently symmetric (A is an ellipsoid for example). Examples of isoperimetric processes in probability theory are presented in [Bo11].

---

* During the school, R. Latała [La] proved that (1.8) holds when only one of the two sets A and B is convex. Thus, due to the preceding comment, the Brunn-Minkowski principle generalizes to the Gaussian setting.

2. THE CONCENTRATION OF MEASURE PHENOMENON

In this section, we present the concentration of measure phenomenon which was most vigourously put forward by V. D. Milman in the local theory of Banach spaces (cf. [Mi2], [Mi3], [Mi-S]). Isoperimetry is concerned with infinitesimal neighborhoods and surface areas and with extremal sets. The concentration of measure phenomenon rather concerns the behavior of "large" isoperimetric neighborhoods. Although of basic isoperimetric inspiration, the concentration of measure phenomenon is a milder property that may be shown, as we will see, to be satisfied in a large number of settings, sometimes rather far from the geometrical frame of isoperimetry. It roughly states that if a set $A \subset X$ has measure at least one half, "most" of the points in X are "close" to A. The main task is to make precise the meaning of the words "most" and "close" in the examples of interest. Moreover, new tools may be used to establish concentration inequalities. In particular, we will present in this chapter simple semigroup and probabilistic proofs of both the concentration inequalities on spheres and in Gauss space. In chapter 8, we further develop the functional approach and try to reach with these tools the full isoperimetric statements.

As we mentioned it at the end of the preceding chapter, isoperimetric inequalities and description of their extremal sets are often rather delicate, if not unknown. However, in almost all the applications presented here, the Gaussian isoperimetric inequality is only used in the form of the corresponding concentration inequality. Since the latter will be established here in an elementary way, it can be freely used in the applications.

In the setting of Theorem 1.1, if A is a set on M with sufficiently large measure, for example if $\mu(A) \geq \frac{1}{2}$, then, by the explicit expression of the measure of a cap, we get that, for every $r \geq 0$

$$\mu(A_r) \geq 1 - \exp\left(-R\,\frac{r^2}{2}\right), \tag{2.1}$$

that is a Gaussian bound, only depending on R, on the complement of the neighborhood of order r of A, uniformly in those A's such that $\mu(A) \geq \frac{1}{2}$. More precisely, if $\mu(A) \geq \frac{1}{2}$, for "most" x's in M, there exists y in A within distance $1/\sqrt{R}$ of x. Of course, the ratio $1/\sqrt{R}$ is in general much smaller than the diameter of the manifold (see below the example of S_1^N). Equivalently, let f be a Lipschitz map on M and let m be a median of f for μ (i.e. $\mu(f \geq m) \geq \frac{1}{2}$ and $\mu(f \leq m) \geq \frac{1}{2}$). If we apply (2.1) to the set $A = \{f \leq m\}$, it easily follows that, for every $r \geq 0$,

$$\mu(f \geq m + r) \leq \exp\left(-\frac{R r^2}{2\|f\|_{\mathrm{Lip}}^2}\right).$$

Together with the corresponding inequality for $A = \{f \geq m\}$, for every $r \geq 0$,

$$\mu\big(|f - m| \geq r\big) \leq 2\exp\left(-\frac{R r^2}{2\|f\|_{\mathrm{Lip}}^2}\right). \tag{2.2}$$

Thus, f is concentrated around some mean value with a large probability depending on some exponential of the ratio $R/\|f\|_{\mathrm{Lip}}^2$. This property has taken the name of concentration of measure phenomenon (cf. [G-M], [Mi-S]).

The preceding bounds are of particular interest for families of probability measures such as for example the measures σ_1^N on the unit spheres S_1^N as N tends to infinity for which (2.2) becomes (since $R(S_1^N) = N - 1$),

$$\sigma_1^N\big(|f - m| \geq r\big) \leq 2\exp\left(-\frac{(N-1)r^2}{2\|f\|_{\mathrm{Lip}}^2}\right).$$

Think thus of the dimension N to be large. Of course, if $\|f\|_{\mathrm{Lip}} \leq 1$, for every x, y in S_1^N, $|f(x) - f(y)| \leq \pi$. But the preceding concentration inequality tells us that, already for r of the order of $1/\sqrt{N}$, $|f(x) - m| \leq r$ on a large set (in the sense of the measure) of x's. It is then from the interplay, in this inequality, between N large, r of the order of $1/\sqrt{N}$ and the respective values of m and $\|f\|_{\mathrm{Lip}}$ for f the gauge of a convex body that V. D. Milman draws the information in order to choose at random the Euclidean sections of the convex body and to prove in this way Dvoretzky's theorem (see [Mi1], [Mi3], [Mi-S]).

Another (this time noncompact) concentration example is of course the Gaussian measure γ_n on $\mathbb{R}^n$ (the canonical Gaussian measure on $\mathbb{R}^n$ with density with respect to Lebesgue measure $(2\pi)^{-n/2}\exp(-|x|^2/2)$). If $\gamma_n(A) \geq \frac{1}{2}$, we may take $a = 0$ in (1.5) and thus, for every $r \geq 0$,

$$\gamma_n(A_r) \geq \Phi(r) \geq 1 - \frac{1}{2}\,\mathrm{e}^{-r^2/2}. \tag{2.3}$$

Let f be a Lipschitz function on $\mathbb{R}^n$ with Lipschitz (semi-) norm

$$\|f\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$$

(where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^n$) and let m be a median of f for γ_n. As efore, it follows from (2.3) that for every $r \geq 0$,

$$\gamma_n\big(|f - m| \geq r\big) \leq 2\big(1 - \Phi(r/\|f\|_{\mathrm{Lip}})\big) \leq \exp\left(-\frac{r^2}{2\|f\|_{\mathrm{Lip}}^2}\right). \tag{2.4}$$

Thus, for r of the order of $\|f\|_{\mathrm{Lip}}$, $|f-m| \le r$ on "most" of the space. The word "most" is described here by a Gaussian bound.

Isoperimetric and concentration inequalities involve both a measure and a metric structure (to define the isoperimetric neighborhoods or enlargements). On an abstract (Polish) metric space (X,d) equipped with a probability measure μ, or a family of probability measures, the concentration of measure phenomenon may be described via the concentration function

$$\alpha(r) = \alpha\big((X,d;\mu);r\big) = \sup\big\{1-\mu(A_r);\ A\subset X, \mu(A)\ge \tfrac{1}{2}\big\}, \quad r\ge 0.$$

It is a remarkable property that this concentration function may be controlled in a rather large number of cases, and very often by a Gaussian decay as above. Isoperimetric tools are one of the most important and powerful arguments used to establish concentration inequalities. However, since we are concerned here with enlargements A_r for (relatively) large values of r rather than infinitesimal values, the study of the concentration phenomenon can be quite different from the study of isoperimetric inequalities, both in establishing new concentration inequalities and in applying them. Indeed, the framework of concentration inequalities is less restrictive than the isoperimetric setting as we will see for example in the next chapter, due mainly to the fact that we are not looking here for the extremal sets.

New tools to establish concentration inequalities were thus developed. For example, M. Gromov and V. D. Milman [G-M] showed that if X is a compact Riemannian manifold, for every $r\ge 0$,

$$\alpha(r) \le C\exp\big(-c\sqrt{\lambda_1}\,r\big)$$

(with $C=\frac{3}{4}$ and $c=\log(\frac{3}{2})$) where λ_1 is the first nontrivial eigenvalue of the Laplacian on X (see also [A-M-S] for a similar result in an abstract setting). In case $R(X)>0$, this is however weaker than (2.1). They also developed in this paper [G-M] several topological applications of concentration such as fixed point theorems. On the probabilistic side, some martingale inequalities put forward by B. Maurey [Ma1] have been used in the local theory of Banach spaces in extensions of Dvoretzky's theorem (cf. [Ma2], [Mi-S], [Pi1]). The main idea consists in writing, for a well-behaved function f, the difference $f-\mathbb{E}(f)$ as a sum of martingale differences $d_i = \mathbb{E}(f|\mathcal{F}_i) - \mathbb{E}(f|\mathcal{F}_{i-1})$ where $(\mathcal{F}_i)_i$ is some (finite) filtration. The classical arguments on sums of independent random variables then show in the same way that if $\sum_i \|d_i\|_\infty^2 \le 1$, for every $r\ge 0$,

$$\mathbb{P}\big\{|f-\mathbb{E}(f)|\ge r\big\} \le 2\mathrm{e}^{-r^2/2} \tag{2.5}$$

([Azu], [Ma1]). As a corollary, one can deduce from this result the concentration of Haar measure μ on $\{0,1\}^n$ equipped with the Hamming metric as

$$\alpha(r) \le C\exp\bigg(-\frac{r^2}{Cn}\bigg)$$

for some numerical constant $C>0$. This property may be established from the corresponding isoperimetric inequality ([Har], [W-W]), but V. D. Milman et G. Schechtman [Mi-S] deduce it from inequality (2.5) (see Chapter 3).

Our first aim here will be to give a simple proof of the concentration inequality (2.4). The proof is based on the Hermite or Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$ defined, for every well-behaved function f on $\mathbb{R}^n$, by (Mehler's formula)

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

and more precisely on its commutation property

$$\nabla P_t f = \mathrm{e}^{-t} P_t(\nabla f). \tag{2.6}$$

The generator L of $(P_t)_{t\geq 0}$ is given by $Lf(x) = \Delta f(x) - \langle x, \nabla f(x)\rangle$, f smooth enough on $\mathbb{R}^n$, and we have the integration by parts formula

$$\int f(-Lg)d\gamma_n = \int \langle \nabla f, \nabla g\rangle d\gamma_n$$

for all smooth functions f and g on $\mathbb{R}^n$.

Proposition 2.1. *Let f be a Lipschitz function on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \leq 1$ and $\int f d\gamma_n = 0$. Then, for every real number λ,*

$$\int \mathrm{e}^{\lambda f} d\gamma_n \leq \mathrm{e}^{\lambda^2/2}. \tag{2.7}$$

Before turning to the proof of this proposition, let us briefly indicate how to deduce a concentration inequality from (2.7). Let f be any Lipschitz function on $\mathbb{R}^n$. As a consequence of (2.7), for every real number λ,

$$\int \exp\big(\lambda\big(f - \textstyle\int f d\gamma_n\big)\big) d\gamma_n \leq \exp\Big(\frac{1}{2}\,\lambda^2 \|f\|_{\mathrm{Lip}}^2\Big).$$

By Chebyshev's inequality, for every λ and $r \geq 0$,

$$\gamma_n\big(f \geq \textstyle\int f d\gamma_n + r\big) \leq \exp\Big(-\lambda r + \frac{1}{2}\,\lambda^2 \|f\|_{\mathrm{Lip}}^2\Big)$$

and, optimizing in λ,

$$\gamma_n\big(f \geq \textstyle\int f d\gamma_n + r\big) \leq \exp\Big(-\frac{r^2}{2\|f\|_{\mathrm{Lip}}^2}\Big). \tag{2.8}$$

Applying (2.8) to both f and $-f$, we get a concentration inequality similar to (2.4) (around the mean rather than the median)

$$\gamma_n\big(\big|f - \textstyle\int f d\gamma_n\big| \geq r\big) \leq 2\exp\Big(-\frac{r^2}{2\|f\|_{\mathrm{Lip}}^2}\Big). \tag{2.9}$$

Two parameters are thus entering those concentration inequalities, the median or the mean of a function f with respect to γ_n and its Lipschitz norm. These can be very different. For example, if f is the Euclidean norm on $\mathbb{R}^n$, the median or the mean of f are of the order of $\sqrt{n}$ while $\|f\|_{\mathrm{Lip}} = 1$. This is one of the main features of these inequalities (cf. [L-T2]) and is at the basis of the Gaussian proofs of Dvoretzky's theorem (see [Pi1], [Pi3]).

Proof of Proposition 2.1. Let f be smooth enough on $\mathbb{R}^n$ with mean zero and $\|f\|_{\mathrm{Lip}} \leq 1$. For λ fixed, set $G(t) = \int \exp(\lambda P_t f) d\gamma_n$, $t \geq 0$. Since $\|f\|_{\mathrm{Lip}} \leq 1$, it follows from (2.6) that $|\nabla(P_s f)|^2 \leq \mathrm{e}^{-2s}$ almost everywhere for every s. Since $\int f d\gamma_n = 0$, $G(\infty) = 1$. Hence, for every $t \geq 0$,

$$\begin{aligned} G(t) = 1 - \int_t^\infty G'(s)\, ds &= 1 - \lambda \int_t^\infty \Big(\int L(P_s f) \exp\big(\lambda P_s f\big) d\gamma_n \Big) ds \\ &= 1 + \lambda^2 \int_t^\infty \Big(\int \big|\nabla\big(P_s f\big)\big|^2 \exp\big(\lambda P_s f\big) d\gamma_n \Big) ds \\ &\leq 1 + \lambda^2 \int_t^\infty \mathrm{e}^{-2s}\, G(s) ds \end{aligned}$$

where we used integration by parts in the space variable. Let $H(t)$ be the logarithm of the right hand side of this inequality. Then the preceding inequality tells us that $-H'(t) \leq \lambda^2 \mathrm{e}^{-2t}$ for every $t \geq 0$. Therefore

$$\log G(0) \leq H(0) = -\int_0^\infty H'(t)\, dt \leq \frac{\lambda^2}{2}$$

which is the claim of the proposition, at least for a smooth function f. The general case follows from a standard approximation, by considering for example $P_\varepsilon f$ instead of f and by letting then ε tend to zero. The proof is complete. □

Inequalities (2.8) and (2.9) will be our key argument in the study of integrability properties and tail behavior of Gaussian random vectors, as well as in the various applications throughout these notes. While the concentration inequalities (2.4) of isoperimetric nature may of course be used equivalently, we would like to emphasize here the simple proof of Proposition 2.1 from which (2.8) and (2.9) follow. Proposition 2.1 is due to I. A. Ibragimov, V. N. Sudakov and B. S. Tsirel'son [I-S-T] (see also B. Maurey [Pi1, p. 181]). Their argument is actually of more probabilistic nature. For every smooth enough function f on $\mathbb{R}^n$, write

$$f\big(W(1)\big) - \mathbb{E} f\big(W(1)\big) = \int_0^1 \langle \nabla T_{1-t} f\big(W(t)\big) dW(t) \rangle$$

where $(W(t))_{t\geq 0}$ is Brownian motion on $\mathbb{R}^n$ starting at the origin and where $(T_t)_{t\geq 0}$ denotes its associated semigroup (the heat semigroup), with the probabilistic normalization. Note then that the above stochastic integral has the same distribution as $\beta(\tau)$ where $(\beta(t))_{t\geq 0}$ is a one-dimensional Brownian motion and where

$\tau = \int_0^1 |\nabla T_{1-t} f(W(t))|^2 \, dt$. Therefore, for every Lipschitz function f such that $\|f\|_{\mathrm{Lip}} \leq 1$, $\tau \leq 1$ almost surely so that, for all $r \geq 0$,

$$\begin{aligned} \mathbb{P}\big\{ f\big(W(1)\big) - \mathbb{E} f\big(W(1)\big) \geq r \big\} &\leq \mathbb{P}\big\{ \max_{0 \leq t \leq 1} \beta(t) \geq r \big\} \\ &= 2 \int_r^\infty \mathrm{e}^{-x^2/2} \, \frac{dx}{\sqrt{2\pi}} \leq \mathrm{e}^{-r^2/2}. \end{aligned}$$

Since $W(1)$ has distribution γ_n, this is thus simply (2.8).

Proposition 2.1 and its proof may actually be extended to a larger setting to yield for example a simple proof of the concentration (2.2) (up to some numerical constants) on spheres or on Riemannian manifolds M with positive curvature $R(M)$. The proof uses the heat semigroup on M and Bochner's formula. It is inspired by the work of D. Bakry and M. Émery [B-É] (cf. also [D-S]) on hypercontractivity and logarithmic Sobolev inequalities. We will come back to this observation in Chapter 8. We establish the following fact (cf. [Led4]).

Proposition 2.2. *Let M be a compact Riemannian manifold of dimension N (≥ 2) and with $R(M) = R > 0$. Let f be a Lipschitz function on M with $\|f\|_{\mathrm{Lip}} \leq 1$ and assume that $\int f d\mu = 0$. Then, for every real number λ,*

$$\int \mathrm{e}^{\lambda f} \, d\mu \leq \mathrm{e}^{\lambda^2/2R}.$$

Proof. Let ∇ be the gradient on M and Δ be the Laplace-Beltrami operator. By Bochner's formula (see e.g. [G-H-L]), for every smooth function f on M, pointwise

$$\frac{1}{2} \, \Delta\big(|\nabla f|^2\big) - \langle \nabla f, \nabla(\Delta f) \rangle = \mathrm{Ric}(\nabla f, \nabla f) + \big\|\mathrm{Hess}(f)\big\|_2^2.$$

In particular,

$$\frac{1}{2} \, \Delta\big(|\nabla f|^2\big) - \langle \nabla f, \nabla(\Delta f) \rangle \geq R \, |\nabla f|^2 + \frac{1}{N} \, (\Delta f)^2. \tag{2.10}$$

The dimensional term in this inequality will actually not be used and we will only be concerned here with the inequality

$$\frac{1}{2} \, \Delta\big(|\nabla f|^2\big) - \langle \nabla f, \nabla(\Delta f) \rangle \geq R \, |\nabla f|^2. \tag{2.11}$$

Now, consider the heat semigroup $P_t = \mathrm{e}^{t\Delta}$, $t \geq 0$, and let f be smooth on M. Let further $s > 0$ be fixed and set, for every $t \leq s$, $F(t) = P_t(|\nabla(P_{s-t} f)|^2)$. It is an immediate consequence of (2.11) applied to $P_{s-t} f$ that $F'(t) \geq 2RF(t)$, $t \leq s$. Hence, the function $\mathrm{e}^{-2Rt} F(t)$ is increasing on the interval $[0, s]$ and we have thus that, for every $s \geq 0$,

$$\big|\nabla\big(P_s f\big)\big|^2 \leq \mathrm{e}^{-2Rs} P_s\big(|\nabla f|^2\big). \tag{2.12}$$

This relation, which is actually equivalent to (2.11), expresses a commutation property between the respective actions of the semigroup and the gradient (cf. (2.6)). It is the only property which is used in the proof itself which is now exactly as the proof of Proposition 2.1. Let f be smooth on M with $\|f\|_{\rm Lip} \leq 1$ and $\int f d\mu = 0$. For λ fixed, set $G(t) = \int \exp(\lambda P_t f) d\mu$, $t \geq 0$. Since $\|f\|_{\rm Lip} \leq 1$, it follows from (2.12) that $|\nabla(P_s f)|^2 \leq {\rm e}^{-2Rs}$ almost everywhere for every s. Since $\int f d\mu = 0$, $G(\infty) = 1$. Hence, for every $t \geq 0$,

$$\begin{aligned} G(t) = 1 - \int_t^\infty G'(s)\, ds &= 1 - \lambda \int_t^\infty \bigg(\int \Delta\big(P_s f\big) \exp\big(\lambda P_s f\big) d\mu \bigg) ds \\ &= 1 + \lambda^2 \int_t^\infty \bigg(\int \big|\nabla\big(P_s f\big)\big|^2 \exp\big(\lambda P_s f\big) d\mu \bigg) ds \\ &\leq 1 + \lambda^2 \int_t^\infty {\rm e}^{-2Rs}\, G(s) ds \end{aligned}$$

where we used integration by parts in the space variable. The proof is completed as in Proposition 2.1. □

The commutation formula $\nabla P_t f = {\rm e}^{-t} P_t(\nabla f)$ of the Ornstein-Uhlenbeck semigroup expresses equivalently a Bochner formula for the second order generator L of infinite dimension ($N = \infty$) and constant curvature 1 (limit of $R(S^N_{\sqrt{N}})$ when N goes to infinity) of the type (2.10) or (2.11)

$$\frac{1}{2}\, L\big(|\nabla f|^2\big) - \langle \nabla f, \nabla(Lf) \rangle \geq (Lf)^2.$$

The geometry of the Ornstein-Uhlenbeck generator is thus purely infinite dimensional, even on a finite dimensional state space (as the Gaussian isoperimetric inequality itself, cf. Chapter 1). The abstract consequences of these observations are at the origin of the study by D. Bakry and M. Émery of hypercontractive diffusions under curvature-dimension hypotheses [B-É], [Bak]. We will come back to this question in Chapter 8 and actually show, according to [A-M-S], that (2.7) can be deduced directly from hypercontractivity.

At this point, we realized that simple semigroup arguments may be used to establish concentration properties, however on Lipschitz functions rather than sets. It is not difficult however to deduce from Propositions 2.1 and 2.2 inequalities on sets very close to the inequalities which follow from isoperimetry (but still for "large" neighborhoods). We briefly indicate the procedure in the Gaussian setting.

Let A be a Borel set in $\mathbb{R}^n$ with canonical Gaussian measure $\gamma_n(A) > 0$. For every $u \geq 0$, let

$$f_{A,u}(x) = \min\big(d(x,A), u\big)$$

where $d(x,A)$ is the Euclidean distance from the point x to the set A. Clearly $\|f_{A,u}\|_{\rm Lip} \leq 1$ so that we may apply inequality (2.8) to this family of Lipschitz functions when u varies. Let $E_{A,u} = \int f_{A,u} d\gamma_n$. Inequality (2.8) applied to $f_{A,u}$ and $r = u - E_{A,u}$ yields

$$\gamma_n\big(x \in \mathbb{R}^n; \min\big(d(x,A), u\big) \geq u\big) \leq \exp\bigg(-\frac{1}{2}\,(u - E_{A,u})^2\bigg),$$

that is

$$\gamma_n\big(x; x \notin A_u\big) \le \exp\Big(-\frac{1}{2}\,(u - E_{A,u})^2\Big) \tag{2.13}$$

since $d(x, A) > u$ if and only if $x \notin A_u$. We have now to appropriately control the expectations $E_{A,u} = \int f_{A,u} d\gamma_n$, possibly only with u and the measure of A. A first bound is simply $E_{A,u} \le u\,\gamma(A^c)$ which already yields,

$$\gamma_n\big(x; x \notin A_u\big) \le \exp\Big(-\frac{1}{2}\,u^2 \gamma_n(A)^2\Big)$$

for every $u \ge 0$. This inequality may already be compared to (2.3). However, if we use this estimate recursively, we get

$$E_{A,u} = \int_0^u \gamma_n\big(x; d(x, A) > t\big)dt \le \int_0^u \min\big(\gamma_n(A^c), \mathrm{e}^{-t^2\gamma_n(A)^2/2}\big)dt.$$

If we let then $\delta(v)$ be the decreasing function on $(0, 1]$ defined by

$$\delta(v) = \int_0^\infty \min\big(1 - v, \mathrm{e}^{-t^2 v^2/2}\big)dt, \tag{2.14}$$

we have $E_{A,u} \le \delta(\gamma_n(A))$ uniformly in u. Hence, from (2.13), for every $u \ge 0$,

$$\gamma_n\big(x; x \notin A_u\big) \le \exp\Big(-\frac{u^2}{2} + uE_{A,u}\Big) \le \exp\Big(-\frac{u^2}{2} + u\delta\big(\gamma_n(A)\big)\Big).$$

In conclusion, we obtained from Proposition 2.1 and inequality (2.8) that, for every $r \ge 0$,

$$\gamma_n(A_r) \ge 1 - \exp\Big(-\frac{r^2}{2} + r\delta\big(\gamma_n(A)\big)\Big). \tag{2.15}$$

This simple argument thus yields an inequality very similar in nature to the isoperimetric bound (2.3), with however the extra factor $r\delta(\gamma_n(A))$. (Using the preceding recursive argument, one may of course improve further and further this estimate.) Due to the fact that $\delta(\gamma_n(A)) \to 0$ as $\gamma_n(A) \to 1$, this result can be used exactly as the isoperimetric inequality in almost all the applications presented in these notes. We will come back to this in Chapter 4 for example, and we will always use (2.15) rather than isoperimetry in the applications.

We conclude this chapter with a proposition closely related to Proposition 2.1 and the proof of which is similar. It will be used in Chapter 4 in some large deviation statement for the Ornstein-Uhlenbeck process. We only consider the Gaussian setting.

Proposition 2.3. *Let f be a Lipschitz function on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \le 1$. Then, for every real number λ and every $t \ge 0$,*

$$\int\!\!\int \exp\bigl(\lambda\bigl[f\bigl(\mathrm{e}^{-t}x + (1-\mathrm{e}^{-2t})^{1/2}y\bigr) - f(x)\bigr]\bigr)d\gamma_n(x)d\gamma_n(y) \le \exp\bigl(\lambda^2(1-\mathrm{e}^{-t})\bigr).$$

Proposition 2.3 will be used for the small values of the time t. When $t \to \infty$, it is somewhat weaker than Proposition 2.1. The stochastic version of this proposition is inspired from the forward and backward martingales of T. Lyons and W. Zheng [L-Z] (see [Tak], [Fa]).

Proof. The left hand side of the inequality of Proposition 2.3 may be rewritten as

$$G(t) = \int \mathrm{e}^{-\lambda f} P_t\bigl(\mathrm{e}^{\lambda f}\bigr) d\gamma_n.$$

Let λ be fixed and f be smooth enough. For every $t \ge 0$,

$$\begin{aligned} G(t) = 1 + \int_0^t G'(s)\,ds &= 1 + \int_0^t \biggl(\int \mathrm{e}^{-\lambda f} LP_s\bigl(\mathrm{e}^{\lambda f}\bigr)d\gamma_n\biggr)ds \\ &= 1 + \lambda^2 \int_0^t \mathrm{e}^{-s}\biggl(\int \mathrm{e}^{-\lambda f}\langle \nabla f, P_s\bigl(\mathrm{e}^{\lambda f}\nabla f\bigr)\rangle d\gamma_n\biggr) \\ &\le 1 + \lambda^2 \int_0^t \mathrm{e}^{-s} G(s)ds \end{aligned}$$

since $|\nabla f| \le 1$ almost everywhere. Let $H(t) = \log(1 + \lambda^2 \int_0^t \mathrm{e}^{-s}G(s)ds)$, $t \ge 0$. We just showed that $H'(t) \le \lambda^2 \mathrm{e}^{-t}$ for every $t \ge 0$. Hence,

$$H(t) = \int_0^t H'(s)\,ds \le \lambda^2 \int_0^t \mathrm{e}^{-s}\,ds = \lambda^2(1-\mathrm{e}^{-t})$$

and the proof is complete. □

If A and B are subsets of $\mathbb{R}^n$, and if $t \ge 0$, set

$$K_t(A,B) = \int_A P_t(I_B)\,d\gamma_n \quad \biggl(= \int_B P_t(I_A)\,d\gamma_n\biggr)$$

where I_A is the indicator function of the set A. Assume that $d(A,B) > r > 0$ (for the Euclidean distance on $\mathbb{R}^n$). In particular, $B \subset (A_r)^c$ so that

$$K_t(A,B) \le K_t\bigl(A,(A_r)^c\bigr).$$

Apply then Proposition 2.3 to the Lipschitz map $f(x) = d(x,A)$. For every $t \ge 0$ and every $\lambda \ge 0$,

$$K_t\bigl(A,(A_r)^c\bigr) = \int_A P_t\bigl(I_{(A_r)^c}\bigr)d\gamma_n \le \mathrm{e}^{-\lambda r}\int_A \mathrm{e}^{-\lambda f} P_t\bigl(\mathrm{e}^{\lambda f}\bigr)\,d\gamma_n$$

since $I_{(A_r)^c} \leq \mathrm{e}^{-\lambda r}\mathrm{e}^{\lambda f}$. Hence

$$K_t\big(A, (A_r)^c\big) \leq \mathrm{e}^{-\lambda r}\mathrm{e}^{\lambda^2(1-\mathrm{e}^{-t})}.$$

Optimizing in λ yields

$$K_t(A, B) \leq K_t\big(A, (A_r)^c\big) \leq \exp\left(-\frac{r^2}{4(1-\mathrm{e}^{-t})}\right). \tag{2.16}$$

Formula (2.16) will thus be used in Chapter 4 in applications to large deviations for the Ornstein-Uhlenbeck process.

3. ISOPERIMETRIC AND CONCENTRATION INEQUALITIES FOR PRODUCT MEASURES

In this chapter, we present several isoperimetric and concentration inequalities for product measures due to M. Talagrand. On the basis of the product structure of the canonical Gaussian measure γ_n and various open problems on sums of independent vector valued random variables, M. Talagrand developed in the past years new inequalities in products of probability spaces by defining several different notions of isoperimetric enlargement in this setting. These results appear as a striking illustration of the power of abstract concentration ideas which can be developed far beyond the framework of the classical geometrical isoperimetric inequalities. One of the main applications of his powerful techniques and results concerns tail behaviors and limit properties of sums of independent Banach space valued random variables. It partly motivated the writing of the book [L-T2] and we thus refer the interested reader to this reference for this kind of applications. New applications concern geometric probabilities, percolation, statistical mechanics... We will concentrate here on some of the theoretical inequalities and their relations with the Gaussian isoperimetric inequality, as well as on some recent and new aspects of the work of M. Talagrand [Ta16]. We actually refer to [Ta16] for complete proofs and details of some of the main results we present here. The reader that is interested first in the applications of the Gaussian isoperimetric and concentration inequalities may skip this chapter and come back to it after Chapter 7.

One first example studied by M. Talagrand is uniform measure on $\{0,1\}^{\mathbb{N}}$. For this example, he established a concentration inequality independent of the dimension [Ta3]. More importantly, he developed a new powerful scheme of proof based on induction on the number of coordinates. This technique allowed him to investigate isoperimetric and concentration inequalities in abstract product spaces.

Let (Ω, Σ, μ) be a (fixed but arbitrary) probability space and let P be the product measure $\mu^{\otimes n}$ on Ω^n. A point x in Ω^n has coordinates $x = (x_1, \ldots, x_n)$. (In the results which we present, one should notice that one does not increase the generality with arbitrary products spaces $\big(\Pi_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mu_i\big)$. Since the crucial inequalities will not depend on n, we need simply towork on products of $(\widetilde{\Omega}, \widetilde{\mu}) = \big(\Pi_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mu_i\big)$

with itself and consider the coordinate map

$$\widetilde{x} = (\widetilde{x}_1, \ldots, \widetilde{x}_n) \in \widetilde{\Omega}^n \to (\widetilde{x}_i)_i \in \Omega_i, \quad \widetilde{x}_i = ((\widetilde{x}_i)_1, \ldots, (\widetilde{x}_i)_n) \in \widetilde{\Omega},$$

that only depends on the i-th factor.)

The Hamming distance on Ω^n is defined by

$$d(x, y) = \mathrm{Card}\{1 \le i \le n;\, x_i \ne y_i\}.$$

The concentration function α of Ω^n for d satisfies, for every product probability P,

$$\alpha(r) \le C \exp\left(-\frac{r^2}{Cn}\right), \quad r \ge 0, \tag{3.1}$$

where $C > 0$ is numerical. In particular, if $P(A) \ge \frac{1}{2}$, for most of the elements x in Ω^n, there exists $y \in A$ within distance $\sqrt{n}$ of x. On the two point space, this may be shown to follow from the corresponding isoperimetric inequality [Har], [W-W]. A proof using the martingale inequality (2.4) is given in [Mi-S] (see also [MD] for a version with a better constant). As we will see later on, one can actually give an elementary proof of this result by induction on n. It might be important for the sequel to briefly indicate the procedure. If A is a subset of Ω^n and $x \in \Omega^n$, denote by $\varphi_A^1(x)$ the Hamming distance from x to A thus defined by

$$\varphi_A^1(x) = \inf\bigl\{k \ge 0;\, \exists\, y \in A,\, \mathrm{Card}\{1 \le i \le n;\, x_i \ne y_i\} \le k\bigr\}.$$

(Although this is nothing more than $d(x, A)$, this notation will be of better use in the subsequent developments.) M. Talagrand's approach [Ta3], [Ta16] then consists in showing that, for every $\lambda > 0$ and every product probability P,

$$\int \mathrm{e}^{\lambda \varphi_A^1}\, dP \le \frac{1}{P(A)}\, \mathrm{e}^{n\lambda^2/4}. \tag{3.2}$$

In particular, by Chebyshev's inequality, for every integer k,

$$P(\varphi_A^1 \ge k) \le \frac{1}{P(A)}\, \mathrm{e}^{-k^2/n},$$

that is the concentration (3.1). The same proof actually applies to all the Hamming metrics

$$d_a(x, y) = \sum_{i=1}^n a_i I_{\{x_i \ne y_i\}}, \quad a = (a_1, \ldots, a_n) \in \mathbb{R}_+^n,$$

with $|a|^2 = \sum_{i=1}^n a_i^2$ instead of n in the right hand side of (3.2). One can improve this result by studying functions of the probability of A in (3.2) such as $P(A)^{-\gamma}$. Optimizing in $\gamma > 0$, it is then shown in [Ta16] that for $k \ge \bigl(\frac{n}{2} \log \frac{1}{P(A)}\bigr)^{1/2}$,

$$P(\varphi_A^1 \ge k) \le \exp\left(-\frac{2}{n}\left[k - \left(\frac{n}{2} \log \frac{1}{P(A)}\right)^{1/2}\right]^2\right),$$

which is close to the best exponent $-2k^2/n$ ([MD]).

Note that various measurability questions arise on φ_A^1. These are actually completely unessential and will be ignored in what follows (start for example with a finite probability space Ω).

The previous definition of φ_A^1 allows one to investigate various and very different ways to measure the "distance" of a point x to a set A. In particular, this need not anymore be metric. The functional φ_A^1 controls a point x in Ω^n by a single point in A. Besides this function, M. Talagrand defines two new main controls, or enlargement functions: one using several points in A, and one using a convex hull procedure. In each case, a Gaussian concentration will be proved.

The convex hull control is defined with the metric $\varphi_A^c(x) = \sup_{|a|=1} d_a(x, A)$. However, this definition somewhat hides the convexity properties of the functional φ_A^c which will be needed in its investigation. For a subset $A \subset \Omega^n$ and $x \in \Omega^n$, let

$$U_A(x) = \big\{ s = (s_i)_{1 \le i \le n} \in \{0,1\}^n ;\, \exists\, y \in A \text{ such that } y_i = x_i \text{ if } s_i = 0 \big\}.$$

(One can use instead the collection of the indicator functions $I_{\{x_i \neq y_i\}}$, $y \in A$.) Denote by $V_A(x)$ the convex hull of $U_A(x)$ as a subset of $\mathbb{R}^n$. Note that $0 \in V_A(x)$ if and only if $x \in A$. One may then measure the distance from x to A by the Euclidean distance $d(0, V_A(x))$ from 0 to $V_A(x)$. It is easily seen that $d(0, V_A(x)) = \varphi_A^c(x)$. If $d(0, V_A(x)) \le r$, there exists z in $V_A(x)$ with $|z| \le r$. Let $a \in \mathbb{R}_+^n$ with $|a| = 1$. Then

$$\inf_{y \in V_A(x)} \langle a, y \rangle \le \langle a, z \rangle \le |z| \le r.$$

Since

$$\inf_{y \in V_A(x)} \langle a, y \rangle = \inf_{s \in U_A(x)} \langle a, s \rangle = d_a(x, A),$$

$\varphi_A^c(x) \le r$. The converse, that is not needed below, follows from Hahn-Banach's theorem.

The functional $\varphi_A^c(x)$ is a kind of uniform control in the Hamming metrics d_a, $|a| = 1$. The next theorem [Ta6], [Ta16] extends the concentration (3.2) to this uniformity.

Theorem 3.1. *For every subset A of Ω^n, and every product probability P,*

$$\int \exp\Big(\frac{1}{4} \left(\varphi_A^c\right)^2 \Big) dP \le \frac{1}{P(A)} \,.$$

In particular, for every $r \ge 0$,

$$P(\varphi_A^c \ge r) \le \frac{1}{P(A)}\, \mathrm{e}^{-r^2/4}.$$

To briefly describe the general scheme of proofs by induction on the number of coordinates, we present the proof of Theorem 3.1. The main difficulty in this type of statements is to find the adapted recurrence hypothesis expressed here by the exponential integral inequalities.

Proof. The case $n = 1$ is easy. To go from n to $n+1$, let A be a subset of Ω^{n+1} and let B be its projection on Ω^n. Let furthermore, for $\omega \in \Omega$, $A(\omega)$ be the section of A along ω. If $x \in \Omega^n$ and $\omega \in \Omega$, set $z = (x, \omega)$. The key observation is the following: if $s \in U_{A(\omega)}(x)$, then $(s, 0) \in U_A(z)$, and if $t \in U_B(x)$, then $(t, 1) \in U_A(z)$. It follows that if $u \in V_{A(\omega)}(x)$, $v \in V_B(x)$ and $0 \le \lambda \le 1$, then $(\lambda u + (1-\lambda)v, 1-\lambda) \in V_A(z)$. By definition of φ_A^c and convexity of the square function,

$$\varphi_A^c(z)^2 \le (1-\lambda)^2 + \big|\lambda u + (1-\lambda)v\big|^2 \le (1-\lambda)^2 + \lambda|u|^2 + (1-\lambda)|v|^2.$$

Hence,

$$\varphi_A^c(z)^2 \le (1-\lambda)^2 + \lambda\varphi_{A(\omega)}^c(x)^2 + (1-\lambda)\varphi_B^c(x)^2.$$

Now, by Hölder's inequality and the induction hypothesis, for every ω in Ω,

$$\begin{aligned}
&\int_{\Omega^n} \exp\Big(\frac{1}{4}\,\big(\varphi_A^c(x,\omega)\big)^2\Big)dP(x) \\
&\qquad \le \mathrm{e}^{(1-\lambda)^2/4}\Big(\int_{\Omega^n} \exp\Big(\frac{1}{4}\,\big(\varphi_{A(\omega)}^c\big)^2\Big)dP\Big)^{\lambda}\Big(\int_{\Omega^n} \exp\Big(\frac{1}{4}\,\big(\varphi_B^c\big)^2\Big)dP\Big)^{1-\lambda} \\
&\qquad \le \mathrm{e}^{(1-\lambda)^2/4}\Big(\frac{1}{P(A(\omega))}\Big)^{\lambda}\Big(\frac{1}{P(B)}\Big)^{1-\lambda}
\end{aligned}$$

that is,

$$\int_{\Omega^n} \exp\Big(\frac{1}{4}\,\big(\varphi_A^c(x,\omega)\big)^2\Big)dP(x) \le \frac{1}{P(B)}\,\mathrm{e}^{(1-\lambda)^2/4}\Big(\frac{P(A(\omega))}{P(B)}\Big)^{-\lambda}.$$

Optimize now in λ (cf. [Ta16]) to get that

$$\int_{\Omega^n} \exp\Big(\frac{1}{4}\,\big(\varphi_A^c(x,\omega)\big)^2\Big)dP(x) \le \frac{1}{P(B)}\Big(2 - \frac{P(A(\omega))}{P(B)}\Big).$$

To conclude, integrate in ω, and, by Fubini's theorem,

$$\int_{\Omega^{n+1}} \exp\Big(\frac{1}{4}\,\big(\varphi_A^c(x,\omega)\big)^2\Big)dP(x)d\mu(\omega) \le \frac{1}{P(B)}\Big(2 - \frac{P\otimes\mu(A)}{P(B)}\Big) \le \frac{1}{P\otimes\mu(A)}$$

since $u(2-u) \le 1$ for every real number u. Theorem 3.1 is established. □

It is easy to check that if $\Omega = [0,1]$ and if d_A is the Euclidean distance to the convex hull $\mathrm{Conv}(A)$ of A, then $d_A \le \varphi_A^c$. Let then f be a convex function on $[0,1]^n$ such that $\|f\|_{\mathrm{Lip}} \le 1$, and let m be a median of f for P and $A = \{f \le m\}$. Since f is convex, $f \le m$ on $\mathrm{Conv}(A)$. Using that $\|f\|_{\mathrm{Lip}} \le 1$, we see that $f(x) < m + r$ for every x such that $d_A(x) < r$, $r \ge 0$. Hence, by Theorem 3.1, $P(f \ge m+r) \le 2\mathrm{e}^{-r^2/4}$.

On the other hand, let $B = \{f \leq m - r\}$. As above, $d_B(x) < r$ implies $f(x) < m$. By definition of the median, it follows from Theorem 3.1 again that

$$1 - \frac{1}{P(B)}\,\mathrm{e}^{-r^2/4} \leq P(d_B < r) \leq P(f < m) \leq \frac{1}{2}\,.$$

Hence $P(f \leq m - r) \leq 2\mathrm{e}^{-r^2/4}$. Therefore

$$P\big(|f - m| \geq r\big) \leq 4\,\mathrm{e}^{-r^2/4} \tag{3.3}$$

for every $r \geq 0$ and every probability measure μ on $[0,1]$. The numerical constant 4 in the exponent may be improved to get close to the best possible value 2. This concentration inequality (3.3) is very similar to the Gaussian concentration (2.4) or (2.9), with however f convex. It applies to norms of vector valued sums $\sum_i \varphi_i e_i$ with coefficients e_i in a Banach space E where the φ_i's are independent real valued uniformly bounded random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. This applies in particular to independent symmetric Bernoulli (or Rademacher) random variables and (3.3) allows us in particular to recover and improve the pioneer inequalities by J.-P. Kahane [Ka1], [Ka2] (cf. also Chapter 4). More precisely, if $\|\varphi_i\|_\infty \leq 1$ for all i's and if $S = \sum_i \varphi_i e_i$ is almost surely convergent in E, for every $r \geq 0$,

$$\mathbb{P}\big(\big|\|S\| - m\big| \geq r\big) \leq 4\,\mathrm{e}^{-r^2/16\sigma^2} \tag{3.4}$$

where m is a median of $\|S\|$ and where

$$\sigma = \sup_{\xi \in E^*, \|\xi\| \leq 1} \bigg(\sum_i \langle \xi, e_i\rangle^2\bigg)^{1/2}.$$

Typically, the martingale inequality (2.5) would only yield a similar inequality but with σ replaced by the larger quantity $\sum_i \|e_i\|^2$. This result is the exact analogue of what we will obtain on Gaussian series in the next chapter through isoperimetric and concentration inequalities. It shows how powerful the preceding induction techniques can be. In particular, we may integrate by parts (3.4) to see that $\mathbb{E}\exp(\alpha\|S\|^2) < \infty$ for every α (cf. [L-T2]). Furthermore, for every $p > 0$,

$$\big(\mathbb{E}\|S\|^p\big)^{1/p} \leq m + C_p\sigma \tag{3.5}$$

where C_p is of the order of $\sqrt{p}$ as $p \to \infty$.

When the φ_i's are Rademacher random variables, by the classical Khintchine inequalities, one easily sees that $\sigma \leq 2\sqrt{2}m'$ for every m' such that $\mathbb{P}\{\|S\| \geq m'\} \leq \frac{1}{8}$ (see [L-T2], p. 99). Since $m \leq m' \leq (8\mathbb{E}\|S\|^q)^{1/q}$ for every $0 < q < \infty$, we also deduce from (3.5) the moment equivalences for $\|S\|$: for every $0 < p, q < \infty$, there exists $C_{p,q} > 0$ only depending on p and q such that

$$\big(\mathbb{E}\|S\|^p\big)^{1/p} \leq C_{p,q}\big(\mathbb{E}\|S\|^q\big)^{1/q}. \tag{3.6}$$

By the classical central limit theorem, these inequalities imply the corresponding ones for Gaussian averages (see (4.5)). In the case of the two point space, the method of proof by induction on the dimension is very similar to hypercontractivity techniques [Bon], [Gr3], [Be1] (which, in particular, also show (3.6) [Bo6]). Some further connections on the basis of this observation are developed in [Ta11] in analogy with the Gaussian example (Chapter 8). See also [Bo8]. It was recently shown in [L-O] that $C_{2,1} = \sqrt{2}$ as in the real case [Sz].

We turn to the control by a finite number of points. If q is an integer ≥ 2 and if $A^1, \ldots, A^q$ are subsets of Ω^n, let, for every $x = (x_1, \ldots, x_n)$ in Ω^n,

$$\varphi^q(x) = \varphi^q_{A^1,\ldots,A^q}(x) = \inf\bigl\{k \geq 0;\, \exists\, y^1 \in A^1, \ldots, \exists\, y^q \in A^q \text{ such that } \operatorname{Card}\bigl\{1 \leq i \leq n;\, x_i \notin \{y_i^1, \ldots, y_i^q\}\bigr\} \leq k\bigr\}.$$

(We agree that $\varphi^q = \infty$ if one of the A_i's is empty.) If, for every $i = 1, \ldots, n$, $A^i = A$ for some $A \subset \Omega^n$, $\varphi^q(x) \leq k$ means that the coordinates of x may be copied, at the exception of k of them, by the coordinates of q elements in A. Using again a proof by induction on the number of coordinates, M. Talagrand [Ta16] established the following result.

Theorem 3.2. *Under the previous notations,*

$$\int q^{\varphi^q(x)}\, dP(x) \leq \prod_{i=1}^q \frac{1}{P(A^i)}\,.$$

In particular, for every integer k,

$$P(\varphi^q \geq k) \leq q^{-k} \prod_{i=1}^q \frac{1}{P(A^i)}\,.$$

Proof. One first observes that if g is a function on Ω such that $\frac{1}{q} \leq g \leq 1$, then

$$\int \frac{1}{g}\, d\mu \left(\int g d\mu\right)^q \leq 1. \tag{3.7}$$

Since $\log u \leq u - 1$, it suffices to show that

$$\int \frac{1}{g}\, d\mu + q\int g d\mu = \int \Big(\frac{1}{g} + qg\Big) d\mu \leq q + 1.$$

But this is obvious since $\frac{1}{u} + qu \leq q + 1$ for $\frac{1}{q} \leq u \leq 1$.

Let g_i, $i = 1, \ldots, q$, be functions on Ω such that $0 \leq g_i \leq 1$. Applying (3.7) to g given by $\frac{1}{g} = \min(q, \min_{1 \leq i \leq q} \frac{1}{g_i})$ yields

$$\int \min\Big(q, \min_{1 \leq i \leq q} \frac{1}{g_i}\Big) d\mu \left(\prod_{i=1}^q \int g_i d\mu\right) \leq 1 \tag{3.8}$$

since $g_i \le g$ for every $i = 1, \ldots, q$.

We prove the theorem by induction over n. If $n = 1$, the result follows from (3.8) by taking $g_i = I_{A^i}$. Assume Theorem 3.2 has been proved for n and let us prove it for $n+1$. Consider sets $A^1, \ldots, A^q$ of Ω^{n+1}. For $\omega \in \Omega$, consider $A^i(\omega)$, $i = 1, \ldots, q$, as well as the projections B^i of A^i on Ω^n, $i = 1, \ldots, q$. Note that if we set $g_i = P(A^i(\omega))/P(B^i)$ in (3.8), we get by Fubini's theorem that

$$\int \min\biggl(q \prod_{i=1}^q \frac{1}{P(B^i)}, \min_{1 \le j \le q} \prod_{i=1}^q \frac{1}{P(C^{ij})}\biggr) d\mu \le \prod_{i=1}^q \frac{1}{P \otimes \mu(A^i)} \tag{3.9}$$

where $C^{ij} = B^i$ if $i \ne j$ and $C^{ii} = A^i(\omega)$. The basic observation is now the following: for $(x, \omega) \in \Omega^n \times \Omega$,

$$\varphi^q_{A^1,\ldots,A^q}(x,\omega) \le 1 + \varphi^q_{B^1,\ldots,B^q}(x)$$

and, for every $1 \le j \le q$,

$$\varphi^q_{A^1,\ldots,A^q}(x,\omega) \le \varphi^q_{C^{1j},\ldots,C^{qj}}(x).$$

It follows that

$$\begin{aligned}
&\int_{\Omega^{n+1}} q^{\varphi^q_{A^1,\ldots,A^q}(x,\omega)} dP(x) d\mu(\omega) \\
&\quad\le \int_{\Omega^{n+1}} \min\bigl(q\, q^{\varphi^q_{B^1,\ldots,B^q}(x)}, \min_{1\le j\le q} q^{\varphi^q_{C^{1j},\ldots,C^{qj}}(x)}\bigr) dP(x) d\mu(\omega) \\
&\quad\le \int_{\Omega} \min\biggl(q \int_{\Omega^n} q^{\varphi^q_{B^1,\ldots,B^q}(x)} dP(x), \min_{1\le j\le q} \int_{\Omega^n} q^{\varphi^q_{C^{1j},\ldots,C^{qj}}(x)} dP(x)\biggr) d\mu(\omega) \\
&\quad\le \int_{\Omega} \min\biggl(q \prod_{i=1}^q \frac{1}{P(B^i)}, \min_{1\le j\le q} \prod_{i=1}^q \frac{1}{P(C^{ij})}\biggr) d\mu(\omega)
\end{aligned}$$

by the recurrence hypothesis. The conclusion follows from (3.9). □

In the applications, q is usually fixed, for example equal to 2. Theorem 3.2 then shows how to control, with a fixed subset A, arbitrary samples with an exponential decay of the probability in the number of coordinates which are neglected. Theorem 3.2 was first proved by delicate rearrangement techniques (of isoperimetric flavor) in [Ta5]. It allowed M. Talagrand to solve a number of open questions in probability in Banach spaces (and may be considered at the origin of the subsequent abstract developments, see [Ta5], [Ta16], [L-T2]). To briefly illustrate how Theorem 3.2 is used in the applications, let us consider a sum $S = X_1 + \cdots + X_n$ of independent nonnegative random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In the preceding language, we may simply equip $[0,\infty)^n$ with the product P of the laws of the X_i's. Let $A = \{\sum_{i=1}^n x_i \le m\}$ where m is such that, for example, $P(A) \ge \frac{1}{2}$. Let $\varphi^q = \varphi^q_{A,\ldots,A}$. If $x \in \{\varphi^q \le k\}$, there exist $y^1, \ldots, y^q$ in A such that $\mathrm{Card}\, I \le k$ where $I = \{1 \le i \le n;\ x_i \notin \{y^1_i, \ldots, y^q_i\}\}$. Take then a partition $(J_j)_{1\le j\le q}$ of $\{1, \ldots, n\} \setminus I$ such that $x_i = y^j_i$ if $i \in J_j$. Then,

$$\sum_{i \notin I} x_i = \sum_{j=1}^q \sum_{i \in J_j} y^j_i \le \sum_{j=1}^q \sum_{i=1}^n y^j_i \le qm$$

where we are using a crucial monotonicity property since the coordinates are nonnegative. It follows that

$$\sum_{i=1}^{n} x_i \le qm + \sum_{i=1}^{k} x_i^*$$

where $\{x_1^*, \dots, x_n^*\}$ is the nonincreasing rearrangement of the sample $\{x_1, \dots, x_n\}$. Hence, according to Theorem 3.2, for every integers $k, q \ge 1$, and every $t \ge 0$,

$$\mathbb{P}\{S \ge qm + t\} \le 2^q q^{-(k+1)} + \mathbb{P}\left\{\sum_{i=1}^{k} X_i^* \ge t\right\}. \tag{3.10}$$

Let $\mathcal{F}$ be a family of n-tuples $\alpha = (\alpha_i)_{1 \le i \le n}$, $\alpha_i \ge 0$. It is plain that the preceding argument leading to (3.10) applies in the same way to

$$S = \sup_{\alpha \in \mathcal{F}} \sum_{i=1}^{n} \alpha_i X_i$$

to yield

$$\mathbb{P}\{S \ge qm + t\} \le 2^q q^{-(k+1)} + \mathbb{P}\left\{\sigma \sum_{i=1}^{k} X_i^* \ge t\right\}$$

where $\sigma = \sup\{\alpha_i; 1 \le i \le n, \alpha \in \mathcal{F}\}$.

Now, in probability in Banach spaces or in the study of empirical processes, one does not usually deal with nonnegative summands. One general situation is the following (cf. [L-T2], [Ta13] for the notations and further details). Let $X_1, \dots, X_n$ be independent random variables taking values in some space $\mathcal{S}$ and consider say a countable family $\mathcal{F}$ of (measurable) real valued functions on $\mathcal{S}$. We are interested in bounds on the tail of

$$\left\| \sum_{i=1}^{n} f(X_i) \right\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} f(X_i) \right|.$$

If $\mathbb{E} f(X_i) = 0$ for every $1 \le i \le n$ and every $f \in \mathcal{F}$, standard symmetrization techniques (cf. [L-T2]) reduce to the investigation of

$$\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{F}}$$

where $(\varepsilon_i)_{1 \le i \le n}$ are independent symmetric Bernoulli random variables independent of the X_i's. Although the isoperimetric approach applies in the same way, we may not use directly here the crucial monotonicity property on the coordinates. We turn over this difficulty via a symmetrization procedure with Rademacher random variables which was developed first in the study of the law of the iterated logarithm [L-T1]. One writes

$$\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{F}} = \left(\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{F}} - \mathbb{E}_\varepsilon \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{F}} \right) + \mathbb{E}_\varepsilon \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{F}}$$

where $\mathbb{E}_\varepsilon$ is partial integration with respect to the Bernoulli variables $\varepsilon_1, \ldots, \varepsilon_n$. Now, on $\mathbb{E}_\varepsilon \| \sum_{i=1}^n \varepsilon_i f(X_i) \|_{\mathcal{F}}$ the monotonicity property is satisfied, since, by Jensen's inequality and independence, for every subset $I \subset \{1, \ldots, n\}$,

$$\mathbb{E}_\varepsilon \Big\| \sum_{i \in I} \varepsilon_i f(X_i) \Big\|_{\mathcal{F}} \leq \mathbb{E}_\varepsilon \Big\| \sum_{i=1}^n \varepsilon_i f(X_i) \Big\|_{\mathcal{F}}.$$

Therefore, the isoperimetric method may be used efficiently on this term. The remainder term

$$\Big\| \sum_{i=1}^n \varepsilon_i f(X_i) \Big\|_{\mathcal{F}} - \mathbb{E}_\varepsilon \Big\| \sum_{i=1}^n \varepsilon_i f(X_i) \Big\|_{\mathcal{F}}$$

is bounded, conditionally on the X_i's, with the deviation inequality (3.4) by a Gaussian tail involving

$$\sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i)^2$$

which will again satisfy this monotonicity property. The proper details are presented in [L-T2], p. 166-169. Combining the arguments yields the inequality, for nonnegative integers k, q and real numbers $s, t \geq 0$,

$$\mathbb{P}\bigg\{ \Big\| \sum_{i=1}^n \varepsilon_i f(X_i) \Big\|_{\mathcal{F}} \geq 8qM + s + t \bigg\} \leq 2^q q^{-k} + \mathbb{P}\bigg\{ \sum_{i=1}^k \|f(X_i)\|_{\mathcal{F}}^* \geq s \bigg\} + 2 \exp\Big(-\frac{t^2}{128qm^2} \Big)$$

where

$$M = \mathbb{E} \Big\| \sum_{i=1}^n \varepsilon_i f(X_i) \Big\|_{\mathcal{F}}, \quad m = \mathbb{E}\bigg(\sup_{f \in \mathcal{F}} \Big(\sum_{i=1}^n f(X_i)^2 \Big)^{1/2} \bigg)$$

and where $(\|f(X_i)\|_{\mathcal{F}}^*)_{1 \leq i \leq n}$ denotes the nonincreasing rearrangement of the sample $(\|f(X_i)\|_{\mathcal{F}})_{1 \leq i \leq n}$. (Of course, if the functions f of $\mathcal{F}$ are such that $|f| \leq 1$, one may choose for example $s = k$.) We find again in this type of inequalities the basic parameters of concentration inequalities of Gaussian type.

This approach to bounds on sums of independent Banach space valued random variables (or empirical processes) is today one of the main successful tools in the study of integrability and limit properties of these sums. The results which may be obtained with this isoperimetric technique are rather sharp and often improve even the scalar case. The range of applications appears to be much broader than what can be obtained for example from the martingale inequalities (2.5). We refer to the monograph [L-T2] for a complete exposition of these applications in the context of probability in Banach spaces.

In his recent developments, M. Talagrand further analyzes the control functionals φ^1, φ^q and φ^c and extends their potential use and interest by a new concept of penalty. Indeed, in the functional φ^1 for example, the coordinates of x which differ from the coordinates of a point in A are accounted for one. One may therefore imagine a more precise measure of this control with some adapted weight. Let, for a nonnegative function h on $\Omega \times \Omega$ such that $h(\omega, \omega) = 0$, and for $A \subset \Omega^n$, $x \in \Omega^n$,

$$\varphi_A^{1,h}(x) = \inf\Big\{\sum_{i=1}^n h(x_i, y_i);\ y \in A\Big\}.$$

When $h(\omega, \omega') = 1$ if $\omega \neq \omega'$, we simply recover the Hamming metric φ_A^1. The new functional $\varphi_A^{1,h}$ thus puts a variable penalty $h(x, y)$ on the coordinates of x and y which differ.

Provided with these functionals, one may therefore take again the preceding study and obtain, by the same method of proof based on induction on the number of coordinates, several new and important concentration inequalities. The first result resembles Bernstein's classical exponential bound. Denote by $\|h\|_2$ and $\|h\|_\infty$ respectively the L^2 and L^∞-norms of h with respect to $\mu \otimes \mu$.

Theorem 3.3. *For each subset A of Ω^n and every product probability P, and every $r \geq 0$,*

$$P\big(\varphi_A^{1,h} \geq r\big) \leq \frac{1}{P(A)} \exp\bigg(-\min\bigg(\frac{r^2}{8n\|h\|_2^2}, \frac{r}{2\|h\|_\infty}\bigg)\bigg).$$

To better analyze the conditions on the penalty function h, set, for $B \subset \Omega$ and $\omega \in \Omega$,

$$h(\omega, B) = \inf\big\{h(\omega, \omega');\ \omega' \in B\big\}.$$

Assume that for all $B \subset \Omega$,

$$\int \mathrm{e}^{2h(\omega,B)}\, d\mu(\omega) \leq \frac{\mathrm{e}}{\mu(B)}\,.$$

A typical statement of [Ta16] is then that, for every $0 \leq \lambda \leq 1$,

$$\int \mathrm{e}^{\lambda \varphi_A^{1,h}}\, dP \leq \frac{1}{P(A)}\, \mathrm{e}^{Cn\lambda^2} \tag{3.11}$$

where $C > 0$ is a numerical constant. With respect to (3.2), we easily see how successful (3.11) can be for an appropriate choice of the penalty function h. One may also prove extensions where, as we already mentioned it, the probability of A is replaced by more complicated functions of this probability (related of course to h.) The penalty or interacting functions h which are used in such a result are of various types. For example, on $\mathbb{R}$, one may take $h(\omega, \omega') = |\omega - \omega'|$ or $h(\omega, \omega') = (\omega - \omega')^+$. One of the striking observations by M. Talagrand is the dissymmetric behavior of the two variables of h, that is on the point x that we would like to control and the point y in the fixed set A. For example, if h only depends on the first coordinate, then it should be bounded; if it only depends on the second coordinate, only weak integrability properties (with respect to μ) are required.

These extensions can also be performed on the functionals φ^q and φ^c, the latter being probably the most interesting for the applications. For a nonnegative penalty function h as before, let, for $A \subset \Omega^n$ and $x \in \Omega$,

$$U_A(x) = \big\{ s = (s_i)_{1 \le i \le n} \in \mathbb{R}^n_+ ; \exists\, y \in A \text{ such that } s_i \ge h(x_i, y_i) \text{ for every } i = 1, \ldots, n \big\}.$$

Denote by $V_A(x)$ the convex hull of $U_A(x)$. To measure the “distance” from 0 to $V_A(x)$, let us consider a function ψ on $\mathbb{R}$ with $\psi(0) = 0$ and such that $\psi(t) \le t^2$ if $t \le 1$ and $\psi(t) \ge t$ if $t \ge 1$. Then, let

$$\varphi_A^{c,h,\psi}(x) = \inf \bigg\{ \sum_{i=1}^n \psi(s_i);\ s = (s_i)_{1 \le i \le n} \in V_A(x) \bigg\}.$$

The metric φ^c thus simply corresponds to $h(\omega, \omega') = 1$ if $\omega \ne \omega'$ and $\psi(t) = t^2$. Again by induction on the dimension, M. Talagrand [Ta16] then establishes a general form of Theorem 3.2. He shows that, for some constant $\alpha > 0$,

$$\int \exp\big(\alpha\, \varphi_A^{c,h,\psi}\big) \le \exp\big(\theta\big(P(A)\big)\big)$$

under various conditions connecting μ, h and ψ to the function θ of the probability of A. The proof is of course more involved due to the level of generality.

This abstract study of isoperimetry and concentration in product spaces is motivated by the large number of applications, both in theoretical and more applied probabilistic topics proposed today by M. Talagrand [Ta16]. Most often, the preceding inequalities allow one to establish a concentration inequality once an appropriate mean or median is known. To briefly present such an example of application, let us deal with first passage time in percolation theory. Let $G = (V, \mathcal{E})$ be a graph with vertices V and edges $\mathcal{E}$. Let, on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $(X_e)_{e \in \mathcal{E}}$ be a family of nonnegative independent and identically distributed random variables with the same distribution as X. X_e represents the passage time through the edge e. Let $\mathcal{T}$ be a family of (finite) subsets of $\mathcal{E}$, and, for $T \in \mathcal{T}$, set $X_T = \sum_{e \in T} X_e$. If T is made of contiguous edges, X_T represents the passage time through the path T. Set $Z_{\mathcal{T}} = \inf_{T \in \mathcal{T}} X_T$ and $D = \sup_{T \in \mathcal{T}} \mathrm{Card}(T)$, and let m be a median of $Z_{\mathcal{T}}$. As a corollary of his penalty theorems, M. Talagrand [Ta16] proved the following result.

Theorem 3.4. *There exists a numerical constant $c > 0$ such that, if $\mathbb{E}(\mathrm{e}^{cX}) \le 2$, for every $r \ge 0$,*

$$\mathbb{P}\big(|Z_{\mathcal{T}} - m| \ge r\big) \le \exp\bigg(-c \min\Big(\frac{r^2}{D}, r\Big)\bigg).$$

When V is $\mathbb{Z}^2$ and $\mathcal{E}$ the edges connecting two adjacent points, and when $\mathcal{T} = \mathcal{T}_n$ is the set of all selfavoiding paths connecting the origin to the point $(0, n)$,

H. Kesten [Ke] showed that, when $0 \le X \le 1$ almost surely and $\mathbb{P}(X=0) < \frac{1}{2}$ (percolation), one may reduce, in $Z_{\mathcal{T}}$, to paths with length less than some multiple of n. Together with this result, Theorem 3.4 indicates that

$$\mathbb{P}\big(|Z_{\mathcal{T}_n} - m| \ge r\big) \le 5\exp\Big(-\frac{r^2}{Cn}\Big)$$

for every $r \le n/C$ where $C > 0$ is a constant independent of n. This result strengthens the previous estimate by H. Kesten [Ke] which was of the order of $r/C\sqrt{n}$ in the exponent and the proof of which was based on martingale inequalities.

Let us mention to conclude a further application of these methods to spin glasses. Consider a sequence $(\varepsilon_i)_{i\in\mathbb{N}}$ of independent symmetric random variables taking values ± 1. Each ε_i represents the spin of particule i. Consider then interactions H_{ij}, $i < j$, between spins. For some parameter $\beta > 0$ (that plays the role of the inverse of the temperature), the so-called partition function is defined by

$$Z_n = Z_n(\beta) = \mathbb{E}_\varepsilon\bigg(\exp\bigg(\frac{\beta}{\sqrt{n}}\sum_{1\le i<j\le n} H_{ij}\varepsilon_i\varepsilon_j\bigg)\bigg), \quad n \ge 2,$$

where $\mathbb{E}_\varepsilon$ is integration with respect to the $\varepsilon_i's$. In the model we study, the interactions H_{ij} are random and the H_{ij}'s will be assumed independent and identically distributed. We assume that, for every $i < j$,

$$\mathbb{E}(H_{ij}) = \mathbb{E}(H_{ij}^3) = 0, \quad \mathbb{E}(H_{ij}^2) = 1,$$

and

$$\mathbb{E}\big(\exp(\pm H_{ij})\big) \le 2$$

(for normalization purposes). The typical example is of course the example of a standard Gaussian sequence. In this case, it was shown in [A-L-R] and [C-N] that for $\beta < 1$, the sequence

$$\log Z_n - \frac{\beta^2 n}{4}, \quad n \ge 2,$$

converges in distribution to a (nonstandard) centered Gaussian variable. Of equal interest, but of rather different nature, is a concentration result of $\log Z_n$ around $\frac{\beta^2 n}{4}$ for n fixed, that M. Talagrand deduces from its penalty theorems [Ta16].

Theorem 3.5. *There is a numerical constant $C > 1$ such that for $0 \le r \le n/C$ and $\beta < 1$,*

$$\mathbb{P}\bigg\{\Big|\log Z_n - \frac{\beta^2 n}{4}\Big| \ge C\bigg(r + \Big(\log\frac{C}{1-\beta^2}\Big)^{1/2}\bigg)\sqrt{n}\bigg\} \le 4\mathrm{e}^{-r^2}.$$

In particular,

$$-\frac{C}{\sqrt{n}}\Big(\log\frac{C}{1-\beta^2}\Big)^{1/2} \le \frac{1}{n}\,\mathbb{E}(\log Z_n) - \frac{\beta^2}{4} \le \frac{C}{n}\,.$$

In case the interactions H_{ij}, $i < j$, are independent standard Gaussian, Theorem 3.5 immediately follows from the Gaussian concentration inequalities. Let indeed, on $\mathbb{R}^k$, $k = (n(n-1)/2$,

$$f(x) = \log \mathbb{E}_\varepsilon \biggl(\exp\biggl(\frac{\beta}{\sqrt{n}} \sum_{1 \le i < j \le n} x_{ij} \varepsilon_i \varepsilon_j \biggr) \biggr), \quad x = (x_{ij})_{1 \le i < j \le n} .$$

It is easily seen that $\|f\|_{\mathrm{Lip}} \le \beta \sqrt{(n-1)/2}$ so that, by (2.9), for every $r \ge 0$,

$$\mathbb{P}\bigl\{ \bigl| \log Z_n - \mathbb{E}(\log Z_n) \bigr| \ge r \bigr\} \le 2 \exp\biggl(- \frac{r^2}{\beta^2 (n-1)} \biggr). \tag{3.12}$$

Now $\mathbb{E}(\log Z_n) \le \log \mathbb{E}(Z_n) = \beta^2 (n-1)/4$. Conversely, it may easily be shown (cf. [Ta16]) that

$$\mathbb{E}(Z_n^2) = \bigl(\mathbb{E}(Z_n) \bigr)^2 \mathrm{e}^{-\beta^2/2} \mathbb{E}\biggl(\exp\biggl(\frac{\beta^2}{2n} \biggl(\sum_{i=1}^n \varepsilon_i \biggr)^2 \biggr) \biggr). \tag{3.13}$$

In particular (using the subgaussian inequality for sums of Rademacher random variables [L-T2], p. 90), if $\beta < 1$,

$$\mathbb{E}(Z_n^2) \le \frac{3}{1 - \beta^2} \bigl(\mathbb{E}(Z_n) \bigr)^2 .$$

Hence, by the Paley-Zygmund inequality ([L-T2], p. 92),

$$\mathbb{P}\Bigl\{ Z_n \ge \frac{1}{2} \, \mathbb{E}(Z_n) \Bigr\} \ge \frac{\bigl(\mathbb{E}(Z_n) \bigr)^2}{4 \mathbb{E}(Z_n^2)} \ge \frac{1 - \beta^2}{12} \, .$$

Assume first that $r = \log(\frac{1}{2} \mathbb{E}(Z_n)) - \mathbb{E}(\log Z_n) > 0$. Then, by (3.12) applied to this r,

$$\begin{aligned} \frac{1 - \beta^2}{12} &\le \mathbb{P}\Bigl\{ \log Z_n \ge \log\Bigl(\frac{1}{2} \, \mathbb{E}(Z_n) \Bigr) \Bigr\} \\ &\le \mathbb{P}\bigl\{ \log Z_n \ge \mathbb{E}(\log Z_n) + r \bigr\} \le 2 \exp\biggl(- \frac{r^2}{\beta^2 (n-1)} \biggr) \end{aligned}$$

so that

$$r \le \sqrt{n} \biggl(\log \frac{24}{1 - \beta^2} \biggr)^{1/2} .$$

Hence, in any case,

$$\begin{aligned} \frac{\beta^2 (n-1)}{4} \ge \mathbb{E}(\log Z_n) &\ge \log\Bigl(\frac{1}{2} \, \mathbb{E}(Z_n) \Bigr) - \sqrt{n} \biggl(\log \frac{24}{1 - \beta^2} \biggr)^{1/2} \\ &\ge \frac{\beta^2 (n-1)}{4} - 2 \sqrt{n} \biggl(\log \frac{24}{1 - \beta^2} \biggr)^{1/2} \end{aligned}$$

and the theorem follows in this case.

Note that, by (3.12) and the Borel-Cantelli lemma, for any $\beta > 0$,

$$\lim_{n\to\infty} \left| \frac{1}{n} \log Z_n - \frac{1}{n} \mathbb{E}(\log Z_n) \right| = 0$$

almost surely. In particular,

$$0 \leq \limsup_{n\to\infty} \frac{1}{n} \log Z_n \leq \frac{\beta^2}{4}$$

almost surely. This supports the conjecture that $\frac{1}{n} \log Z_n$ should converge in an appropriate sense for every $\beta > 0$ (cf. [A-L-R] and [Co] for precise bounds using different techniques).

As we have seen, the application to probability in Banach spaces is one main topic in which these isoperimetric and concentration inequalities for product measures prove all their strength and efficiency. Besides, M. Talagrand has thus shown how these tools may be used in a variety of problems (random subsequences, random graphs, percolation, geometric probability, spin glasses...). We refer the interested reader to his important contribution [Ta16].

Notes for further reading. As already mentioned, the interested reader may find in the book [L-T2] an extensive description of the application of the isoperimetric inequalities for product measures to probability in Banach spaces (integrability of the norm of sums of independent Banach space valued random variables, strong limit theorems such as laws of large numbers and laws of the iterated logarithm...). Sharper bounds for empirical processes using these methods, and based on Gaussian ideas, are obtained in [Ta13]. The recent paper [Ta16] produces new fields of potential interest for applications of these ideas. [Ta17] provides further sharpenings with approximations by very many points.

4. INTEGRABILITY AND LARGE DEVIATIONS OF GAUSSIAN MEASURES

In this chapter, we make use of the isoperimetric and concentration inequalities of Chapters 1 and 2 to study the integrability properties of functionals of a Gaussian measure as well as large deviation statements. In particular, we will only use in this study the concentration inequalities which were obtained by rather elementary arguments in Chapter 2 so that the results presented here actually proceed from a very simple scheme. We first establish the, by now classical, strong integrability theorems of norms of Gaussian measures. In a second part, we present, on the basis of the Gaussian isoperimetric and concentration inequalities, a large deviation theorem for Gaussian measures without topology. We conclude this chapter with a large deviation statement for the Ornstein-Uhlenbeck process.

A Gaussian measure μ on a real separable Banach space E equipped with its Borel σ-algebra $\mathcal{B}$ and with norm $\|\cdot\|$ is a Borel probability measure on $(E,\mathcal{B})$ such that the law of each continuous linear functional on E is Gaussian. Throughout this work, we only consider centered Gaussian measures or random variables. Although the study of the integrability properties may be developed in a single step from the isoperimetric or concentration inequalities of Chapters 1 and 2, we prefer to decompose the procedure, for pedagogical reasons, in two separate arguments.

Let thus μ be a centered Gaussian measure on $(E,\mathcal{B})$. We first claim that

$$\sigma = \sup_{\xi \in E^*, \|\xi\| \le 1} \left(\int \langle \xi, x\rangle^2 d\mu(x) \right)^{1/2} < \infty. \tag{4.1}$$

Indeed, if we denote by j the injection map from E^* into $\mathrm{L}^2(\mu) = \mathrm{L}^2(E,\mathcal{B},\mu;\mathbb{R})$, $\|j\| = \sigma$ and j is bounded by the closed graph theorem. Alternatively, let $m > 0$ be such that $\mu(x; \|x\| \le m) \ge \frac{1}{2}$. Then, for every element ξ in E^* with $\|\xi\| \le 1$, $\mu(x; |\langle \xi, x\rangle| \le m) \ge \frac{1}{2}$. Now, under μ, $\langle \xi, x\rangle$ is Gaussian with variance $\int \langle \xi, x\rangle^2 d\mu(x)$. Since $2[1-\Phi(\frac{1}{2})] > \frac{1}{2}$, it immediately follows that $(\int \langle \xi, x\rangle^2 d\mu(x))^{1/2} \le 2m$.

Since E is separable, the norm $\|\cdot\|$ on E may be described as a supremum over a countable set $(\xi_n)_{n\ge 1}$ of elements of the unit ball of the dual space E^*, that is, for every x in E,

$$\|x\| = \sup_{n\ge 1} \langle \xi_n, x\rangle.$$

In particular, the norm $\|\cdot\|$ can freely be used as a measurable map on $(E, \mathcal{B})$. Let $\Xi = \{\xi_1, \ldots, \xi_n\}$ be a finite subset of $(\xi_n)_{n\geq 1}$. Denote by $\Gamma = M\,{}^{\mathrm{t}}M$ the (semi-) positive definite covariance matrix of the Gaussian vector $(\langle \xi_1, x\rangle, \ldots, \langle \xi_n, x\rangle)$ on $\mathbb{R}^n$. This random vector has the same distribution as $M\Lambda$ where Λ is distributed according to the canonical Gaussian measure γ_n. Let then $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(z) = \max_{1\leq i\leq n} M(z)_i, \quad z = (z_1, \ldots, z_n) \in \mathbb{R}^n.$$

It is easily seen that the Lipschitz norm $\|f\|_{\mathrm{Lip}}$ of f is less than or equal to the norm $\|M\|$ of M as an operator from $\mathbb{R}^n$ equipped with the Euclidean norm into $\mathbb{R}^n$ with the supnorm, and that furthermore this operator norm $\|M\|$ is equal, by construction, to

$$\max_{1\leq i\leq n}\left(\int \langle \xi_i, x\rangle^2 \, d\mu(x)\right)^{1/2} \leq \sigma.$$

Therefore, inequality (2.8) applied to this Lipschitz function f yields, for every $r \geq 0$,

$$\mu\left(x; \sup_{\xi\in\Xi}\langle \xi, x\rangle \geq \int \sup_{\xi\in\Xi}\langle \xi, x\rangle d\mu(x) + r\right) \leq \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.2}$$

The same inequality applied to $-f$ yields

$$\mu\left(x; \sup_{\xi\in\Xi}\langle \xi, x\rangle + r \leq \int \sup_{\xi\in\Xi}\langle \xi, x\rangle d\mu(x)\right) \leq \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.3}$$

Let then r_0 be large enough so that $\exp(-r_0^2/2\sigma^2) < \frac{1}{2}$. Let also m be large enough in order that $\mu(x; \|x\| \leq m) \geq \frac{1}{2}$. Intersecting this probability with the one in (4.3) for $r = r_0$, we see that

$$\int \sup_{\xi\in\Xi}\langle \xi, x\rangle \, d\mu(x) \leq r_0 + m.$$

Since m and r_0 have been chosen independently of Ξ, we already notice that

$$\int \|x\| \, d\mu(x) < \infty.$$

Now, one can use monotone convergence in (4.2) and thus one obtains that, for every $r \geq 0$,

$$\mu\big(x; \|x\| \geq \textstyle\int \|x\| d\mu(x) + r\big) \leq \mathrm{e}^{-r^2/2\sigma^2}. \tag{4.4}$$

Note that an entirely similar result may be obtained exactly in the same way (even simpler) from the concentration inequality (2.4) around the median of a Lipschitz function. As an immediate consequence of (4.4), we may already state the basic theorem about the integrability properties of norms of Gaussian measures. The lower bound and necessity part easily follow from the scalar case. As we have seen in Chapter 2, the two parameters $\int \|x\| d\mu(x)$ and σ in inequality (4.4) may be very different so that this inequality is a much stronger result than the following well-known consequence.

Theorem 4.1. *Let μ be a centered Gaussian measure on a separable Banach space E with norm $\|\cdot\|$. Then*

$$\lim_{r\to\infty} \frac{1}{r^2} \log \mu\big(x; \|x\| \geq r\big) = -\frac{1}{2\sigma^2}\,.$$

In other words,

$$\int \exp\big(\alpha\|x\|^2\big)\, d\mu(x) < \infty \quad \textit{if and only if} \quad \alpha < \frac{1}{2\sigma^2}\,.$$

The question of the integrability (actually only the square integrability) of the norm of a Gaussian measure was first raised by L. Gross [Gr1], [Gr2]. In 1969, A. V. Skorohod [Sk] was able to show that $\int \exp(\alpha\|x\|)\, d\mu(x) < \infty$ (for every $\alpha > 0$) using the strong Markov property of Brownian motion. The existence of some $\alpha > 0$ for which $\int \exp(\alpha\|x\|^2)\, d\mu(x) < \infty$ was then established independently by X. Fernique [Fe2] and H. J. Landau and L. A. Shepp [L-S] (with a proof already isoperimetric in nature). The best possible value for α was first obtained in [M-S]. Recently, S. Kwapień mentioned to me that J.-P. Kahane, back in 1964 [Ka1] (cf. [Ka2]), proved an inequality on norms of Rademacher series which, together with a simple central limit theorem argument, already implied that $\int \exp(\alpha\|x\|)\, d\mu(x) < \infty$ for every $\alpha > 0$.

From inequality (4.4), we may also mention the equivalence of all moments of norms of Gaussian measures: for every $0 < p, q < \infty$, there exists a constant $C_{p,q} > 0$ only depending on p and q such that

$$\left(\int \|x\|^p\, d\mu(x)\right)^{1/p} \leq C_{p,q}\left(\int \|x\|^q\, d\mu(x)\right)^{1/q}. \tag{4.5}$$

For the proof, simply integrate by parts inequality (4.4) together with the fact that $\sigma \leq C_q(\int \|x\|^q\, d\mu(x))^{1/q}$ for every $q > 0$ by the one-dimensional equivalence of Gaussian moments. This yields (4.5) for every $q \geq 1$. When $0 < q \leq 1$, simply note that if $m = (2C_{2,1})^{2/q}\big(\int \|x\|^q d\mu(x)\big)^{1/q}$,

$$\begin{aligned}
\int \|x\| d\mu(x) &\leq m + \mu\big(x; \|x\| \geq m\big)^{1/2}\left(\int \|x\|^2\, d\mu(x)\right)^{1/2} \\
&\leq m + C_{2,1}\mu\big(x; \|x\| \geq m\big)^{1/2} \int \|x\| d\mu(x) \\
&\leq 2(2C_{2,1})^{2/q}\left(\int \|x\|^q d\mu(x)\right)^{1/q}
\end{aligned}$$

since $C_{2,1}\mu(x; \|x\| \geq m)^{1/2} \leq \frac{1}{2}$. Note that $C_{p,2}$ (for example) is of the order of $\sqrt{p}$ as p goes to infinity. We will come back to this remark in the last chapter where we will relate (4.5) to hypercontractivity. It is conjectured that $C_{2,1} = \sqrt{\pi/2}$ (that is, the constant of the real case). S. Szarek recently noticed that if conjecture (1.11) holds, then the best possible $C_{p,q}$ are given by the real case.

The preceding integrability properties may also be applied in the context of almost surely bounded Gaussian processes. Let $X = (X_t)_{t\in T}$ be a centered Gaussian process indexed by a set T on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $\sup_{t\in T} X_t(\omega) < \infty$ for almost all ω in Ω (or $\sup_{t\in T} |X_t(\omega)| < \infty$, which, by symmetry, is equivalent to the preceding, at least if the process is separable). Then, the same proof as above shows in particular that

$$\sup\big\{\mathbb{E}\big(\sup_{t\in U} X_t\big); U \text{ finite in } T\big\} < \infty.$$

We will actually take this as the definition of an almost surely bounded Gaussian process in Chapter 6. Under a separability assumption on the process, one can actually formulate the analogue of Theorem 4.1 in this context. Assume there exists a countable subset S of T such that the set $\{\omega; \sup_{t\in T} X_t \neq \sup_{t\in S} X_t\}$ is negligible. Set $\|X\| = \sup_{t\in S} X_t$. Then, provided $\|X\| < \infty$ almost surely,

$$\mathbb{E}\big(\exp\big(\alpha\|X\|^2\big)\big) < \infty \quad \text{if and only if} \quad \alpha < \frac{1}{2\sigma^2}$$

with $\sigma^2 = \sup_{t\in S} \mathbb{E}(X_t^2)$ $(= \sup_{t\in T} \mathbb{E}(X_t^2))$.

As still another remark, notice that the proof of Theorem 4.1 also shows that whenever $X = (X_1, \ldots, X_n)$ is a centered Gaussian random vector in $\mathbb{R}^n$, then

$$\mathrm{var}\big(\max_{1\le i\le n} X_i\big) \le \max_{1\le i\le n} \mathrm{var}(X_i).$$

(Use again the Lipschitz map $f(z) = \max_{1\le i\le n} M(z)_i$ where $\Gamma = M\,{}^{\mathrm{t}}M$ is the covariance matrix of X with however (2.4) instead of (2.8).) This inequality may however be deduced directly from the Poincaré type inequality

$$\int \Big|f - \int f d\gamma_n\Big|^2 d\gamma_n \le \int |\nabla f|^2 d\gamma_n \quad (\le \|f\|_{\mathrm{Lip}}^2)$$

which is elementary (by an expansion in Hermite polynomials for example).

Our aim will now be to extend the isoperimetric and concentration inequalities to the setting of an infinite dimensional Gaussian measure μ as before. Let us mention however before that the fundamental inequalities are the ones in finite dimension and that the infinite dimensional extensions we will present actually follow in a rather classical and straightforward manner from the finite dimensional case. The main tool will be the concept of abstract Wiener space and reproducing kernel Hilbert space which will define the isoperimetric neighborhoods or enlargements in this framework. We follow essentially C. Borell [Bo3] in the construction below.

Let μ be a mean zero Gaussian measure on a real separable Banach space E. Consider then the abstract Wiener space factorization [Gr1], [B-C], [Ku], [Bo3] (for a recent account, cf. [Lif3]),

$$E^* \xrightarrow{\ j\ } \mathrm{L}^2(\mu) \xrightarrow{\ j^*\ } E.$$

First note that since E is separable and μ is a Borel probability measure on E, μ is Radon, that is, for every $\varepsilon > 0$ there is a compact set K in E such that $\mu(K) \geq 1-\varepsilon$. Let $(K_n)_{n\in\mathbb{N}}$ be a sequence of compact sets such that $\mu(K_n) \to 1$. If φ is an element of $\mathrm{L}^2(\mu)$, $j^*(\varphi I_{K_n})$ belongs to E since it may be identified with the expectation, in the strong sense, $\int_{K_n} x\varphi(x)d\mu(x)$. Now, the sequence $\left(\int_{K_n} x\varphi(x)d\mu(x)\right)_{n\in\mathbb{N}}$ is Cauchy in E since,

$$\sup_{\xi\in E^*, \|\xi\|\leq 1} \langle \xi, \textstyle\int_{K_n} x\varphi(x)d\mu(x) - \int_{K_m} x\varphi(x)d\mu(x)\rangle \leq \sigma\left(\displaystyle\int \varphi^2 |I_{K_n} - I_{K_m}| d\mu\right)^{1/2} \to 0.$$

It therefore converges in E to the weak integral $\int x\varphi(x)d\mu(x) = j^*(\varphi) \in E$.

Define now the reproducing kernel Hilbert space $\mathcal{H}$ of μ as the subspace $j^*(\mathrm{L}^2(\mu))$ of E. Since $j(E^*)^\perp = \mathrm{Ker}(j^*)$, j^* restricted to the closure E_2^* of E^* in $\mathrm{L}^2(\mu)$ is linear and bijective onto $\mathcal{H}$. For simplicity in the notation, we set below $\widetilde{h} = ({j^*}_{|E_2^*})^{-1}(h)$. Under μ, $\widetilde{h}$ is Gaussian with variance $|h|^2$. Note that σ of (4.1) is then also $\sup_{x\in\mathcal{K}} \|x\|$ where $\mathcal{K}$ is the closed unit ball of $\mathcal{H}$ for its Hilbert space scalar product given by

$$\langle j^*(\varphi), j^*(\psi)\rangle_{\mathcal{H}} = \langle \varphi, \psi\rangle_{\mathrm{L}^2(\mu)}, \quad \varphi, \psi \in \mathrm{L}^2(\mu).$$

In particular, for every x in $\mathcal{H}$, $\|x\| \leq \sigma|x|$ where $|x| = |x|_{\mathcal{H}} = \langle x, x\rangle_{\mathcal{H}}^{1/2}$. Moreover, $\mathcal{K}$ is a compact subset of E. Indeed, if $(\xi_n)_{n\in\mathbb{N}}$ is a sequence in the unit ball of E^*, there is a subsequence $(\xi_{n'})_{n'\in\mathbb{N}}$ which converges weakly to some ξ in E^*. Now, since the ξ_n are Gaussian under μ, $\xi_{n'} \to \xi$ in $\mathrm{L}^2(\mu)$ so that j is a compact operator. Hence j^* is also a compact operator which is the claim.

For γ_n the canonical Gaussian measure on $\mathbb{R}^n$ (equipped with some arbitrary norm), it is plain that $\mathcal{H} = \mathbb{R}^n$ with its Euclidean structure, that is $\mathcal{K}$ is the Euclidean unit ball $B(0,1)$. If $X = (X_1, \ldots, X_n)$ is a centered Gaussian measure on $\mathbb{R}^n$ with nondegenerate covariance matrix $\Gamma = M\,{}^tM$, it is easily seen that the unit ball $\mathcal{K}$ of the reproducing kernel Hilbert space associated to the distribution of X is the ellipsoid $M(B(0,1))$. As another example, let us mention the classical Wiener space associated with Brownian motion, say on [0,1] and with real values for simplicity. Let thus E be the Banach space $C_0([0,1])$ of all real continuous functions x on [0,1] vanishing at the origin equipped with the supnorm (the Wiener space) and let μ be the distribution of a standard Brownian motion, or Wiener process, $W = (W(t))_{t\in[0,1]}$ starting at the origin (the Wiener measure). If m is a finitely supported measure on [0,1], $m = \sum_i c_i\delta_{t_i}$, $c_i \in \mathbb{R}$, $t_i \in [0,1]$, clearly $h = j^*j(m)$ is the element of E given by

$$h(t) = \sum_i c_i(t_i \wedge t), \quad t \in [0,1];$$

it satisfies

$$\int_0^1 h'(t)^2\,dt = \int \langle m, x\rangle^2 d\mu(x) = |h|_{\mathcal{H}}^2.$$

By a standard extension, the reproducing kernel Hilbert space $\mathcal{H}$ associated to the Wiener measure μ on E may then be identified with the Cameron-Martin Hilbert space of the absolutely continuous elements h of $C_0([0,1])$ such that

$$\int_0^1 h'(t)^2\,dt < \infty.$$

Moreover, if $h \in \mathcal{H}$, $\widetilde{h} = (j^*{}_{|E_2^*})^{-1}(h) = \int_0^1 h'(t)dW(t)$. While we equipped the Wiener space $C_0([0,1])$ with the uniform topology, other choices are possible. Let F be a separable Banach space such that the Wiener process W belongs almost surely to F. Using probabilistic notation, we know from the previous abstract Wiener space theory that if φ is a real valued random variable with $\mathbb{E}(\varphi^2) < \infty$, then $h = \mathbb{E}(W\varphi) \in F$. Since $\mathbb{P}\{W \in F \cap C_0([0,1])\} = 1$, it immediately follows that the Cameron-Martin Hilbert space may be identified with a subset of F and is also the reproducing kernel Hilbert space of Wiener measure on F. For h in the Cameron-Martin space, $\widetilde{h} = (j^*{}_{|F_2^*})^{-1}(h)$ may be identified with $\int_0^1 h'(t)dW(t)$ as soon as there is a sequence $(\xi_n)_{n\in\mathbb{N}}$ in F^* such that

$$\mathbb{E}\left(\left|\int_0^1 h'(t)dW(t) - \langle \xi_n, W\rangle\right|^2\right) \to 0.$$

This is the case if, for every $t \in [0,1]$, there is $(\xi_n)_{n\in\mathbb{N}}$ in F^* with

$$\mathbb{E}\big(\big|W(t) - \langle \xi_n, W\rangle\big|^2\big) \to 0.$$

Examples include the Lebesgue spaces $\mathrm{L}^p([0,1])$, $1 \le p < \infty$, or the Hölder spaces (see below). Actually, since the preceding holds for the L^1-norm, this will be the case for a norm $\|\cdot\|$ on $C_0([0,1])$ as soon as, for some constant $C > 0$, $\|x\| \ge C\int_0^1 |x(t)|dt$ for every x in $C_0([0,1])$.

The next proposition is a useful series representation of Gaussian measures and random vectors which can be used efficiently in proofs by finite dimensional approximation. This proposition puts forward the fundamental Gaussian measurable structure consisting of the canonical Gaussian product measure on $\mathbb{R}^{\mathbb{N}}$ with reproducing kernel Hilbert space ℓ^2.

Proposition 4.2. *Let μ be as before. Let $(g_i)_{i\ge1}$ denote an orthonormal basis of the closure E_2^* of E^* in $\mathrm{L}^2(\mu)$ and set $e_i = j^*(g_i)$, $i \ge 1$. Then $(e_i)_{i\ge1}$ defines an orthonormal basis of $\mathcal{H}$ and the series $X = \sum_{i=1}^\infty g_i e_i$ converges in E μ-almost everywhere and in every L^p and is distributed as μ.*

Proof. Since μ is a Radon measure, the space $\mathrm{L}^2(\mu)$ is separable and E_2^* consists of Gaussian random variables on the probability space $(E, \mathcal{B}, \mu)$. Hence, $(g_i)_{i\ge1}$ defines on this space a sequence of independent standard Gaussian random variables. The sequence $(e_i)_{i\ge1}$ is clearly a basis in $\mathcal{H}$. Recall from Theorem 4.1 that the integral $\int \|x\|d\mu(x)$ is finite. Denote then by $\mathcal{B}_n$ the σ-algebra generated by $g_1, \ldots, g_n$. It is easily seen that the conditional expectation of the identity map on (E,μ) with respect to $\mathcal{B}_n$ is equal to $X_n = \sum_{i=1}^n g_i e_i$. By the vector valued martingale convergence theorem (cf. [Ne2]), the series $\sum_{i=1}^\infty g_i e_i$ converges almost surely. Since $\int \|x\|^p d\mu(x) < \infty$ for every $p > 0$, the convergence also takes place in any L^p-space. Since moreover

$$\int \langle \xi, X\rangle^2 d\mu = \sum_{i=1}^\infty \langle \xi, e_i\rangle^2 = \sum_{i=1}^\infty \langle j(\xi), g_i\rangle^2 = \int \langle \xi, x\rangle^2 d\mu(x)$$

for every ξ in E^*, X has law μ. Proposition 4.2 is proved. □

According to Proposition 4.2, we use from time to time below more convenient probabilistic notation and consider $(g_i)_{i\geq 1}$ as a sequence of independent standard Gaussian random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and X as a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in E and law μ.

As a consequence of Proposition 4.2, note that the closure $\overline{\mathcal{H}}$ of $\mathcal{H}$ in E coincides with the support of μ (for the topology given by the norm on E). Indeed, by Proposition 4.2, $\mathrm{supp}(\mu) \subset \overline{\mathcal{H}}$. Conversely, it suffices to prove that $\mu(B(h,\eta)) > 0$ for every h in $\mathcal{H}$ and every $\eta > 0$ where $B(h,\eta)$ is the ball in E with center h and radius η. By the Cameron-Martin translation formula (see below), it suffices to prove it for $h = 0$. Now, for every $a \in E$, by symmetry and independence,

$$\begin{aligned}\mu\big(B(a,\eta)\big)^2 &= \mu\big(x; \|x-a\| \leq \eta\big)\mu\big(x; \|x+a\| \leq \eta\big)\\ &\leq \mu\otimes\mu\big((x,y); \big\|(x-a)+(y+a)\big\| \leq 2\eta\big)\\ &= \mu\big(B(0,\eta\sqrt{2})\big)\end{aligned}$$

since $x+y$ under $\mu\otimes\mu$ is distributed as $\sqrt{2}x$ under μ. Now, assume that $\mu(B(h,\eta_0)) = 0$ for some $\eta_0 > 0$. Since μ is Radon, there is a sequence $(a_n)_{n\in\mathbb{N}}$ in E such that

$$\mu\big(x; \exists\, n, \|x-a_n\| \leq \eta_0/\sqrt{2}\big) = 1.$$

Then,

$$1 \leq \sum_n \mu\big(B(a_n,\eta_0/\sqrt{2})\big) \leq \sum_n \mu\big(B(0,\eta_0)\big)^{1/2} = 0$$

which is a contradiction (cf. also [D-HJ-S]).

To complete this brief description of the reproducing kernel Hilbert space of a Gaussian measure, let us mention the dual point of view more commonly used by analysts on Wiener spaces (cf. [Ku] for further details). Let $\mathcal{H}$ be a real separable Hilbert space with norm $|\cdot|$ and let $e_1, e_2, \ldots$ be an orthormal basis of $\mathcal{H}$. Define a simple additive measure ν on the cylinder sets in $\mathcal{H}$ by

$$\nu\big(x \in \mathcal{H}; \big(\langle x, e_1\rangle, \ldots, \langle x, e_n\rangle\big) \in A\big) = \gamma_n(A)$$

for all Borel sets A in $\mathbb{R}^n$. Let $\|\cdot\|$ be a measurable seminorm on $\mathcal{H}$ and denote by E the completion of $\mathcal{H}$ with respect to $\|\cdot\|$. Then $(E, \|\cdot\|)$ is a real separable Banach space. If $\xi \in E^*$, we consider $\xi_{|\mathcal{H}} : \mathcal{H} \to \mathbb{R}$ that we identify with an element h in $\mathcal{H} = \mathcal{H}^*$ (in our language, $h = j^*j(\xi)$). Let then μ be the (σ-additive) extension of ν

on the Borel sets of E. In particular, the distribution of $\xi \in E^*$ under μ is Gaussian with mean zero and variance $|h|^2$. Therefore, μ is a Gaussian Radon measure on E with reproducing kernel Hilbert space $\mathcal{H}$. With respect to this approach, our construction priviledges the point of view of the measure.

We are now ready to state and prove the isoperimetric inequality in $(E, \mathcal{H}, \mu)$. As announced, the isoperimetric neighborhoods A_r, $r \geq 0$, of a set A in E will be understood in this setting as the Minkowski sum $A + r\mathcal{K} = \{x + ry; x \in A, y \in \mathcal{K}\}$ where we recall that $\mathcal{K}$ is the unit ball of the reproducing kernel Hilbert space $\mathcal{H}$ associated to the Gaussian measure μ. In this form, the result is due to C. Borell [Bo2].

Theorem 4.3. *Let A be a Borel set in E such that $\mu(A) \geq \Phi(a)$ for some real number a. Then, for every $r \geq 0$*

$$\mu_*(A + r\mathcal{K}) \geq \Phi(a + r).$$

It might be worthwhile mentioning that if the support of μ is infinite dimensional, $\mu(\mathcal{H}) = 0$ so that the infinite dimensional version of the Gaussian isoperimetric inequality might be somewhat more surprising than its finite dimensional statement. The inner measure in Theorem 4.3 is necessary since $A + r\mathcal{K}$ need not always be measurable.

Proof. As announced, it is based on a classical finite dimensional approximation procedure. We use the series representation $X = \sum_{i=1}^{\infty} g_i e_i$ of Proposition 4.2 and, accordingly, probabilistic notations. We may assume that $-\infty < a < +\infty$. Let $r \geq 0$ be fixed. Let also $\varepsilon > 0$. Since μ is a Radon measure, there exists a compact set $K \subset A$ such that

$$\mathbb{P}\{X \in K\} = \mu(K) \geq \Phi(a - \varepsilon).$$

For every $\eta > 0$, let $K^\eta = \{x \in E; \inf_{y \in K} \|x - y\| \leq \eta\}$. Recall $X_n = \sum_{i=1}^{n} g_i e_i$. Since $\mathbb{P}\{\|X - X_n\| > \eta\} \to 0$, for some n_0 and every $n \geq n_0$, $\mathbb{P}\{X_n \in K^\eta\} \geq \Phi(a - 2\varepsilon)$ and

$$\mathbb{P}\{X \in K^{3\eta} + r\mathcal{K}\} \geq \mathbb{P}\{X_n \in K^{2\eta} + r\mathcal{K}\} - \varepsilon.$$

Now, let $\mathcal{K}_n$ be the unit ball of the reproducing kernel Hilbert space of the (finite dimensional) Gaussian random vector X_n, or rather of its distribution on E. $\mathcal{K}_n$ consists of those elements in E of the form $\mathbb{E}(X_n \varphi)$ with $\|\varphi\|_2 \leq 1$. Clearly,

$$\big\|\mathbb{E}(X\varphi) - \mathbb{E}(X_n\varphi)\big\| \leq \big(\mathbb{E}\|X - X_n\|^2\big)^{1/2} \to 0$$

independently of φ, $\|\varphi\|_2 \leq 1$. Hence, for some $n_1 \geq n_0$, and every $n \geq n_1$,

$$\mathbb{P}\{X \in K^{3\eta} + r\mathcal{K}\} \geq \mathbb{P}\{X_n \in K^{\eta} + r\mathcal{K}_n\} - \varepsilon.$$

Let Q be the map from $\mathbb{R}^n$ into E defined by $Q(z) = \sum_{i=1}^{n} z_i e_i$, $z = (z_1, \ldots, z_n)$. Therefore

$$\gamma_n\big(Q^{-1}(K^\eta)\big) = \mathbb{P}\{X_n \in K^\eta\} \geq \Phi(a - 2\varepsilon).$$

Since the distribution of X_n is the image by Q of γ_n and since similarly $\mathcal{K}_n$ is the image by Q of the Euclidean unit ball, it follows from Theorem 1.3 that

$$\mathbb{P}\{X_n \in K^\eta + r\mathcal{K}_n\} = \gamma_n\big(\big(Q^{-1}(K^\eta)\big)_r\big) \geq \Phi(a - 2\varepsilon + r).$$

Summarizing, for every $\eta > 0$,

$$\mu(K^{3\eta} + r\mathcal{K}) \geq \Phi(a - 2\varepsilon + r) - \varepsilon.$$

Since K and $\mathcal{K}$ are compact in E, letting η decrease to zero yields

$$\mu_*(A + r\mathcal{K}) \geq \mu(K + r\mathcal{K}) \geq \Phi(a - 2\varepsilon + r) - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the theorem is proved. □

The approximation procedure developed in the proof of Theorem 4.3 may be used exactly in the same way on the basis of inequality (2.15) to show that, for every $r \geq 0$,

$$\mu_*(A + r\mathcal{K}) \geq 1 - \exp\bigg(-\frac{r^2}{2} + r\delta\big(\mu(A)\big)\bigg) \tag{4.6}$$

where we recall that

$$\delta(v) = \int_0^\infty \min\big(1 - v, \mathrm{e}^{-t^2v^2/2}\big)dt, \quad 0 \leq v \leq 1.$$

The point here is that inequality (2.15) (and thus also inequality (4.6)) was obtained at the very cheap price of Proposition 2.1. In what follows, inequality (4.6) will be good enough for almost all the applications we have in mind.

Theorem 4.3, or inequality (4.6), of course allows us to recover the integrability properties described in Theorem 4.1. For example, if $f : E \to \mathbb{R}$ is measurable and Lipschitz in the direction of $\mathcal{H}$, that is

$$\big|f(x+h) - f(x)\big| \leq |h| \quad \text{for all } x \in E,\ h \in \mathcal{H}, \tag{4.7}$$

and if m is median of f for μ, exactly as in the finite dimensional case (2.4),

$$\mu(f \geq m + r) \leq 1 - \Phi(r) \leq \mathrm{e}^{-r^2/2} \tag{4.8}$$

for every $r \geq 0$. In the same way, a finite dimensional argument on (2.8) shows that $\int f d\mu$ exists and that, for all $r \geq 0$,

$$\mu\big(f \geq \textstyle\int f d\mu + r\big) \leq \mathrm{e}^{-r^2/2}. \tag{4.9}$$

Indeed, assume first that f is bounded. We follow Proposition 4.2 and its notation. Let f_n, $n \geq 1$, be the conditional expectation of f with respect to $\mathcal{B}_n$. Define $\widetilde{f}_n : \mathbb{R}^n \to \mathbb{R}$ by

$$\widetilde{f}_n(z) = \int f\bigg(\sum_{i=1}^n z_i e_i + y\bigg) d\mu^n(y), \quad z = (z_1, \ldots, z_n) \in \mathbb{R}^n,$$

where μ^n is the distribution of $\sum_{i=n+1}^{\infty} g_i e_i$. Then f_n under μ has the same distribution as $\widetilde{f}_n$ under γ_n. Moreover, it is clear by (4.7) that $\widetilde{f}_n$ is Lipschitz in the usual sense on $\mathbb{R}^n$ with $\|\widetilde{f}_n\|_{\mathrm{Lip}} \leq 1$. Therefore, by (2.8) applied to $\widetilde{f}_n$, for every $r \geq 0$,

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

Letting n tend to infinity, we see that (4.9) is satisfied for bounded functionals f on E satisfying (4.7). When f is not bounded, set, for every integer N,

$$f^N = \min\big(\max(f, -N), N\big).$$

Then f^N still satisfies (4.7) for each N so that

$$\mu\big(f^N \geq \textstyle\int f^N d\mu + r\big) \leq \mathrm{e}^{-r^2/2}$$

for every $r \geq 0$. Of course, the same result holds for $|f^N|$. Let then m be such that $\mu(|f| \leq m) \geq \frac{3}{4}$. There exists N_0 such that for every $N \geq N_0$, $\mu(|f^N| \leq m+1) \geq \frac{1}{2}$. Let $r_0 \geq 0$ be such that $\mathrm{e}^{-r_0^2/2} < \frac{1}{2}$. Together with the preceding inequality for $|f^N|$, we thus get that for every $N \geq N_0$,

$$\int |f^N| d\mu \leq m + 1 + r_0.$$

Moreover, $\mu(|f^N| \geq m + 1 + r_0 + r) \leq \mathrm{e}^{-r^2/2}$, $r \geq 0$. Hence, in particular, the supremum $\sup_N \int |f^N|^2 d\mu$ is finite. The announced claim (4.9) now easily follows by uniform integrability.

Let us also mention that the preceding inequalities (4.8) and (4.9) may of course be applied to $f(x) = \|x\|$, $x \in E$, since, as we have seen,

$$\big|\|x+h\| - \|x\|\big| \leq \|h\| \leq \sigma|h|, \quad x \in E,\, h \in \mathcal{H}.$$

It should be noticed that the $\mathcal{H}$-Lipschitz hypothesis (4.7) has recently been shown [E-S] to be equivalent to the fact that the Malliavin derivative Df of f exists and satisfies $\||Df|_{\mathcal{H}}\|_\infty \leq 1$. (Due to the preceding simple arguments, the hypothesis that f be in $\mathrm{L}^2(\mu)$ in the paper [E-S] is easily seen to be superfluous.) But actually, that (4.7) holds when $\||Df|_{\mathcal{H}}\|_\infty \leq 1$ is the easy part of the argument so that the preceding result is as general as possible. One could also prove (4.9) along the lines of Proposition 2.1 in infinite dimension with the Ornstein-Uhlenbeck semigroup associated to μ. One however runs into the question of differentiability in infinite dimension (Gross-Malliavin derivatives) that is not really needed here.

In the preceding spirit, it might be worthwhile to briefly describe some related inequalities due to B. Maurey and G. Pisier [Pi1]. Let f be of class C^1 on $\mathbb{R}^n$ with gradient ∇f. Let furthermore V be a convex function on $\mathbb{R}$. To avoid integrability questions, assume first that f is bounded. By Jensen's inequality,

$$\int V\big(f - \textstyle\int f d\gamma_n\big) d\gamma_n \leq \displaystyle\iint V\big(f(x) - f(y)\big) d\gamma_n(x) d\gamma_n(y).$$

Now, for x, y in $\mathbb{R}^n$, and every real number θ, set

$$x(\theta) = x\sin\theta + y\cos\theta, \quad x'(\theta) = x\cos\theta - y\sin\theta.$$

We have

$$f(x) - f(y) = \int_0^{\pi/2} \frac{d}{d\theta} f\big(x(\theta)\big)d\theta = \int_0^{\pi/2} \langle \nabla f\big(x(\theta)\big), x'(\theta)\rangle d\theta.$$

Hence, using Jensen's inequality one more time but now with respect to the variable θ,

$$\int V\big(f - \textstyle\int f d\gamma_n\big)d\gamma_n \le \frac{2}{\pi}\int_0^{\pi/2}\iint V\Big(\frac{\pi}{2}\langle\nabla f\big(x(\theta)\big), x'(\theta)\rangle\Big)d\gamma_n(x)d\gamma_n(y)d\theta.$$

By the fundamental rotational invariance of Gaussian measures, for any θ, the couple $(x(\theta), x'(\theta))$ has the same distribution as the original independent couple (x, y). Therefore, we obtained that

$$\int V\big(f - \textstyle\int f d\gamma_n\big)d\gamma_n \le \iint V\Big(\frac{\pi}{2}\langle\nabla f(x), y\rangle\Big)d\gamma_n(x)d\gamma_n(y). \tag{4.10}$$

We leave it to the interested reader to properly extend this type of inequality to unbounded functions. It also easily extends to infinite dimensional Gaussian measures μ. Indeed, let f be smooth enough, more precisely differentiable in the direction of $\mathcal{H}$ or in the sense of Gross-Malliavin (cf. e.g. [Bel], [Wa], [Nu]...). With the same notation as in the proof of (4.9),

$$\nabla\widetilde{f}_n = (D_{e_1}f, \ldots, D_{e_n}f),$$

where $D_h f$ is the derivative of f in the direction of $h \in \mathcal{H}$. Therefore, for every n,

$$\int V\big(f_n - \textstyle\int f_n d\mu\big)d\mu \le \iint V\Big(\frac{\pi}{2}\sum_{i=1}^n y_i D_{e_i}f(x)\Big)d\mu(x)d\gamma_n(y).$$

Hence, by Fatou's lemma and Jensen's inequality, (4.10) yields in an infinite dimensional setting that

$$\int V\big(f - \textstyle\int f d\mu\big)d\mu \le \iint V\Big(\frac{\pi}{2}\sum_{i=1}^\infty y_i D_{e_i}f(x)\Big)d\mu(x)d\gamma_\infty(y)$$

where γ_∞ is the canonical Gaussian product measure on $\mathbb{R}^{\mathbb{N}}$. If V is an exponential function $e^{\lambda x}$, we may perform partial integration in the variable y to get that

$$\int \exp\big[\lambda\big(f - \textstyle\int f d\mu\big)\big]d\mu \le \int \exp\Big(\frac{\lambda^2\pi^2}{4}\,|Df|^2_{\mathcal{H}}\Big)d\mu.$$

In particular, if f is Lipschitz, we recover in this way an inequality similar to (2.8) (or (4.9)) with however a worse constant. Inequality (4.10) is however more general and applies moreover to vector valued functions (cf. [Pi1]).

So far, we only used isoperimetry and concentration in a very mild way for the application to the integrability properties. As we have seen, there is however a strong difference between these integrability properties (Theorem 4.1) and, for example, inequalities (4.4), (4.8) or (4.9). In these inequalities indeed, two parameters, and not only one, on the Gaussian measure enter into the problem, namely the median or the mean of the $\mathcal{H}$-Lipschitz map f and its Lipschitz norm (the supremum σ of weak variances in the case of a norm). These can be very different even in simple examples.

We now present another application of the Gaussian isoperimetric inequality due to M. Talagrand [Ta1]. It is a powerful strengthening on Theorem 4.1 that makes critical use of the preceding comment. (See also [G-K1] for some refinement.) More on Theorem 4.4 may be found in Chapter 7.

Theorem 4.4. *Let μ be a Gaussian measure on E. For every $\varepsilon > 0$, there exists $r_0 = r_0(\varepsilon)$ such that for every $r \geq r_0$,*

$$\mu\big(x; \|x\| \geq \varepsilon + \sigma r\big) \leq \exp\biggl(-\frac{r^2}{2} + \varepsilon r\biggr).$$

Ehrhard's inequality (1.8) (or rather its infinite dimensional extension) indicates that the map $F(r) = \Phi^{-1}(\mu(x; \|x\| \leq r))$, $r \geq 0$, is concave. While Theorem 4.1 expresses that $\lim_{r\to\infty} F(r)/r = 0$, Theorem 4.4 yields $\lim_{r\to\infty}[F(r) - (r/\sigma)] = \frac{1}{\sigma}$. In other words, the line r/σ is an asymptote at infinity to F. Notice furthermore that Theorem 4.4 implies (is equivalent to saying) that

$$\int \exp\biggl(\frac{1}{2\sigma^2}\,\big(\|x\| - \varepsilon\big)^2\biggr) d\mu(x) < \infty$$

for all $\varepsilon > 0$.

Proof. Recall the series $X = \sum_{i=1}^{\infty} g_i e_i$ of Proposition 4.2 which we consider on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $X_n = \sum_{i=1}^{n} g_i e_i$ and $X^n = X - X_n$, $n \geq 1$. Let $\varepsilon > 0$ be fixed and set $A = \{x \in E; \|x\| < \varepsilon\}$. For every $r \geq 0$ and every integer $n \geq 1$, we can write

$$\begin{aligned} \mathbb{P}\big\{\|X\| \geq \varepsilon + \sigma r\big\} &\leq \mathbb{P}\{X \notin A + r\mathcal{K}\} \\ &\leq \mathbb{P}\big\{|X_n| > r\big\} + \mathbb{P}\big\{|X_n| \leq r, X \notin A + r\mathcal{K}\big\}. \end{aligned}$$

On the set $\{|X_n| \leq r\}$, $X \notin A + r\mathcal{K}$ implies that

$$X^n \notin A + \big(r^2 - |X_n|^2\big)^{1/2}\mathcal{K}^n$$

where $\mathcal{K}^n$ is the unit ball of the reproducing kernel Hilbert space associated to the distribution of X^n. Indeed, if this is not the case,

$$X^n = a + \big(r^2 - |X_n|^2\big)^{1/2} h^n$$

for some $a \in A$ and $h^n \in \mathcal{K}^n$. This would imply that

$$X = X_n + X^n = a + X_n + \left(r^2 - |X_n|^2\right)^{1/2} h^n = a + k$$

where, by orthogonality, $|k| \leq r$. Therefore,

$$\mathbb{P}\{X \notin A + r\mathcal{K}\} \leq \mathbb{P}\big\{|X_n| > r\big\} + \mathbb{P}\big\{|X_n| \leq r, X^n \notin A + \left(r^2 - |X_n|^2\right)^{1/2}\mathcal{K}^n\big\}.$$

Recall now the function δ of (2.12) or (4.6) and choose n large enough in order that $\delta(\mathbb{P}\{\|X^n\| < \varepsilon\}) \leq \varepsilon$. Now, X_n and X^n are independent and $|X_n| = \left(\sum_{i=1}^n g_i^2\right)^{1/2}$. Hence, by inequality (4.6),

$$\begin{aligned}
\mathbb{P}\big\{|X_n| \leq r, X^n \notin & A + \left(r^2 - |X_n|^2\right)^{1/2}\mathcal{K}^n\big\} \\
&\leq \int_{\{|X_n| \leq r\}} \exp\left(-\frac{1}{2}\left(r^2 - |X_n|^2\right) + \varepsilon\left(r^2 - |X_n|^2\right)^{1/2}\right) d\mathbb{P} \\
&\leq C_n r^n \exp\left(-\frac{r^2}{2} + \varepsilon r\right)
\end{aligned}$$

where $C_n > 0$ only depends on n. In summary,

$$\mathbb{P}\big\{\|X\| \geq \varepsilon + \sigma r\big\} \leq \mathbb{P}\left\{\sum_{i=1}^n g_i^2 > r^2\right\} + C_n r^n \exp\left(-\frac{r^2}{2} + \varepsilon r\right)$$

from which the conclusion immediately follows. Theorem 4.4 is established. □

We now present some further applications of isoperimetry and concentration to the study of large deviations of Gaussian measures. As an introduction to these ideas, we first present the elementary concentration proof, due to S. Chevet [Che], of the upper bound in the large deviation principle for Gaussian measures.

Let μ be as before a mean zero Gaussian measure on a separable Banach space E with reproducing kernel Hilbert space $\mathcal{H}$. For a subset A of E, let

$$\mathcal{I}(A) = \inf\big\{\tfrac{1}{2}|h|^2; h \in A \cap \mathcal{H}\big\}$$

be the classical large deviation rate functional in this setting. Set $\mu_\varepsilon(\cdot) = \mu(\varepsilon^{-1}(\cdot))$, $\varepsilon > 0$. Let now A be closed in E and take r such that $0 < r < \mathcal{I}(A)$. By the very definition of $\mathcal{I}(A)$,

$$A \cap \sqrt{2r}\mathcal{K} = \emptyset.$$

Since A is closed and the balls in $\mathcal{H}$ are compact in E, there exists $\eta > 0$ such that we still have

$$A \cap \left[\sqrt{2r}\mathcal{K} + B_E(0, \eta)\right] = \emptyset$$

where $B_E(0, \eta)$ is the ball with center the origin and with radius η for the norm $\|\cdot\|$ in E. Since

$$\lim_{\varepsilon \to 0} \mu\big(B_E(0, \varepsilon^{-1}\eta)\big) = \lim_{\varepsilon \to 0} \mu_\varepsilon\big(B_E(0, \eta)\big) = 1,$$

it is then an immediate consequence of (4.6) (or Theorem 4.3) that for every $\varepsilon > 0$ small enough

$$\mu_\varepsilon(A) \le \mu\big(\big[\varepsilon^{-1}\sqrt{2r}\mathcal{K} + B_E(0,\varepsilon^{-1}\eta)\big]^c\big) \le \exp\bigg(-\frac{r}{\varepsilon^2} + \frac{\sqrt{2r}}{\varepsilon}\bigg).$$

Therefore, since $r < \mathcal{I}(A)$ is arbitrary,

$$\limsup_{\varepsilon \to 0} \varepsilon^2 \log \mu_\varepsilon(A) \le -\mathcal{I}(A).$$

This simple proof may easily be modified to yield some version of the large deviation theorem with only "measurable operations" on the sets. One may indeed ask about the role of the topology in a large deviation statement. As we will see, the isoperimetric and concentration ideas in this Gaussian setting are powerful enough to state a large deviation principle without any topological operations of closure or interior.

Let, as before, $(E, \mathcal{H}, \mu)$ be an abstract Wiener space. If A and B are subsets of E, and if λ is a real number, we set

$$\lambda A + B = \{\lambda x + y; x \in A, y \in B\},$$
$$A \ominus B = \{x \in A; x + B \subset A\}.$$

Crucial to the approach is the class $\mathcal{V}$ of all Borel subsets V of E such that

$$\liminf_{\varepsilon \to 0} \mu_\varepsilon(V) > 0.$$

Notice that if $V \in \mathcal{V}$, then $\lambda V \in \mathcal{V}$ for every $\lambda > 0$. Typically, the balls $B_E(0,\eta)$, $\eta > 0$, for the norm $\|\cdot\|$ on E belong to $\mathcal{V}$ while the balls in the reproducing kernel Hilbert space $\mathcal{H}$ do not (when the support of μ is infinite dimensional). A starlike subset V of E of positive measure belongs to $\mathcal{V}$.

In the example of Wiener measure on $C_0([0,1])$, the balls centered at the origin for the Hölder norm $\|\cdot\|_\alpha$ of exponent α, $0 < \alpha < \frac{1}{2}$, given by

$$\|x\|_\alpha = \sup_{0 \le s \ne t \le 1} \frac{|x(s) - x(t)|}{|s-t|^\alpha}, \quad x \in C_0([0,1]),$$

do belong to the class $\mathcal{V}$. Actually, the balls of any reasonable norm on Wiener space for which Wiener measure is Radon are in $\mathcal{V}$. Using the properties of the Brownian paths, many other examples of elements of $\mathcal{V}$ may be imagined (cf. [B-BA-K]).

Provided with the preceding notation, we introduce new rate functionals, on the subsets of E rather than the points. For a Borel subset A of E, set

$$r(A) = \sup\big\{r \ge 0; \exists\, V \in \mathcal{V}, (V + r\mathcal{K}) \cap A = \emptyset\big\}$$

($r(A) = 0$ if $\{\ \} = \emptyset$) and

$$s(A) = \inf\big\{s \ge 0; \exists\, V \in \mathcal{V}, (A \ominus V) \cap (s\mathcal{K}) \ne \emptyset\big\}$$

$(s(A) = \infty$ if $\{ \ \} = \emptyset)$. The functionals $r(\cdot)$ and $s(\cdot)$ are decreasing for the inclusion. Furthermore, it is elementary to check that $\frac{1}{2}r(A)^2 \geq \mathcal{I}(A)$ when A is closed in $(E, \|\cdot\|)$ and that $\frac{1}{2}\, s(A)^2 \leq \mathcal{I}(A)$ when A is open. These inequalities correspond to the choice of a ball $B_E(0, \eta)$ as an element of $\mathcal{V}$ in the definitions of $r(A)$ and $s(A)$ (cf. also the previous elementary proof of the classical large deviation principle). Let us briefly verify this claim. Assume first that A is closed and let r be such that $0 < r < \mathcal{I}(A)$ (there is nothing to prove if $\mathcal{I}(A) = 0$). Then $A \cap r\mathcal{K} = \emptyset$ and since A is closed and $\mathcal{K}$ is compact in E, there exists $\eta > 0$ such that $A \cap (r\mathcal{K} + B_E(0, \eta))$ is still empty. Now $B_E(0, \eta) \in \mathcal{V}$ so that $r(A) \geq \sqrt{2r}$. Since $r < \mathcal{I}(A)$ is arbitrary, the first assertion follows. When A is open, let h be in $A \cap \mathcal{H}$ (there is nothing to prove if there is no such h). There exists $\eta > 0$ such that $B_E(h, \eta) \subset A$ which means that

$$\big(A \ominus B_E(0, \eta)\big) \cap \big(|h|\mathcal{K}\big) \neq \emptyset.$$

Therefore, $s(A) \leq |h|$ and since h is arbitrary in $A \cap \mathcal{H}$, $\frac{1}{2}s(A)^2 \leq \mathcal{I}(A)$. It should be noticed that the compactness of $\mathcal{K}$ is only used in the argument concerning the functional $r(\cdot)$. One may also note that if we restrict (without loss of generality) the class $\mathcal{V}$ to those elements V for which $0 \in V$, then for any set A, $\frac{1}{2}r(A)^2 \leq \mathcal{I}(A) \leq \frac{1}{2}s(A)^2$. In particular, $\frac{1}{2}r(A)^2$ (respectively $\frac{1}{2}s(A)^2$) coincide with $\mathcal{I}(A)$ if A is closed (respectively open).

The next theorem [BA-L1] is the main result concerning the measurable large deviation principle. The proof of the upper bound is entirely similar to the preceeding sketch of proof of the classical large deviation theorem. The lower bound amounts to the classical argument based on Cameron-Martin translates. Recall that the Cameron-Martin translation formula [C-M] (cf. [Ne1], [Ku], [Fe5], [Lif3]...) indicates that, for any h in $\mathcal{H}$, the probability measure $\mu(h + \cdot)$ is absolutely continuous with respect to μ with density given by the formula

$$\mu(h + A) = \exp\left(-\frac{|h|^2}{2}\right) \int_A \exp\big(-\widetilde{h}\big) d\mu \tag{4.11}$$

for every Borel set A in E (where we recall that $\widetilde{h} = ({j^*}_{|E_2^*})^{-1}(h)$).

Theorem 4.5. *For every Borel set A in E,*

$$\limsup_{\varepsilon \to 0} \varepsilon^2 \log \mu_\varepsilon(A) \leq -\tfrac{1}{2} r(A)^2 \tag{4.12}$$

and

$$\liminf_{\varepsilon \to 0} \varepsilon^2 \log \mu_\varepsilon(A) \geq -\tfrac{1}{2} s(A)^2. \tag{4.13}$$

By the preceding comments, this result generalizes the classical large deviations theorem for the Gaussian measure μ (due to M. Schilder [Sc] for Wiener measure and to M. Donsker and S. R. S. Varadhan [D-V] in general – see e.g. [Az], [D-S], [Var]...) which expresses that

$$\limsup_{\varepsilon \to 0} \varepsilon^2 \log \mu_\varepsilon(A) \leq -\mathcal{I}(\bar{A}), \tag{4.14}$$

where $\bar{A}$ is the closure of A (in $(E, \|\cdot\|)$) and

$$\liminf_{\varepsilon\to 0} \varepsilon^2 \log \mu_\varepsilon(A) \geq -\mathcal{I}(\mathring{A}) \tag{4.15}$$

where $\mathring{A}$ is the interior of A. It is rather easy to find examples of sets A such that $\frac{1}{2}r(A)^2 > \mathcal{I}(\bar{A})$ and $\frac{1}{2}s(A)^2 < \mathcal{I}(\mathring{A})$. (For example, if we fix the uniform topology on Wiener space, and if $A = \{x; \|x\|_\alpha \geq 1\}$ where $\|\cdot\|_\alpha$ is the Hölder norm of index α, then $r(A) > 0$ but $\mathcal{I}(\bar{A}) = 0$. (In this case of course, one can simply consider Wiener measure on the corresponding Hölder space.) More significant examples are described in [B-BA-K].) Therefore, Theorem 4.5 improves upon the classical large deviations for Gaussian measures.

Proof of Theorem 4.5. We start with (4.12). Let $r \geq 0$ be such that $(V + r\mathcal{K}) \cap A = \emptyset$ for some V in $\mathcal{V}$. Then

$$\mu_\varepsilon(A) = \mu(\varepsilon^{-1}A) \leq 1 - \mu_*\big(\varepsilon^{-1}V + \varepsilon^{-1}r\mathcal{K}\big).$$

Since $V \in \mathcal{V}$, there exists $\alpha > 0$ such that $\mu(\varepsilon^{-1}V) \geq \alpha$ for every $\varepsilon > 0$ small enough. Hence, by (4.6) (or Theorem 4.3),

$$\mu_\varepsilon(A) \leq \exp\bigg(-\frac{r^2}{2\varepsilon^2} + \frac{r}{\varepsilon}\,\delta(\alpha)\bigg)$$

from which (4.12) immediately follows in the limit.

As announced, the proof of (4.13) is classical. Let $s \geq 0$ be such that

$$(A \ominus V) \cap (s\mathcal{K}) \neq \emptyset$$

for some V in $\mathcal{V}$. Therefore, there exists h in $\mathcal{H}$ with $|h| \leq s$ such that $h + V \subset A$. Hence, for every $\varepsilon > 0$,

$$\mu_\varepsilon(A) = \mu(\varepsilon^{-1}A) \geq \mu\big(\varepsilon^{-1}(h + V)\big).$$

By Cameron-Martin's formula (4.11) (one could also use (1.10) in this argument),

$$\mu\big(\varepsilon^{-1}(h+V)\big) = \exp\bigg(-\frac{|h|^2}{2\varepsilon^2}\bigg) \int_{\varepsilon^{-1}V} \exp\bigg(-\frac{\widetilde{h}}{\varepsilon}\bigg) d\mu.$$

Since $V \in \mathcal{V}$, there exists $\alpha > 0$ such that $\mu(\varepsilon^{-1}V) \geq \alpha$ for every $\varepsilon > 0$ small enough. By Jensen's inequality,

$$\int_{\varepsilon^{-1}V} \exp\bigg(-\frac{\widetilde{h}}{\varepsilon}\bigg) d\mu \geq \mu(\varepsilon^{-1}V) \exp\bigg(-\int_{\varepsilon^{-1}V} \frac{\widetilde{h}}{\varepsilon} \cdot \frac{d\mu}{\mu(\varepsilon^{-1}V)}\bigg).$$

Now,

$$\int_{\varepsilon^{-1}V} \widetilde{h}\, d\mu \leq \int |\widetilde{h}| d\mu \leq \bigg(\int \widetilde{h}^2 d\mu\bigg)^{1/2} = |h|.$$

We have thus obtained that, for every $\varepsilon > 0$ small enough,

$$\mu_\varepsilon(A) \geq \mu\big(\varepsilon^{-1}(h+V)\big) \geq \alpha \exp\bigg(-\frac{|h|^2}{2\varepsilon^2} - \frac{|h|}{\alpha\varepsilon}\bigg)$$

from which we deduce that

$$\liminf_{\varepsilon\to 0} \varepsilon^2 \log \mu_\varepsilon(A) \geq -\frac{1}{2}\,|h|^2 \geq -\frac{1}{2}\,s^2.$$

The claim (4.13) follows since s may be chosen arbitrary less than $s(A)$. The proof of Theorem 4.5 is complete. □

It is a classical result in the theory of large deviations, due to S. R. S. Varadhan (cf. [Az], [D-S], [Var]...), that the statements (4.14) and (4.15) on sets may be translated essentially equivalently on functions. More precisely, if $F : E \to \mathbb{R}$ is bounded and continuous on E,

$$\lim_{\varepsilon\to 0} \varepsilon^2 \log\bigg(\int \exp\Big(-\frac{1}{\varepsilon^2}\,F(\varepsilon x)\Big)d\mu(x)\bigg) = -\inf_{x\in E}\big(F(x)+\mathcal{I}(x)\big).$$

One consequence of measurable large deviations is that it allows us to weaken the continuity hypothesis into a continuity "in probability".

Corollary 4.6. *Let $F : E \to \mathbb{R}$ be measurable and bounded on E and such that, for every $r > 0$ and every $\eta > 0$,*

$$\limsup_{\varepsilon\to 0} \mu\Big(x;\ \sup_{|h|\leq r}\big|F(h+\varepsilon x) - F(h)\big| > \eta\Big) < 1.$$

Then,

$$\lim_{\varepsilon\to 0} \varepsilon^2 \log\bigg(\int \exp\Big(-\frac{1}{\varepsilon^2}\,F(\varepsilon x)\Big)d\mu(x)\bigg) = -\inf_{x\in E}\big(F(x)+\mathcal{I}(x)\big).$$

It has to be mentioned that the continuity assumption in Corollary 4.6 is not of the Malliavin calculus type since limits are taken along the elements of E and not the elements of $\mathcal{H}$.

Proof. Set

$$L(\varepsilon) = \int \exp\Big(-\frac{1}{\varepsilon^2}\,F(\varepsilon x)\Big)d\mu(x), \quad \varepsilon > 0.$$

By a simple translation, we may assume that $F \geq 0$. For simplicity in the notation, let us assume moreover that $0 \leq F \leq 1$. For every integer $n \geq 1$, set

$$A_k^n = \big\{\tfrac{k-1}{n} < F \leq \tfrac{k}{n}\big\}, \quad k = 2,\ldots,n, \quad A_1^n = \big\{F \leq \tfrac{1}{n}\big\}.$$

Since

$$L(\varepsilon) \leq \sum_{k=1}^n \exp\Big(-\frac{k-1}{\varepsilon^2 n}\Big)\mu^\varepsilon\big(A_k^n\big),$$

we get that

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log L(\varepsilon) \le -\min_k \Big(\frac{k-1}{n} + \frac{1}{2}\, r\big(A_k^n\big)^2 \Big).$$

Since $r(A_k^n) \ge r(\{F \le \frac{k}{n}\})$ and since n is arbitrary, it follows that

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log L(\varepsilon) \le -\inf_{t\in\mathbb{R}} \Big(t + \frac{1}{2}\, r\big(\{F \le t\}\big)^2 \Big). \tag{4.16}$$

Now, we show that the right hand side of (4.16) is less than or equal to

$$-\inf_{x\in E} (F(x) + \mathcal{I}(x)).$$

Let $t \in \mathbb{R}$ be fixed. Let $\eta > 0$ and $r > r(\{F \le t\})$ (assumed to be finite). Set

$$V = \big\{ x;\ \sup_{|h|\le r} \big| F(h+x) - F(h) \big| \le \eta \big\}.$$

By the hypothesis, $V \in \mathcal{V}$. By the definition of r,

$$\big(V + r\mathcal{K}\big) \cap \{F \le t\} \ne \emptyset.$$

Therefore, there exist v in V and $|h| \le r$ such that $F(h+v) \le t$. By definition of V, $F(h) \le t + \eta$. Hence

$$\inf_{x\in E} \big(F(x) + \mathcal{I}(x) \big) \le t + \eta + \frac{r^2}{2}.$$

Since $\eta > 0$ and $r > r(\{F \le t\})$ are arbitrary, the claim follows and thus, together with (4.16),

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log L(\varepsilon) \le -\inf_{x\in E} \big(F(x) + \mathcal{I}(x) \big). \tag{4.17}$$

The proof of the lower bound is similar. We have, for every $n \ge 1$,

$$\begin{aligned} L(\varepsilon) &\ge \sum_{k=1}^{n} \exp\Big(-\frac{k}{\varepsilon^2 n} \Big) \mu^\varepsilon\big(A_k^n\big) \\ &\ge \sum_{k=1}^{n-1} \Big[\exp\Big(-\frac{k}{\varepsilon^2 n} \Big) - \exp\Big(-\frac{k+1}{\varepsilon^2 n} \Big) \Big] \mu^\varepsilon\big(\{F \le \tfrac{k}{n}\}\big) \\ &\ge \frac{1}{2} \sum_{k=1}^{n-1} \exp\Big(-\frac{k}{\varepsilon^2 n} \Big) \mu^\varepsilon\big(\{F \le \tfrac{k}{n}\}\big) \end{aligned}$$

at least for all $\varepsilon > 0$ small enough. Therefore,

$$\liminf_{\varepsilon\to 0} L(\varepsilon) \ge -\inf_{t\in\mathbb{R}} \Big(t + \frac{1}{2}\, s\big(\{F \le t\}\big)^2 \Big). \tag{4.18}$$

Now, let h be in $\mathcal{H}$ and set $t = F(h)$. Let $\eta > 0$ and $0 < s < s(\{F \le t + \eta\})$. We will show that $|h| > s$. Let

$$V = \{x; F(h + x) \le F(h) + \eta\}.$$

By the hypothesis, $V \in \mathcal{V}$ and by the definition of s,

$$(\{F \le t + \eta\} \ominus V) \cap (s\mathcal{K}) = \emptyset.$$

It is clear that $h \in \{F \le t + \eta\} \ominus V$. Hence $|h| > s$, and since s is arbitrary, we have $|h| \ge s(\{F \le t + \eta\})$. Now, if $t > -\infty$,

$$t + \frac{1}{2} s(\{F \le t + \eta\})^2 \le F(h) + \mathcal{I}(h).$$

If $t = -\infty$, $0 \le s(\{F = -\infty\}) \le |h| < \infty$, and the preceding also holds. In any case,

$$-\inf_{t \in \mathbb{R}} \Big(t + \frac{1}{2}\, s(\{F \le t\})^2\Big) \le \inf_{x \in E} \big(F(x) + \mathcal{I}(x)\big).$$

Together with (4.18) and (4.17), the proof of Corollary 4.6 is complete. □

In the last part of this chapter, we prove a large deviation principle for the Ornstein-Uhlenbeck process due to S. Kusuoka [Kus]. If μ is a Gaussian measure on E, define, for every say bounded measurable function f on E, and every $x \in E$ and $t \ge 0$,

$$P_t f(x) = \int_E f\big(\mathrm{e}^{-t} x + (1 - \mathrm{e}^{-2t})^{1/2} y\big) d\mu(y).$$

If A and B are Borel subsets of E, set then, as at the end of Chapter 2,

$$K_t(A, B) = \int_A P_t(I_B) d\mu, \quad t \ge 0.$$

We will be interested in the large deviation behavior of $K_t(A, B)$ in terms of the $\mathcal{H}$-distance between A and B. Set indeed

$$d_{\mathcal{H}}(A, B) = \inf\{|h - k|; h \in A, k \in B, h - k \in \mathcal{H}\}.$$

One defines in the same way $d_{\mathcal{H}}(x, A)$, $x \in E$, and notices that $d_{\mathcal{H}}(x, A) < \infty$ μ-almost everywhere if and only if $\mu(A + \mathcal{H}) = 1$. By the isoperimetric inequality, this is immediately the case as soon as $\mu(A) > 0$.

The main result is the following. S. Kusuoka's proof uses the wave equation. We folllow here the approach by S. Fang [Fa] (who actually establishes a somewhat stronger statement by using a slightly different distance on the subsets of E).

Theorem 4.7. *Let A and B be Borel subsets in E such that $\mu(A) > 0$ and $\mu(B) > 0$. Then*

$$\limsup_{t \to 0} 4t \log K_t(A, B) \le -d_{\mathcal{H}}(A, B)^2.$$

If moreover A and B are open, then

$$\liminf_{t\to 0} 4t \log K_t(A,B) \geq -d_{\mathcal{H}}(A,B)^2.$$

Proof. We start with the upper bound which is thus based on Proposition 2.3. We use the same approximation procedure as the one described in the proof of Theorem 4.3 and, accordingly the probabilistic notation put forward in Proposition 4.2. Denote in particular by Y an independent copy of X (with thus distribution μ). Assume that $d_{\mathcal{H}}(A,B) > r > 0$. Choose $K \subset A$ and $L \subset B$ compact subsets of positive measure. Let $t \geq 0$ be fixed and let $\varepsilon > 0$. We can write that, for every $n \geq n_0$ large enough,

$$\begin{aligned}
&\mathbb{P}\big\{X \in K, \mathrm{e}^{-t}X + (1-\mathrm{e}^{-2t})^{1/2}Y \notin K^{3\varepsilon} + r\mathcal{K}\big\} \\
&\qquad \leq \mathbb{P}\big\{X_n \in K^{\varepsilon}, \mathrm{e}^{-t}X_n + (1-\mathrm{e}^{-2t})^{1/2}Y_n \notin K^{2\varepsilon} + r\mathcal{K}\big\} + \varepsilon \\
&\qquad \leq \mathbb{P}\big\{X_n \in K^{\varepsilon}, \mathrm{e}^{-t}X_n + (1-\mathrm{e}^{-2t})^{1/2}Y_n \notin K^{\varepsilon} + r\mathcal{K}_n\big\} + \varepsilon.
\end{aligned}$$

Hence, according to (2.16) of Chapter 2,

$$\mathbb{P}\big\{X \in K, \mathrm{e}^{-t}X + (1-\mathrm{e}^{-2t})^{1/2}Y \notin K^{3\varepsilon} + r\mathcal{K}\big\} \leq \exp\Big(-\frac{r^2}{4(1-\mathrm{e}^{-t})}\Big) + \varepsilon.$$

Letting ε decrease to zero, by compactness,

$$\mathbb{P}\big\{X \in K, \mathrm{e}^{-t}X + (1-\mathrm{e}^{-2t})^{1/2}Y \notin K + r\mathcal{K}\big\} \leq \exp\Big(-\frac{r^2}{4(1-\mathrm{e}^{-t})}\Big)$$

and thus, by definition of r,

$$\mathbb{P}\big\{X \in K, \mathrm{e}^{-t}X + (1-\mathrm{e}^{-2t})^{1/2}Y \in L\big\} \leq \exp\Big(-\frac{r^2}{4(1-\mathrm{e}^{-t})}\Big).$$

The first claim of Theorem 4.7 is proved.

The lower bound relies on Cameron-Martin translates. Let $r = d_{\mathcal{H}}(A,B) \geq 0$ (assumed to be finite). Let also $h \in A \cap \mathcal{H}$ and $k \in B \cap \mathcal{H}$. Since A and B are open, there exists $\eta > 0$ such that $B_E(h, 2\eta) \subset A$ and $B_E(k, 2\eta) \subset B$. Therefore, for every $t \geq 0$,

$$K_t(A,B) \geq K_t\big(B_E(h,2\eta), B_E(k,2\eta)\big).$$

By the Cameron-Martin translation formula (4.11),

$$\begin{aligned}
&K_t\big(B_E(h,2\eta), B_E(k,2\eta)\big) \\
&\qquad = \exp\Big(-\frac{|h-k|^2}{2(1-\mathrm{e}^{-2t})}\Big) \int_{(x,y)\in C} \exp\Big(\frac{\widetilde{h}(y) - \widetilde{k}(y)}{(1-\mathrm{e}^{-2t})^{1/2}}\Big) d\mu(x) d\mu(y)
\end{aligned}$$

where $C = \{(x,y) \in E \times E; x \in B_E(h, 2\eta), \mathrm{e}^{-t}x + (1-\mathrm{e}^{-2t})^{1/2}y \in B_E(h, 2\eta)\}$. Now, for $t \leq t_0(\eta, h)$ small enough,

$$B_E(h,\eta) \times B_E(0,1) \subset C$$

so that

$$\begin{aligned} K_t(A,B) &\geq \exp\bigg(-\frac{|h-k|^2}{2(1-\mathrm{e}^{-2t})}\bigg)\mu\big(B_E(h,\eta)\big)\int_{B_E(0,1)} \exp\bigg(\frac{\widetilde{h}-\widetilde{k}}{(1-\mathrm{e}^{-2t})^{1/2}}\bigg)d\mu \\ &\geq \exp\bigg(-\frac{|h-k|^2}{2(1-\mathrm{e}^{-2t})}\bigg)\mu\big(B_E(h,\eta)\big)\mu\big(B_E(0,1)\big) \end{aligned}$$

by Jensen's inequality. Therefore,

$$\liminf_{t\to 0} 4t \log K_t(A,B) \geq -|h-k|^2$$

and the result follows since h and k are arbitrary in A and B respectively. The proof of Theorem 4.7 is complete. □

Notes for further reading. There is an extensive literature on precise estimates on the tail behavior of norms of Gaussian random vectors (involving in particular the tool of entropy – cf. Chapter 6). We refer in particular the interested reader to the works [Ta4], [Ta13], [Lif2] and the references therein (see also [Lif3]). In the paper [Lif2], a Laplace method is developed to yield some unexpected irregular behaviors. Large deviations without topology may be applied to Strassen's law of the iterated logarithm for Brownian motion [B-BA-K], [BA-L1], [D-L]. In [D-L], a complete description of norms on Wiener space for which the law of the iterated logarithm holds is provided.

5. LARGE DEVIATIONS OF WIENER CHAOS

The purpose of this chapter is to further demonstrate the usefulness and interest of isoperimetric and concentration methods in large deviation theorems in the context of Wiener chaos. This chapter intends actually to present some aspects of the remarkable work of C. Borell on homogeneous chaos whose early ideas strongly influenced the subsequent developments. We present here, closely following the material in [Bo5], [Bo9], a simple isoperimetric proof of the large deviations properties of homogeneous Gaussian chaos (even vector valued). We take again the exposition of [Led2].

Let, as in the preceding chapter, $(E, \mathcal{H}, \mu)$ be an abstract Wiener space. According to Proposition 4.2, for any orthonormal basis $(g_i)_{i\in\mathbb{N}}$ of the closure E_2^* of E^* in $\mathrm{L}^2(\mu)$, μ has the same distribution as the series $\sum_i g_i j^*(g_i)$. It will be convenient here (although this is not strictly necessary) to consider this basis in E^*. Let thus $(\xi_i)_{i\in\mathbb{N}} \subset E^*$ be any fixed orthonormal basis of E_2^* (take any weak-star dense sequence of the unit ball of E^* and orthonormalize it with respect to μ using the Gram-Schmidt procedure). Denote by $(h_k)_{k\in\mathbb{N}}$ the sequence of the Hermite polynomials defined from the generating series

$$\mathrm{e}^{\lambda x - \lambda^2/2} = \sum_{k=0}^{\infty} \lambda^k h_k(x), \quad \lambda, x \in \mathbb{R}.$$

$(\sqrt{k!}\, h_k)_{k\in\mathbb{N}}$ is an orthonormal basis of $\mathrm{L}^2(\gamma_1)$ where γ_1 is the canonical Gaussian measure on $\mathbb{R}$. If $\alpha = (\alpha_0, \alpha_1, \dots) \in \mathbb{N}^{(\mathbb{N})}$, i.e. $|\alpha| = \alpha_0 + \alpha_1 + \cdots < \infty$, set

$$H_\alpha = \sqrt{\alpha!} \prod_i h_{\alpha_i} \circ \xi_i$$

(where $\alpha! = \alpha_0!\alpha_1! \cdots$). Then the family (H_α) constitutes an orthonormal basis of $L^2(\mu)$.

Let now B be a real separable Banach space with norm $\|\cdot\|$ (we denote in the same way the norm on E and the norm on B). $\mathrm{L}^p((E, \mathcal{B}, \mu); B) = \mathrm{L}^p(\mu; B)$

$(0 \le p < \infty)$ is the space of all Bochner measurable functions F on (E, μ) with values in B $(p = 0)$ such that $\int \|F\|^p d\mu < \infty$ $(0 < p < \infty)$. For each integer $d \ge 1$, set

$$\mathcal{W}^{(d)}(\mu; B) = \big\{F \in \mathrm{L}^2(\mu; B); \langle F, H_\alpha \rangle = \int F H_\alpha d\mu = 0 \text{ for all } \alpha \text{ such that } |\alpha| \neq d\big\}.$$

$\mathcal{W}^{(d)}(\mu; B)$ defines the B-valued homogeneous Wiener chaos of degree d [Wi]. An element Ψ of $\mathcal{W}^{(d)}(\mu; B)$ can be written as

$$\Psi = \sum_{|\alpha| = d} \langle \Psi, H_\alpha \rangle \, H_\alpha$$

where the multiple sum is convergent (for any finite filtering) μ-almost everywhere and in $\mathrm{L}^2(\mu; B)$. (Actually, as a consequence of [Bo5], [Bo9] (see also [L-T2], or the subsequent main result), this convergence also takes place in $\mathrm{L}^p(\mu; B)$ for any p.) To see it, we simply follow the proof of Proposition 4.2. Let, for each n, $\mathcal{B}_n$ be the sub-σ-algebra of $\mathcal{B}$ generated by the functions $\xi_0, \ldots, \xi_n$ on E and let Ψ_n be the conditional expectation of Ψ with respect to $\mathcal{B}_n$. Recall that $\mathcal{B}$ may be assumed to be generated by $(\xi_i)_{i \in \mathbb{N}}$. Then

$$\Psi_n = \sum_{\substack{|\alpha| = d \\ \alpha_k = 0, k > n}} \langle \Psi, H_\alpha \rangle \, H_\alpha \tag{5.1}$$

as can be checked on linear functionals, and therefore, by the vector valued martingale convergence theorem (cf. [Ne2]), the claim follows. One could actually take this series representation as the definition of a homogeneous chaos, which would avoid the assumption $F \in \mathrm{L}^2(\mu; B)$ in $\mathcal{W}^{(d)}(\mu; B)$. By the preceding comment, both definitions actually agree (cf. [Bo5], [Bo9]).

As a consequence of the Cameron-Martin formula, we may define for every F in $L^0(\mu; B)$ and every h in $\mathcal{H}$, a new element $F(\cdot + h)$ of $\mathrm{L}^0(\mu; B)$. Furthermore, if F is in $\mathrm{L}^2(\mu; B)$, for any $h \in \mathcal{H}$,

$$\int \big\|F(x+h)\big\| \, d\mu(x) \le \exp\bigg(\frac{|h|^2}{2}\bigg) \bigg(\int \big\|F(x)\big\|^2 d\mu(x)\bigg)^{1/2}. \tag{5.2}$$

Indeed,

$$\int \big\|F(x+h)\big\| \, d\mu(x) = \int \exp\bigg(-\widetilde{h}(x) - \frac{|h|^2}{2}\bigg) \, \big\|F(x)\big\| d\mu(x)$$

from which (5.2) follows by Cauchy-Schwarz inequality and the fact that $\widetilde{h} = (j^*_{|E_2^*})^{-1}(h)$ is Gaussian with variance $|h|^2$.

Let F be in $\mathrm{L}^2(\mu; B)$. By (5.2), for any h in $\mathcal{H}$, we can define an element $F^{(d)}(h)$ of B by setting

$$F^{(d)}(h) = \int F(x+h) d\mu(x).$$

If $\Psi \in \mathcal{W}^{(d)}(\mu;B)$, $\Psi^{(d)}(h)$ is homogeneous of degree d. To see it, we can work by approximation on the Ψ_n's and use then the easy fact (checked on the generating series for example) that, for any real number λ and any integer k,

$$\int h_k(x+\lambda)d\gamma_1(x) = \frac{1}{k!}\,\lambda^k.$$

Actually, $\Psi^{(d)}(h)$ can be written as the convergent multiple sum

$$\Psi^{(d)}(h) = \sum_{|\alpha|=d} \frac{1}{\alpha!}\,\langle \Psi, H_\alpha\rangle\, h^\alpha$$

where h^α is meant as $\langle \xi_0, h\rangle^{\alpha_0}\langle \xi_1, h\rangle^{\alpha_1}\cdots$.

Given thus Ψ in $\mathcal{W}^{(d)}(\mu;B)$, for any s in B, set $\mathcal{I}_\Psi(s) = \inf\{\frac{1}{2}|h|^2; s = \Psi^{(d)}(h)\}$ if there exists h in $\mathcal{H}$ such that $s = \Psi^{(d)}(h)$, $\mathcal{I}_\Psi(s) = \infty$ otherwise. For a subset A of B, set $\mathcal{I}_\Psi(A) = \inf_{s\in A}\mathcal{I}_\Psi(s)$.

We can now state the large deviation properties for the elements Ψ of $\mathcal{W}^{(d)}(\mu;B)$. The case $d=1$ of course corresponds to the classical large deviation result for Gaussian measures (cf. (4.14) and (4.15) for $B=E$ and Ψ the identity map on E). From the point of view of isoperimetry and concentration, the proof for higher order chaos is actually only the appropriate extension of the case $d=1$.

Theorem 5.1. *Let $\mu_\varepsilon(\cdot) = \mu(\varepsilon^{-1}(\cdot))$, $\varepsilon > 0$. Let d be an integer and let Ψ be an element of $\mathcal{W}^{(d)}(\mu;B)$. Then, if A is a closed subset of B,*

$$\limsup_{\varepsilon\to 0}\varepsilon^2\log\mu_\varepsilon\big(x;\Psi(x)\in A\big) \le -\mathcal{I}_\Psi(A). \tag{5.3}$$

If A is an open subset of B,

$$\liminf_{\varepsilon\to 0}\varepsilon^2\log\mu_\varepsilon\big(x;\Psi(x)\in A\big) \ge -\mathcal{I}_\Psi(A). \tag{5.4}$$

The proof of (5.4) follows rather easily from the Cameron-Martin translation formula. (5.3) is rather easy too, but our approach thus rests on the tool of isoperimetric and concentration inequalities. The proof of (5.3) also sheds some light on the structure of Gaussian polynomials as developed by C. Borell, and in particular the homogeneous structures. As it is clear indeed from [Bo5] (and the proof below), the theorem may be shown to hold for all Gaussian polynomials, i.e. elements of the closure in $\mathrm{L}^0(\mu;B)$ of all continuous polynomials from E into B of degree less than or equal to d. As we will see, $\mathcal{W}^{(d)}(\mu;B)$ may be considered as a subspace of the closure of all homogeneous Gaussian polynomials of degree d (at least if the support of μ is infinite dimensional), and hence, the elements of $\mathcal{W}^{(d)}(\mu;B)$ are μ-almost everywhere d-homogeneous. In particular, (5.3) and (5.4) of the theorem are equivalent to saying that (changing moreover ε into t^{-1})

$$\limsup_{t\to\infty}\frac{1}{t^2}\log\mu\big(x;\Psi(x)\in t^dA\big) \le -\mathcal{I}_\Psi(A) \tag{5.5}$$

(A closed) and

$$\liminf_{t\to\infty} \frac{1}{t^2} \log \mu\big(x; \Psi(x) \in t^d A\big) \geq -\mathcal{I}_\Psi(A), \tag{5.6}$$

(A open) and these are the properties we will actually establish.

Before turning to the proof of Theorem 5.1, let us mention some application. If we take A in the theorem to be the complement U^c of the (open or closed) unit ball U of B, one immediately checks that

$$\mathcal{I}_\Psi(U^c) = \frac{1}{2}\big(\sup_{h\in\mathcal{K}} \|\Psi^{(d)}(h)\|\big)^{-2/d}.$$

We may therefore state the following corollary of Theorem 5.1 which was actually established directly from the isoperimetric inequality by C. Borell [Bo5] (see also [Bo8], [L-T2]). It is the analogue for chaos of Theorem 4.1.

Corollary 5.2. *Let Ψ be an element of $\mathcal{W}^{(d)}(\mu; B)$. Then*

$$\lim_{t\to\infty} \frac{1}{t^{2/d}} \log \mu\big(x; \|\Psi(x)\| \geq t\big) = -\frac{1}{2}\big(\sup_{h\in\mathcal{K}} \|\Psi^{(d)}(h)\|\big)^{-2/d}.$$

As in Theorem 4.1, we have that

$$\int \exp\big(\alpha\|\Psi\|^{2/d}\big) d\mu < \infty \quad \text{if and only if} \quad \alpha < \frac{1}{2}\big(\sup_{h\in\mathcal{K}} \|\Psi^{(d)}(h)\|\big)^{-2/d}.$$

Furthermore, the proof of the theorem will show that all moments of Ψ are equivalent (see also Chapter 8, (8.23)).

In the setting of the classical Wiener space $E = C_0([0,1])$ equipped with the Wiener measure μ, and when $B = E$, K. Itô [It] (see also [Ne1] and the recent approach [Str]) identified the elements Ψ of $\mathcal{W}^{(d)}(\mu; E)$ with the multiple stochastic integrals

$$\Psi = \left(\int_0^t \int_0^{t_1} \cdots \int_0^{t_{d-1}} k(t_1, \ldots, t_d)\, dW(t_1) \cdots dW(t_d)\right)_{t\in[0,1]}$$

where k deterministic is such that

$$\int_0^1 \int_0^{t_1} \cdots \int_0^{t_{d-1}} k(t_1, \ldots, t_d)^2 dt_1 \cdots dt_d < \infty.$$

If h belongs to the reproducing kernel Hilbert space of the Wiener measure, then

$$\Psi^{(d)}(h) = \left(\int_0^t \int_0^{t_1} \cdots \int_0^{t_{d-1}} k(t_1, \ldots, t_d)\, h'(t_1) \cdots h'(t_d)\, dt_1 \cdots dt_d\right)_{t\in[0,1]}.$$

Proof of Theorem 5.1. Let us start with the simpler property (5.4). Recall Ψ_n from (5.1). We can write (explicitly on the Hermite polynomials), for all x in E, h in $\mathcal{H}$ and t real number,

$$\Psi_n(x+th) = \sum_{k=0}^{d} t^k \Psi_n^{(k)}(x,h).$$

If $P(t) = a_0 + a_1 t + \cdots + a_d t^d$ is a polynomial of degree d in $t \in \mathbb{R}$ with vector coefficients $a_0, a_1, \ldots, a_d$, there exist real constants $c(i,k,d)$, $0 \le i, k \le d$, independent of P, such that, for every $k = 0, \ldots, d$,

$$a_k = c(0,k,d)P(0) + \sum_{i=1}^{d} c(i,k,d)P(2^{i-1}).$$

Hence, for every $h \in \mathcal{H}$,

$$\Psi_n^{(k)}(\cdot,h) = c(0,k,d)\Psi_n(\cdot) + \sum_{i=1}^{d} c(i,k,d)\Psi_n(\cdot + 2^{i-1}h)$$

from which we deduce together with (5.2) that, for every $k = 0, \ldots, d$,

$$\int \big\| \Psi_n^{(k)}(x,h) \big\| \, d\mu(x) \le C(k,d;h) \left(\int \big\| \Psi_n(x) \big\|^2 d\mu(x) \right)^{1/2}$$

for some constants $C(k,d;h)$ thus only depending on k, d and $h \in \mathcal{H}$. In the limit, we conclude that there exist, for every h in $\mathcal{H}$ and $k = 0, \ldots, d$, elements $\Psi^{(k)}(\cdot,h)$ of $\mathrm{L}^1(\mu; B)$ such that

$$\Psi(\cdot + th) = \sum_{k=0}^{d} t^k \Psi^{(k)}(\cdot,h)$$

for every $t \in \mathbb{R}$, with

$$\int \big\| \Psi^{(k)}(x,h) \big\| \, d\mu(x) \le C(k,d;h) \left(\int \big\| \Psi(x) \big\|^2 d\mu(x) \right)^{1/2}$$

and $\Psi^{(0)}(\cdot,h) = f(\cdot)$, $\Psi^{(d)}(\cdot,h) = \Psi^{(d)}(h)$ (since $\int f(x+th)\, d\mu(x) = t^d f^{(d)}(h)$). As a main consequence, we get that, for every h in $\mathcal{H}$,

$$\lim_{t\to\infty} \frac{1}{t^d} \int \big\| \Psi(x+th) - t^d \Psi^{(d)}(h) \big\| d\mu(x) = 0. \tag{5.7}$$

This limit can be made uniform in $h \in \mathcal{K}$ but we will not use this observation in this form later (that is in the proof of (5.3); we use instead a stronger property, (5.9) below).

To establish (5.4), let A be open in B and let $s = \Psi^{(d)}(h)$, $h \in \mathcal{H}$, belong to A (if no such s exists, then $\mathcal{I}_\Psi(A) = \infty$ and (5.4) then holds trivially). Since A is open, there is $\eta > 0$ such that the ball $B(s,\eta)$ in B with center s and radius η is contained in A. Therefore, if $V = V(t) = \{x \in E; \Psi(x) \in t^d B(s,\eta)\}$, by the Cameron-Martin translation formula (4.11),

$$\mu\big(x; \Psi(x) \in t^d A\big) \geq \mu(V) = \int_{V-th} \exp\bigg(t\widetilde{h} - \frac{t^2|h|^2}{2}\bigg) d\mu.$$

Furthermore, by Jensen's inequality,

$$\mu(V) \geq \exp\bigg(-\frac{t^2|h|^2}{2}\bigg)\mu(V-th) \exp\bigg(\frac{t}{\mu(V-th)} \int_{V-th} \widetilde{h} d\mu\bigg).$$

By (5.7),

$$\mu(V-th) = \mu\big(x; \big\|\Psi(x+th) - t^d \Psi^{(d)}(h)\big\| \leq \eta t^d\big) \geq \frac{1}{2}$$

for all $t \geq t_0$ large enough. We have

$$\int_{V-th} \widetilde{h} d\mu \geq -\int |\widetilde{h}| d\mu \geq -\bigg(\int \widetilde{h}^2 d\mu\bigg)^{1/2} = -|h|.$$

Thus, for all $t \geq t_0$,

$$\frac{t}{\mu(V-th)} \int_{V-th} \widetilde{h} d\mu \geq -2t|h|,$$

and hence, summarizing,

$$\mu\big(x; \Psi(x) \in t^d A\big) \geq \frac{1}{2} \exp\bigg(-\frac{t^2|h|^2}{2} - 2t|h|\bigg).$$

It follows that

$$\liminf_{t\to\infty} \frac{1}{t^2} \log \mu\big(x; \Psi(x) \in t^d A\big) \geq -\frac{1}{2}\,|h|^2 = -\mathcal{I}_\Psi(s)$$

and since s is arbitrary in A, property (5.6) is satisfied. As a consequence of what we will develop now, (5.4) will be satisfied as well.

Now, we turn to (5.3) and in the first part of this investigation, we closely follow C. Borell [Bo5], [Bo9]. We start by showing that every element Ψ of $\mathcal{W}^{(d)}(\mu; B)$ is limit (at least if the dimension of the support of μ is infinite), μ-almost everywhere and in $\mathrm{L}^2(\mu; B)$, of a sequence of d-homogeneous polynomials. In particular, Ψ is μ-almost everywhere d-homogeneous justifying therefore the equivalences between (5.3) and (5.4) and respectively (5.5) and (5.6). Assume thus in the following that μ is infinite dimensional. We can actually always reduce to this case by appropriately tensorizing μ, for example with the canonical Gaussian measure on $\mathbb{R}^{\mathbb{N}}$. Recall that Ψ is limit almost surely and in $\mathrm{L}^2(\mu; B)$ of the Ψ_n's of (5.1). The finite sums Ψ_n can be decomposed into their homogeneous components as

$$\Psi_n = \Psi_n^{(d)} + \Psi_n^{(d-2)} + \cdots,$$

where, for any x in E,

$$\Psi_n^{(k)}(x) = \sum_{i_1,\ldots,i_k=0}^{\infty} b_{i_1,\ldots,i_k} \langle \xi_{i_1}, x\rangle\langle \xi_{i_2}, x\rangle \cdots \langle \xi_{i_k}, x\rangle \tag{5.8}$$

with only finitely many $b_{i_1,\ldots,i_k}$ in B nonzero. The main observation is that the constant 1 is limit of homogeneous polynomials of degree 2: indeed, simply take by the law of large numbers

$$p_n(x) = \frac{1}{n+1}\sum_{k=0}^{n}\langle \xi_k, x\rangle^2.$$

Since p_n and $\Psi_n^{(k)}$ belong to $\mathrm{L}^p(\mu)$ and $\mathrm{L}^p(\mu; B)$ respectively for every p, and since $p_n - 1$ tends there to 0, it is easily seen that there exists a subsequence m_n of the integers such that $(p_{m_n} - 1)(\Psi_n^{(d-2)} + \Psi_n^{(d-4)} + \cdots)$ converges to 0 in $\mathrm{L}^2(\mu; B)$. This means that Ψ is the limit in $\mathrm{L}^2(\mu; B)$ of $\Psi_n^{(d)} + p_{m_n}(\Psi_n^{(d-2)} + \Psi_n^{(d-4)} + \cdots)$, that is limit of a sequence of polynomials Ψ'_n whose decomposition in homogeneous polynomials

$$\Psi'_n = {\Psi'_n}^{(d)} + {\Psi'_n}^{(d-2)} + \cdots$$

is such that ${\Psi'_n}^{(1)}$, or ${\Psi'_n}^{(0)}$ and ${\Psi'_n}^{(2)}$, according as d is odd or even, can be taken to be 0. Repeating this procedure, Ψ is indeed seen to be the limit in $\mathrm{L}^2(\mu; B)$ of a sequence (Ψ'_n) of d-homogeneous polynomials (i.e. polynomials of the type (5.8)).

The important property in order to establish (5.5) is the following. It improves upon (5.7) and claims that, in the preceding notations, i.e. if Ψ is limit of the sequence (Ψ'_n) of d-homogeneous polynomials,

$$\lim_{t\to\infty} \frac{1}{t^{2d}} \sup_n \int \sup_{h\in\mathcal{K}} \big\|\Psi'_n(x+th) - t^d\Psi'_n(h)\big\|^2 \, d\mu(x) = 0 \tag{5.9}$$

where we recall that $\mathcal{K}$ is the unit ball of the reproducing kernel Hilbert space $\mathcal{H}$ of μ. To establish this property, given

$$\Psi'_n(x) = \sum_{i_1,\ldots,i_d=0}^{\infty} b^n_{i_1,\ldots,i_d} \langle \xi_{i_1}, x\rangle\langle \xi_{i_2}, x\rangle \cdots \langle \xi_{i_d}, x\rangle$$

(with only finitely many $b^n_{i_1,\ldots,i_d}$ nonzero), let us consider the (unique) multilinear symmetric polynomial $\widehat{\Psi}'_n$ on E^d such that $\widehat{\Psi}'_n(x,\ldots,x) = \Psi'_n(x)$; $\widehat{\Psi}'_n$ is given by

$$\widehat{\Psi}'_n(x_1,\ldots,x_d) = \sum_{i_1,\ldots,i_d=0}^{\infty} \widehat{b}^n_{i_1,\ldots,i_d} \langle \xi_{i_1}, x_1\rangle \cdots \langle \xi_{i_d}, x_d\rangle, \quad x_1,\ldots,x_d \in E,$$

where

$$\widehat{b}^n_{i_1,\ldots,i_d} = \frac{1}{d!}\sum_{\sigma} b^n_{\sigma(i_1),\ldots,\sigma(i_d)},$$

the sum running over all permutations σ of $\{1,\ldots,d\}$. We use the following polarization formula: letting $\varepsilon_1,\ldots,\varepsilon_d$ be independent random variables taking values ± 1 with probability $\frac{1}{2}$ and denoting by $\mathbb{E}$ expectation with respect to them,

$$\widehat{\Psi}'_n(x_1,\ldots,x_d) = \frac{1}{d!}\,\mathbb{E}\big(\Psi'_n(\varepsilon_1 x_1+\cdots+\varepsilon_d x_d)\,\varepsilon_1\cdots\varepsilon_d\big). \tag{5.10}$$

We adopt the notation $x^{d-k}y^k$ for the element $(x,\ldots,x,y,\ldots,y)$ in E^d where x is repeated $(d-k)$-times and y k-times. Then, for any x,y in E, we have

$$\Psi'_n(x+y) = \sum_{k=0}^{d}\binom{d}{k}\widehat{\Psi}'_n\big(x^{d-k}y^k\big). \tag{5.11}$$

To establish (5.9), we see from (5.11) that it suffices to show that for all $k = 1,\ldots,d-1$,

$$\sup_n \int \sup_{h\in\mathcal{K}}\big\|\widehat{\Psi}'_n(x^{d-k}h^k)\big\|^2 d\mu(x) < \infty. \tag{5.12}$$

Let k be fixed. By orthogonality,

$$\begin{aligned}
&\sup_{h\in\mathcal{K}}\big\|\widehat{\Psi}'_n(x^{d-k}h^k)\big\|^2\\
&\qquad\le \sup_{\|\zeta\|\le 1}\ \sup_{h_1,\ldots,h_k\in\mathcal{K}}\langle\zeta,\widehat{\Psi}'_n(x,\ldots,x,h_1,\ldots,h_k)\rangle^2\\
&\qquad\le \sup_{\|\zeta\|\le 1}\sum_{i_{d-k+1},\ldots,i_d=0}^{\infty}\bigg|\sum_{i_1,\ldots,i_{d-k}=0}^{\infty}\langle\zeta,\widehat{b}^n_{i_1,\ldots,i_d}\rangle\langle\xi_{i_1},x\rangle\cdots\langle\xi_{i_{d-k}},x\rangle\bigg|^2\\
&\qquad= \sup_{\|\zeta\|\le 1}\int\cdots\int\langle\zeta,\widehat{\Psi}'_n(x,\ldots,x,y_1,\ldots,y_k)\rangle^2\,d\mu(y_1)\cdots d\mu(y_k)\\
&\qquad\le \int\cdots\int\big\|\widehat{\Psi}'_n(x,\ldots,x,y_1,\ldots,y_k)\big\|^2\,d\mu(y_1)\cdots d\mu(y_k).
\end{aligned}$$

By the polarization formula (5.10),

$$\begin{aligned}
&\widehat{\Psi}'_n(x,\ldots,x,y_1,\ldots,y_k)\\
&\qquad= \frac{1}{d!}\,\mathbb{E}\Big(\Psi'_n\big((\varepsilon_{k+1}+\cdots+\varepsilon_d)x+\varepsilon_1 y_1+\cdots+\varepsilon_k y_k\big)\,\varepsilon_1\cdots\varepsilon_d\Big).
\end{aligned}$$

Therefore, we obtain from the rotational invariance of Gaussian distributions and homogeneity that

$$\begin{aligned}
&(d\,!)^2\int\sup_{h\in\mathcal{K}}\big\|\widehat{\Psi}'_n(x^{d-k}h^k)\big\|^2\,d\mu(x)\\
\le&\,\mathbb{E}\iint\cdots\int\big\|\Psi'_n\big((\varepsilon_{k+1}+\cdots+\varepsilon_d)x+\varepsilon_1 y_1+\cdots+\varepsilon_k y_k\big)\big\|^2 d\mu(x)d\mu(y_1)\cdots d\mu(y_k)\\
=&\,\mathbb{E}\int\big\|\Psi'_n\big(\big((\varepsilon_{k+1}+\cdots+\varepsilon_d)^2+k\big)^{1/2}x\big)\big\|^2\,d\mu(x)\\
=&\,\mathbb{E}\Big(\big((\varepsilon_{k+1}+\cdots+\varepsilon_d)^2+k\big)^d\Big)\int\big\|\Psi'_n(x)\big\|^2 d\mu(x).
\end{aligned}$$

Hence (5.12) and therefore (5.9) are established.

We can now conclude the proof of (5.5) and thus of the theorem. It is intuitively clear that

$$\lim_{n\to\infty} \sup_{h\in\mathcal{K}} \big\| \Psi'_n(h) - \Psi^{(d)}(h) \big\| = 0. \tag{5.13}$$

This property is an easy consequence of (5.9). Indeed, for all n and $t > 0$,

$$\begin{aligned}
\sup_{h\in\mathcal{K}} &\big\|\Psi'_n(h) - \Psi^{(d)}(h)\big\| \\
&\le \sup_m \sup_{h\in\mathcal{K}} \Big\| \Psi'_m(h) - t^{-d} \int \Psi'_m(x+th) d\mu(x) \Big\| \\
&\qquad + \sup_{h\in\mathcal{K}} t^{-d} \Big\| \int \Psi'_n(x+th) - \Psi(x+th) \Big\| d\mu(x) \\
&\le \sup_m \int \sup_{h\in\mathcal{K}} \big\| \Psi'_m(h) - t^{-d}\Psi'_m(x+th) \big\| d\mu(x) \\
&\qquad + \sup_{h\in\mathcal{K}} t^{-d} \int \big\| \Psi'_n(x+th) - \Psi(x+th) \big\| \, d\mu(x)
\end{aligned}$$

and, using (5.2) and (5.9), the limit in n and then in t yields (5.13). Let now A be closed in B and take $0 < r < \mathcal{I}_\Psi(A)$. The definition of $\mathcal{I}_\Psi(A)$ indicates that $(2r)^{d/2}\Psi^{(d)}(\mathcal{K}) \cap A = \emptyset$ where we recall that the unit ball $\mathcal{K}$ of $\mathcal{H}$ is a compact subset of E. Therefore, since $\Psi^{(d)}(\mathcal{K})$ is clearly seen to be compact in B by (5.13), and since A is closed, one can find $\eta > 0$ such that

$$\big((2r)^{d/2}\Psi^{(d)}(\mathcal{K}) + B(0,2\eta)\big) \cap \big(A + B(0,\eta)\big) = \emptyset. \tag{5.14}$$

By (5.13), there exists $n_0 = n_0(\eta)$ large enough such that for every $n \ge n_0$,

$$(2r)^{d/2}\Psi'_n(\mathcal{K}) \subset (2r)^{d/2}\Psi^{(d)}(\mathcal{K}) + B(0,\eta). \tag{5.15}$$

Let thus $n \ge n_0$. For any $t > 0$, we can write

$$\begin{aligned}
&\mu\big(x; \Psi(x) \in t^d A\big) \\
&\quad \le \mu\big(x; \big\|\Psi(x) - \Psi'_n(x)\big\| > \eta t^d\big) + \mu\big(x; \Psi'_n(x) \in t^d\big(A + B(0,\eta)\big)\big) \\
&\quad \le \mu\big(x; \big\|\Psi(x) - \Psi'_n(x)\big\| > \eta t^d\big) + \mu^*\big(x; x \notin V + t\sqrt{2r}\mathcal{K}\big)
\end{aligned} \tag{5.16}$$

where

$$V = V(t,n) = \big\{v; \sup_{h\in\mathcal{K}} t^{-d} \big\| \Psi'_n\big(v + t\sqrt{2r}h\big) - t^d(2r)^{d/2}\Psi'_n(h)\big\| \le \eta\big\}.$$

To justify the second inequality in (5.16), observe that if $x = v + t\sqrt{2r}h$ with $v \in V$ and $h \in \mathcal{K}$, then

$$t^{-d}\Psi'_n(x) = t^{-d}\big[\Psi'_n\big(v + t\sqrt{2r}h\big) - t^d(2r)^{d/2}\Psi'_n(h)\big] + (2r)^{d/2}\Psi'_n(h),$$

so that the claim follows by (5.14), (5.15) and the definition of V. By (5.9), let now $t_0 = t_0(\eta)$ be large enough so that, for all $t \geq t_0$,

$$\sup_n \frac{1}{t^d} \int \sup_{h \in \mathcal{K}} \big\| \Psi'_n \big(x + t\sqrt{2r}h \big) - t^d (2r)^{d/2} \Psi'_n(h) \big\|^2 d\mu(x) \leq \frac{\eta^2}{2} \, .$$

That is, for every n and every $t \geq t_0$, $\mu(V(t,n)) \geq \frac{1}{2}$. By Theorem 4.3 (one could use equivalently (4.6)), it follows that

$$\mu^* \big(x; x \notin V + t\sqrt{2r}\mathcal{K} \big) \leq \mathrm{e}^{-rt^2} . \tag{5.17}$$

Fix now $t \geq t_0 = t_0(\eta)$. Choose $n = n(t) \geq n_0 = n_0(\eta)$ large enough in order that

$$\mu \big(x; \big\| \Psi(x) - \Psi'_n(x) \big\| > \eta t^d \big) \leq \mathrm{e}^{-rt^2} .$$

Together with (5.16) and (5.17), it follows that for every $t \geq t_0$,

$$\mu \big(x; \Psi(x) \in t^d A \big) \leq 2\mathrm{e}^{-rt^2} .$$

Since $r < \mathcal{I}_\Psi(A)$ is arbitrary, the proof of (5.5) and therefore of Theorem 5.1 is complete. □

Note that it would of course have been possible to work directly on Ψ rather than on the approximating sequence (Ψ'_n) in the preceding proof. This approach however avoids several measurability questions and makes everything more explicit.

It is probably possible to develop, as in Chapter 4, a nontopological approach to large deviations of Wiener chaos.

Notes for further reading. The reader may consult the recent paper [MW-N-PA] for a different approach to the results presented in this chapter, however also based on Borell's main contribution (Corollary 5.2). Borell's articles [Bo8], [Bo9]... contain further interesting results on chaos. See also [A-G], [G-K2], [L-T2]...

6. REGULARITY OF GAUSSIAN PROCESSES

In this chapter, we provide a complete treatment of boundedness and continuity of Gaussian processes via the tool of majorizing measures. After the work of R. M. Dudley, V. Strassen, V. N. Sudakov and X. Fernique on entropy, M. Talagrand [Ta2] gave, in 1987, necessary and sufficient conditions on the covariance structure of a Gaussian process in order that it is almost surely bounded or continuous. These necessary and sufficient conditions are based on the concept of majorizing measure introduced in the early seventies by X. Fernique and C. Preston, and inspired in particular by the "real variable lemma" of A. M. Garsia, E. Rodemich and H. Rumsey Jr. [G-R-R]. Recently, M. Talagrand [Ta7] gave a simple proof of his theorem on necessity of majorizing measures based on the concentration phenomenon for Gaussian measures. We follow this approach here. The aim of this chapter is in fact to demonstrate the actual simplicity of majorizing measures that are usually considered as difficult and obscure.

Let T be a set. A Gaussian random process (or better, random function) $X = (X_t)_{t\in T}$ is a family, indexed by T, of random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that the law of each finite family $(X_{t_1}, \ldots, X_{t_n})$, $t_1, \ldots, t_n \in T$, is centered Gaussian on $\mathbb{R}^n$. Throughout this work, Gaussian will always mean centered Gaussian. In particular, the law (the distributions of the finite dimensional marginals) of the process X is uniquely determined by the covariance structure $\mathbb{E}(X_s X_t)$, $s,t \in T$. Our aim will be to characterize almost sure boundedness and continuity (whenever T is a topological space) of the Gaussian process X in terms of an as simple as possible criterion on this covariance structure. Actually, the main point in this study will be the question of boundedness. As we will see indeed, once the appropriate bounds for the supremum of X are obtained, the characterization of continuity easily follows. Due to the integrability properties of norms of Gaussian random vectors or supremum of Gaussian processes (Theorem 4.1), we will avoid, at a first stage, various cumbersome and unessential measurability questions, by considering the supremum functional

$$F(T) = \sup\big\{\mathbb{E}\big(\sup_{t\in U} X_t\big); U \text{ finite in } T\big\}.$$

(If $S \subset T$, we define in the same way $F(S)$.) Thus, $F(T) < \infty$ if and only if X is almost surely bounded in any reasonable sense. In particular, we already see that the main question will reduce to a uniform control of $F(U)$ over the finite subsets U of T.

After various preliminary results [Fe1], [De]..., the first main idea in the study of regularity of Gaussian processes is the introduction (in the probabilistic area), by R. M. Dudley, V. Strassen and V. N. Sudakov (cf. [Du1], [Du2], [Su1-4]), of the notion of ε-entropy. The idea consists in connecting the regularity of the Gaussian process $X = (X_t)_{t \in T}$ to the size of the parameter set T for the L^2-metric induced by the process itself and given by

$$d(s,t) = \big(\mathbb{E}|X_s - X_t|^2\big)^{1/2}, \quad s,t \in T.$$

Note that this metric is entirely characterized by the covariance structure of the process. It does not necessarily separate points in T but this is of no importance. The size of T is more precisely estimated by the entropy numbers: for every $\varepsilon > 0$, let $N(T,d;\varepsilon)$ denote the minimal number of (open to fix the idea) balls of radius ε for the metric d that are necessary to cover T. The two main results concerning regularity of Gaussian processes under entropy conditions, due to R. M. Dudley [Du1] for the upper bound and V. N. Sudakov [Su3] for the lower bound (cf. [Du2], [Fe4]), are summarized in the following statement.

Theorem 6.1. *There are numerical constants $C_1 > 0$ and $C_2 > 0$ such that for all Gaussian processes $X = (X_t)_{t \in T}$,*

$$C_1^{-1} \sup_{\varepsilon > 0} \varepsilon \big(\log N(T,d;\varepsilon)\big)^{1/2} \le F(T) \le C_2 \int_0^\infty \big(\log N(T,d;\varepsilon)\big)^{1/2} d\varepsilon. \tag{6.1}$$

As possible numerical values for C_1 and C_2, one may take $C_1 = 6$ and $C_2 = 42$ (see below). The convergence of the entropy integral is understood for the small values of ε since it stops at the diameter $D(T) = \sup\{d(s,t); s,t \in T\}$. Actually, if any of the three terms of (6.1) is finite, then (T,d) is totally bounded and in particular $D(T) < \infty$. We will show in more generality below that the process $X = (X_t)_{t \in T}$ actually admits an almost surely continuous version when the entropy integral is finite. Conversely, if $X = (X_t)_{t \in T}$ is continuous, one can show that $\lim_{\varepsilon \to 0} \varepsilon (\log N(T,d;\varepsilon))^{1/2} = 0$ (cf. [Fe4]).

For the matter of comparison with the more refined tool of majorizing measures we will study next, we present a sketch of the proof of Theorem 6.1.

Proof. We start with the upper bound. We may and do assume that T is finite (although this is not strictly necessary). Let $q > 1$ (usually an integer). (We will consider q as a power of discretization; a posteriori, its value is completely arbitrary.) Let n_0 be the largest integer n in $\mathbb{Z}$ such that $N(T,d;q^{-n}) = 1$. For every $n \ge n_0$, we consider a family of cardinality $N(T,d;q^{-n}) = N(n)$ of balls of radius q^{-n} covering T. One may therefore construct a partition $\mathcal{A}_n$ of T of cardinality $N(n)$ on the basis of this covering with sets of diameter less than $2q^{-n}$. In each A of $\mathcal{A}_n$, fix a point of T and denote by T_n the collection of these points. For each t in T, denote by $A_n(t)$

the element of $\mathcal{A}_n$ that contains t. For every t and every n, let then $s_n(t)$ be the element of T_n such that $t \in A_n(s_n(t))$. Note that $d(t, s_n(t)) \leq 2q^{-n}$ for every t and $n \geq n_0$.

The main argument of the proof is the so-called chaining argument (which goes back to A. N. Kolmogorov in his proof of continuity of paths of processes under L^p-control of their increments): for every t,

$$X_t = X_{s_0} + \sum_{n>n_0} \big(X_{s_n(t)} - X_{s_{n-1}(t)}\big) \tag{6.2}$$

where $s_0 = s_{n_0}(t)$ may be chosen independent of $t \in T$. Note that

$$d\big(s_n(t), s_{n-1}(t)\big) \leq 2q^{-n} + 2q^{-n+1} = 2(q+1)q^{-n}.$$

Let $c_n = 4(q+1)q^{-n}(\log N(n))^{1/2}$, $n > n_0$. It follows from (6.2) that

$$\begin{aligned} F(T) = \mathbb{E}\big(\sup_{t\in T} X_t\big) & \\ &\leq \sum_{n>n_0} c_n + \mathbb{E}\bigg(\sup_{t\in T}\sum_{n>n_0}\big|X_{s_n(t)} - X_{s_{n-1}(t)}\big| I_{\{|X_{s_n(t)} - X_{s_{n-1}(t)}|>c_n\}}\bigg) \\ &\leq \sum_{n>n_0} c_n + \mathbb{E}\bigg(\sum_{n>n_0}\sum_{(u,v)\in H_n}\big|X_u - X_v\big| I_{\{|X_u - X_v|>c_n\}}\bigg) \end{aligned}$$

where $H_n = \{(u,v) \in T_n \times T_{n-1}; d(u,v) \leq 2(q+1)q^{-n}\}$. If G is a real centered Gaussian variable with variance less than or equal to σ^2, for every $c > 0$

$$\mathbb{E}\big(|G| I_{\{|G|>c\}}\big) \leq \sigma \mathrm{e}^{-c^2/2\sigma^2}.$$

Hence,

$$\begin{aligned} F(T) &\leq \sum_{n>n_0} c_n + \sum_{n>n_0} \mathrm{Card}(H_n) 2(q+1)q^{-n}\exp\big(-c_n^2/8(q+1)^2q^{-2n}\big) \\ &\leq \sum_{n>n_0} 4(q+1)q^{-n}\big(\log N(n)\big)^{1/2} + \sum_{n>n_0} 2(q+1)q^{-n} \\ &\leq 7(q+1)\sum_{n>n_0} q^{-n}\big(\log N(n)\big)^{1/2} \end{aligned}$$

where we used that $\mathrm{Card}(H_n) \leq N(n)^2$. Since

$$\begin{aligned} \int_0^\infty \big(\log N(T,d;\varepsilon)\big)^{1/2} d\varepsilon &\geq \sum_{n>n_0}\int_{q^{-n-1}}^{q^{-n}} \big(\log N(T,d;\varepsilon)\big)^{1/2} d\varepsilon \\ &\geq (1-q^{-1})\sum_{n>n_0} q^{-n}\big(\log N(n)\big)^{1/2}, \end{aligned}$$

the conclusion follows. If $q = 2$, we may take $C_2 = 42$.

The proof of the lower bound relies on a comparison principle known as Slepian's lemma [Sl]. We use it in the following modified form due to V. N. Sudakov, S. Chevet and X. Fernique (cf. [Su1], [Su2], [Fe4], [L-T2]): if $Y = (Y_1, \ldots, Y_n)$ and $Z = (Z_1, \ldots, Z_n)$ are two Gaussian random vectors in $\mathbb{R}^n$ such that $\mathbb{E}|Y_i - Y_j|^2 \le \mathbb{E}|Z_i - Z_j|^2$ for all i, j, then

$$\mathbb{E}\big(\max_{1\le i\le n} Y_i\big) \le \mathbb{E}\big(\max_{1\le i\le n} Z_i\big). \tag{6.3}$$

Fix $\varepsilon > 0$ and let $n \le N(T, d; \varepsilon)$. There exist therefore $t_1, \ldots, t_n$ in T such that $d(t_i, t_j) \ge \varepsilon$. Let then $g_1, \ldots, g_n$ be independent standard normal random variables. We have, for every $i, j = 1, \ldots, n$,

$$\mathbb{E}\bigg|\frac{\varepsilon}{\sqrt{2}}\, g_i - \frac{\varepsilon}{\sqrt{2}}\, g_j\bigg|^2 = \varepsilon^2 \le d(t_i, t_j) = \mathbb{E}|X_{t_i} - X_{t_j}|^2.$$

Therefore, by (6.3),

$$F(T) \ge \mathbb{E}\big(\max_{1\le i\le n} X_{t_i}\big) \ge \frac{\varepsilon}{\sqrt{2}}\, \mathbb{E}\big(\max_{1\le i\le n} g_i\big).$$

Now, it is classical and easily seen that

$$\mathbb{E}\big(\max_{1\le i\le n} g_i\big) \ge c\,(\log n)^{1/2}$$

for some numerical $c > 0$ (one may choose c such that $\sqrt{2}/c \le 6$). Since n is arbitrary less than or equal to $N(T, d; \varepsilon)$, the conclusion trivially follows. Theorem 6.1 is established. □

As an important remark for further purposes, note that simple proofs of Sudakov's minoration avoiding the rather rigid Slepian's lemma are now available. These are based on duality of entropy numbers [TJ] and are presented in [L-T2]. They allow the investigation of minoration inequalities outside the Gaussian setting (cf. [Ta10], [Ta12]). Note furthermore that we will only use the Sudakov inequality in the proof of the majorizing measure minoration principle (cf. Lemma 6.4).

A simple example of application of Theorem 6.1 is Brownian motion $(W(t))_{0\le t\le 1}$ on $T = [0, 1]$. Since $d(s, t) = \sqrt{|s - t|}$, the entropy numbers $N(T, d; \varepsilon)$ are of the order of ε^{-2} as ε goes to zero and the entropy integral is trivially convergent. Together with the proof of continuity presented below in the framework of majorizing measures, Theorem 6.1 is certainly the shortest way to prove boundedness and continuity of the Brownian paths.

In Theorem 6.1, the difference between the upper and lower bounds is rather tight. It however exists. The examples of a standard orthogaussian sequence or of the canonical Gaussian process indexed by an ellipsoid in a Hilbert space (see [Du1], [Du2], [L-T2], [Ta13]) are already instructive. We will see later on that the convergence of Dudley's entropy integral however characterizes $F(T)$ when T has a group structure and the metric d is translation invariant, an important result of X. Fernique [Fe4].

If one tries to imagine what can be used instead of the entropy numbers in order to sharpen the conclusions of Theorem 6.1, one realizes that one feature of entropy is that is attributes an equal weight to each piece of the parameter set T. One is then naturally led to the possible following definition. Let, as in the proof of Theorem 6.1, q be (an integer) larger than 1. Let $\mathcal{A} = (\mathcal{A}_n)_{n \in \mathbb{Z}}$ be an increasing sequence (i.e. each $A \in \mathcal{A}_{n+1}$ is contained in some $B \in \mathcal{A}_n$) of finite partitions of T such that the diameter $D(A)$ of each element A of $\mathcal{A}_n$ is less than or equal to $2q^{-n}$. If $t \in T$, denote by $A_n(t)$ the element of $\mathcal{A}_n$ that contains t. Now, for each partition $\mathcal{A}_n$, one may consider nonnegative weights $\alpha_n(A)$, $A \in \mathcal{A}_n$, such that $\sum_{A \in \mathcal{A}_n} \alpha_n(A) \leq 1$. Set then

$$\Theta_{\mathcal{A},\alpha} = \Theta_{\mathcal{A},\alpha}(T,d) = \sup_{t \in T} \sum_n q^{-n} \left(\log \frac{1}{\alpha_n(A_n(t))} \right)^{1/2}. \tag{6.4}$$

It is worthwhile mentioning that for $2q^{-n} \geq D(T)$, one can take $\mathcal{A}_n = \{T\}$ and $\alpha_n(T) = 1$. Denote by $\Theta(T,d)$ the infimum of the functional $\Theta_{\mathcal{A},\alpha}$ over all possible choices of partitions $(\mathcal{A}_n)_{n \in \mathbb{Z}}$ and weights $\alpha_n(A)$. In this definition, we may take equivalently

$$\Theta_{\mathcal{A},m} = \sup_{t \in T} \sum_n q^{-n} \left(\log \frac{1}{m(A_n(t))} \right)^{1/2}$$

where m is a probability measure on (T,d). Indeed, if $\Theta_{\mathcal{A},\alpha} < \infty$, it is easily seen that $D(T) < \infty$. Let then n_0 be the largest integer n in $\mathbb{Z}$ such that $2q^{-n} \leq D(T)$. Fix a point in each element of $\mathcal{A}_n$ and denote by T_n, $n \geq n_0$, the collection of these points. It is then clear that if m is a (discrete) probability measure such that

$$m \geq (1 - q^{-1}) \sum_{n \geq n_0} q^{-n+n_0} \sum_{t \in T_n} \alpha_n\big(A_n(t)\big)\delta_t,$$

where δ_t is point mass at t, the functional $\Theta_{\mathcal{A},m}$ is of the same order as $\Theta_{\mathcal{A},\alpha}$ (see also below). We need not actually be concerned with these technical details and consider for simplicity the functionals $\Theta_{\mathcal{A},\alpha}$. Furthermore, the number $q > 1$ should be thought as a universal constant.

The condition $\Theta(T,d) < \infty$ is called a majorizing measure condition and the main result of this section is that $C^{-1}\Theta(T,d) \leq F(T) \leq C\Theta(T,d)$ for some constant $C > 0$ only depending on q. In order to fully appreciate this definition, it is worthwhile comparing it to the entropy integral. As we used it in the proof of Theorem 6.1, the entropy integral is equivalent (for any q) to the series

$$\sum_{n > n_0} q^{-n} \big(\log N(T,d;q^{-n})\big)^{1/2}.$$

We then construct an associated sequence $(\mathcal{A}_n)_{n \in \mathbb{Z}}$ of increasing partitions of T and weights $\alpha_n(A)$ in the following way. Let $\mathcal{A}_n = \{T\}$ and $\alpha_n(T) = 1$ for every $n \leq n_0$. Once $\mathcal{A}_n$ $(n > n_0)$ has been constructed, partition each element A of $\mathcal{A}_n$ with a covering of A of cardinality at most $N(A,d;q^{-n-1}) \leq N(T,d;q^{-n-1})$ and let $\mathcal{A}_{n+1}$ be the collection of all the subsets of T obtained in this way. To each A in $\mathcal{A}_n$, $n > n_0$, we give the weight

$$\alpha_n(A) = \left(\prod_{i=n_0+1}^{n} N(T,d;q^{-i})\right)^{-1}$$

$(\alpha(T) = 1)$. Clearly $\sum_{A\in\mathcal{A}_n} \alpha_n(A) \leq 1$. Moreover, for each t in T,

$$\begin{aligned}\sum_{n>n_0} q^{-n}\left(\log\frac{1}{\alpha(A_n(t))}\right)^{1/2} &\leq \sum_{n>n_0}\sum_{i=n_0+1}^{n} q^{-n}\big(\log N(T,d;q^{-i})\big)^{1/2}\\ &\leq (q-1)^{-1}\sum_{i>n_0} q^{-i}\big(\log N(T,d;q^{-i})\big)^{1/2}.\end{aligned}$$

In other words,

$$\Theta(T) \leq C\int_0^\infty \big(\log N(T,d;\varepsilon)\big)^{1/2} d\varepsilon$$

where $C > 0$ only depends on $q > 1$.

It is clear from this construction how entropy numbers give a uniform weight to each subset of T and how the possible refined tool of majorizing measures can allow a better understanding of the metric properties of T. (Actually, one has rather to think about entropy numbers as the equal weight that is put on each piece of a partition of the parameter set T.) This is what we will investigate now. First however, we would like to briefly comment on the name "majorizing measure" as well as the dependence on $q > 1$ in the definition of the functional $\Theta(T,d)$. Classically, a majorizing measure m on T is a probability measure on the Borel sets of T such that

$$\sup_{t\in T}\int_0^\infty \left(\log\frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon < \infty \tag{6.5}$$

where $B(t,\varepsilon)$ is the ball in T with center t and radius $\varepsilon > 0$. As the definition of the entropy integral, a majorizing measure condition only relies on the metric structure of T and the convergence of the integral is for the small values of ε. In order to connect this definition with the preceding one (6.4), let $q > 1$ and let $(\mathcal{A}_n)_{n\in\mathbb{Z}}$ be an increasing sequence of finite partitions of T such that the diameter $D(A)$ of each element A of $\mathcal{A}_n$ is less than or equal to $2q^{-n}$. Let furthermore m be a probability measure on T. Note that $A_n(t) \subset B(t, 2q^{-n})$ for every t. Therefore

$$\begin{aligned}\int_0^\infty \left(\log\frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon &\leq C\sum_n q^{-n}\left(\log\frac{1}{m(B(t,2q^{-n}))}\right)^{1/2}\\ &\leq C\sum_n q^{-n}\left(\log\frac{1}{m(A_n(t))}\right)^{1/2}\end{aligned}$$

where $C > 0$ only depends on q. Since m is a probability measure, we can set $\alpha_n(A) = m(A)$ for every A in $\mathcal{A}_n$ and every n. It immediately follows that, for every $q > 1$,

$$\inf_m \sup_{t\in T}\int_0^\infty \left(\log\frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon \leq C\Theta(T,d)$$

where C only depends on q. One can prove the reverse inequality in the same spirit with the help however of a somewhat technical and actually nontrivial discretization lemma (cf. [L-T2], Proposition 11.10). In particular, the various functionals $\Theta(T,d)$ when q varies are all equivalent. We actually need not really be concerned with these technical details since our aim is to show that $F(T)$ and $\Theta(T,d)$ are of the same order (for some $q>1$). (It will actually follow from the proofs presented below that the functionals $\Theta(T,d)$ are equivalent up to constants depending only on $q \geq q_0$ for some universal q_0 large enough.)

Now, we start our investigation of the regularity properties of a Gaussian process $X=(X_t)_{t\in T}$ under majorizing measure conditions. The first part of our study concerns upper bounds and sufficient conditions for boundedness and continuity of X. The following theorem is due, in this form and with this proof, to X. Fernique [Fe3], [Fe4]. It follows independently from the work of C. Preston [Pr1], [Pr2].

Theorem 6.2. *Let $X=(X_t)_{t\in T}$ be a Gaussian process indexed by a set T. Then, for every $q>1$,*

$$F(T) \leq C\Theta(T,d)$$

where $C>0$ only depends on q. If, in addition to $\Theta_{\mathcal{A},\alpha}<\infty$ for some partition $\mathcal{A}$ and weights α, one has

$$\lim_{k\to\infty} \sup_{t\in T} \sum_{n\geq k} q^{-n}\left(\log \frac{1}{\alpha_n(A_n(t))}\right)^{1/2} = 0, \tag{6.6}$$

then X admits a version with almost all sample paths bounded and uniformly continuous on (T,d).

Proof. It is very similar to the proof of the upper bound in Theorem 6.1. We first establish the inequality $F(T)\leq C\Theta_{\mathcal{A},\alpha}(T,d)$ for any partition $\mathcal{A}$ and any family of weights α. We may asssume that T is finite. Let n_0 be the largest integer n in $\mathbb{Z}$ such that the diameter $D(T)$ of T is less than or equal to $2q^{-n}$. For every $n\geq n_0$, fix a point in each element of the partition $\mathcal{A}_n$ and denote by T_n the (finite) collection of these points. We may take $T_{n_0}=\{s_0\}$ for some fixed s_0 in T. For every t in T, denote by $s_n(t)$ the element of T_n which belongs to $A_n(t)$. As in (6.2), for every t,

$$X_t = X_{s_0} + \sum_{n>n_0}\big(X_{s_n(t)} - X_{s_{n-1}(t)}\big).$$

Since the partitions $\mathcal{A}_n$ are increasing,

$$s_n(t) \in A_{n-1}\big(s_n(t)\big) = A_{n-1}(t), \quad n>n_0.$$

In particular, $d(s_n(t),s_{n-1}(t))\leq 2q^{-n+1}$. Now, for every t in T and every $n>n_0$, let

$$c_n(t) = 2\sqrt{2}q^{-n+1}\left(\log\frac{1}{\alpha_n(A_n(t))}\right)^{1/2}.$$

With respect to the entropic proof, note here the dependence of c_n on t which is the main feature of the majorizing measure technique. Actually, the partitions $\mathcal{A}$ and weights α are used to bound, in the chaining argument, the "heaviest" portions of the process. We can now write, almost as in the proof of Theorem 6.1,

$$\begin{aligned}
&F(T)\\
&\quad\le \sup_{t\in T}\sum_{n>n_0} c_n(t) + \mathbb{E}\Big(\sup_{t\in T}\sum_{n>n_0}\big|X_{s_n(t)}-X_{s_{n-1}(t)}\big| I_{\{|X_{s_n(t)}-X_{s_{n-1}(t)}|>c_n(t)\}}\Big)\\
&\quad\le \sup_{t\in T}\sum_{n>n_0} c_n(t) + \mathbb{E}\Big(\sum_{n>n_0}\sum_{u\in T_n}\big|X_u-X_{s_{n-1}(u)}\big| I_{\{|X_u-X_{s_{n-1}(u)}|>c_n(u)\}}\Big)\\
&\quad\le \sup_{t\in T}\sum_{n>n_0} c_n(t) + \sum_{n>n_0}\sum_{u\in T_n} 2q^{-n+1}\exp\big(-c_n^2(u)/8q^{-2n+2}\big).
\end{aligned}$$

Therefore

$$\begin{aligned}
F(T) &\le \sup_{t\in T}\sum_{n>n_0} c_n(t) + \sum_{n>n_0} 2q^{-n+1}\sum_{u\in T_n}\alpha_n\big(A_n(u)\big)\\
&\le \sup_{t\in T}\sum_{n>n_0} c_n(t) + 2(q-1)^{-1}q^{-n_0+1}.
\end{aligned}$$

Since

$$\Theta_{\mathcal{A},\alpha} \ge (\log 2)^{1/2} q^{-n_0-1},$$

the first claim of Theorem 6.2 follows.

We turn to the sample path continuity. Let $\eta>0$. For each k $(>n_0)$, set

$$\begin{aligned}
V = V_k = \big\{(x,y)\in T_k\times T_k;\, \exists\, u,v \text{ in } T \text{ such that}\\
d(u,v)\le\eta \text{ and } s_k(u)=x, s_k(v)=y\big\}.
\end{aligned}$$

If $(x,y)\in V$, we fix $u_{x,y}$, $v_{x,y}$ in T such that $s_k(u_{x,y})=x$, $s_k(v_{x,y})=y$ and $d(u_{x,y},v_{x,y})\le\eta$. Now, let s,t in T with $d(s,t)\le\eta$. Set $x=s_k(s)$, $y=s_k(t)$. Clearly $(x,y)\in V$. By the triangle inequality,

$$\begin{aligned}
|X_s-X_t| &\le |X_s-X_{s_k(s)}| + |X_{s_k(s)}-X_{u_{x,y}}| + |X_{u_{x,y}}-X_{v_{x,y}}|\\
&\qquad + |X_{v_{x,y}}-X_{s_k(t)}| + |X_{s_k(t)}-X_t|\\
&\le \sup_{(x,y)\in V}|X_{u_{x,y}}-X_{v_{x,y}}| + 4\sup_{r\in T}|X_r-X_{s_k(r)}|.
\end{aligned}$$

Clearly,

$$\mathbb{E}\big(\sup_{(x,y)\in V}|X_{u_{x,y}}-X_{v_{x,y}}|\big) \le \eta\big(\mathrm{Card}(T_k)\big)^2.$$

Now, the chaining argument in the proof of boundedness similarly shows that

$$\mathbb{E}\big(\sup_{t\in T}|X_t-X_{s_k(t)}|\big) \le C\sup_{t\in T}\sum_{n\ge k} q^{-n}\left(\log\frac{1}{\alpha_n(A_n(t))}\right)^{1/2}$$

for some constant $C > 0$ (independent of k). Therefore, hypothesis (6.6) and the preceding inequalities ensure that for each $\varepsilon > 0$ there exists $\eta > 0$ such that, for every finite and thus also countable subset U of T,

$$\mathbb{E}\Big(\sup_{s,t\in U, d(s,t)\leq\eta} |X_s - X_t|\Big) \leq \varepsilon.$$

Since (T,d) is totally bounded, there exists U countable and dense in T. Then, set $\widetilde{X}_t = X_t$ if $t \in U$ and $\widetilde{X}_t = \lim X_t$ where the limit, in probability or in L^1, is taken for $u \to t$, $u \in U$. Then $(\widetilde{X}_t)_{t\in T}$ is a version of the process X with uniformly continuous sample paths on (T,d). Indeed, let, for each integer n, $\eta_n > 0$ be such that

$$\mathbb{E}\Big(\sup_{d(s,t)\leq\eta_n} \big|\widetilde{X}_s - \widetilde{X}_t\big|\Big) \leq 4^{-n}.$$

Then, if $C_n = \{\sup_{d(s,t)\leq\eta_n} |\widetilde{X}_s - \widetilde{X}_t| \geq 2^{-n}\}$, $\sum_n \mathbb{P}(C_n) < \infty$ and the claim follows from the Borel-Cantelli lemma. The proof of Theorem 6.2 is complete. □

We now turn to the theorem of M. Talagrand [Ta2] on necessity of majorizing measures. This result was conjectured by X. Fernique back in 1974. As announced, we follow the simplified proof of the author [Ta7] based on concentration of Gaussian measures. This new proof moreover allows us to get some insight on the weights α of the "minorizing" measure.

Theorem 6.3. *There exists a universal value $q_0 \geq 2$ such that for every $q \geq q_0$ and every Gaussian process $X = (X_t)_{t\in T}$ indexed by T,*

$$\Theta(T,d) \leq CF(T)$$

where $C > 0$ is a constant only depending on q.

Proof. The key step is provided by the following minoration principle based on concentration and Sudakov's inequality. It may actually be considered as a strengthening of the latter.

Lemma 6.4. *There exists a numerical constant $0 < c < \frac{1}{2}$ with the following property. If $\varepsilon > 0$ and if $t_1, \ldots, t_N$ are points in T such that $d(t_k, t_\ell) \geq \varepsilon$, $k \neq \ell$, $N \geq 2$, and if $B_1, \ldots, B_N$ are subsets of T such that $B_k \subset B(t_k, c\varepsilon)$, $k = 1, \ldots N$, we have*

$$\mathbb{E}\Big(\max_{1\leq k\leq N} \sup_{t\in B_k} X_t\Big) \geq c\varepsilon(\log N)^{1/2} + \min_{1\leq k\leq N} \mathbb{E}\Big(\sup_{t\in B_k} X_t\Big).$$

Proof. We may assume that B_k is finite for every k. Set $Y_k = \sup_{t\in B_k}(X_t - X_{t_k})$, $k = 1, \ldots, N$. Then,

$$\sup_{t\in B_k} X_t = (Y_k - \mathbb{E}Y_k) + \mathbb{E}Y_k + X_{t_k}$$

and thus

$$\max_{1\leq k\leq N} X_{t_k} \leq \max_{1\leq k\leq N} \sup_{t\in B_k} X_t + \max_{1\leq k\leq N} |Y_k - \mathbb{E}Y_k| - \min_{1\leq k\leq N} \mathbb{E}\Big(\sup_{t\in B_k} X_t\Big). \tag{6.7}$$

Integrate both sides of this inequality. By Sudakov's minoration (Theorem 6.1),

$$\mathbb{E}\big(\max_{1\le k\le N} X_{t_k}\big) \ge C_1^{-1}\varepsilon(\log N)^{1/2}.$$

Furthermore, the concentration inequalities, in the form for example of (2.9) or (4.2), (4.3), show that, for every $r\ge 0$, and every k,

$$\mathbb{P}\big\{|Y_k - \mathbb{E}Y_k| \ge r\big\} \le 2\mathrm{e}^{-r^2/2c^2\varepsilon^2}.$$

This estimate easily and classically implies that

$$\mathbb{E}\big(\max_{1\le k\le N} |Y_k - \mathbb{E}Y_k|\big) \le C_3 c\varepsilon(\log N)^{1/2}$$

where $C_3>0$ is numerical. Indeed, by the integration by parts formula, for every $\delta>0$,

$$\begin{aligned}\mathbb{E}\big(\max_{1\le k\le N} |Y_k - \mathbb{E}Y_k|\big) &\le \delta + \int_\delta^\infty \mathbb{P}\big\{\max_{1\le k\le N} |Y_k - \mathbb{E}Y_k| \ge r\big\}dr\\ &\le \delta + 2N\int_\delta^\infty \mathrm{e}^{-r^2/2c^2\varepsilon^2}dr\end{aligned}$$

and the conclusion follows by letting δ be of the order of $c\varepsilon(\log N)^{1/2}$. Hence, coming back to (6.7), we see that if $c>0$ is such that $\frac{1}{C_1} - cC_3 = c$, the minoration inequality of the lemma holds. The value of q_0 in Theorem 6.3 only depends on this choice. (Since we may take $C_1 = 6$ and $C_3 = 20$ (for example), we see that $c = .007$ will work.) Lemma 6.4 is proved. □

We now start the proof of Theorem 6.3 itself and the construction of a partition $\mathcal{A}$ and weights α. Assume that $F(T)<\infty$ otherwise there is nothing to prove. In particular, (T,d) is totally bounded. We further assume that $q\ge q_0$ where $q_0 = c^{-1}$ has been determined by Lemma 6.4.

For each n and each subset of T of diameter less than or equal to $2q^{-n}$, we will construct an associated partition in sets of diameter less than or equal to $2q^{-n-1}$. Let thus S be a subset of T with $D(S)\le 2q^{-n}$. We first construct by induction a (finite) sequence $(t_k)_{k\ge 1}$ of points in S. t_1 is chosen so that $F(S\cap B(t_1, q^{-n-2}))$ is maximal. Assume that $t_1,\dots,t_{k-1}$ have been constructed and set

$$H_k = \bigcup_{\ell<k}\big(S\cap B(t_\ell, q^{-n-1})\big).$$

If $H_k = S$, the construction stops (and it will eventually stop since (T,d) is totally bounded). If not, choose t_k in $S\setminus H_k$ such that $F(B_k)$ is maximal where we set $B_k = (S\setminus H_k)\cap B(t_k, q^{-n-2})$. For every k, let

$$A_k = (S\setminus H_k)\cap B(t_k, q^{-n-1}).$$

Clearly $D(A_k) \le 2q^{-n-1}$ and the A_k's define a partition of S. One important feature of this construction is that, for every t in A_k,

$$F\big(A_k \cap B(t, q^{-n-2})\big) \le F(B_k). \tag{6.8}$$

On the other hand, the minoration lemma 6.4 applied with $\varepsilon = q^{-n-1}$ yields (since $q \ge c^{-1}$), for every k,

$$F(S) \ge cq^{-n-1}(\log k)^{1/2} + F(B_k). \tag{6.9}$$

We denote by $\mathcal{A}(S)$ this ordered finite partition $\{A_1, \ldots, A_k, \ldots\}$ of S. (6.8) and (6.9) together yield: for every $A_k \in \mathcal{A}(S)$ and every $U \in \mathcal{A}(A_k)$,

$$F(S) \ge cq^{-n-1}(\log k)^{1/2} + F(U). \tag{6.10}$$

We now complete the construction. Let n_0 be the largest in $\mathbb{Z}$ with $D(T) \le 2q^{-n_0}$. Set $\mathcal{A}_n = \{T\}$ and $\alpha_n(T) = 1$ for every $n \le n_0$. Suppose that $\mathcal{A}_n$ and $\alpha_n(S)$, $S \in \mathcal{A}_n$, $n > n_0$, have been constructed. We define

$$\mathcal{A}_{n+1} = \bigcup \big\{\mathcal{A}(S); S \in \mathcal{A}_n\big\}.$$

If $U \in \mathcal{A}_{n+1}$, there exists $S \in \mathcal{A}_n$ such that $U = A_k \in \mathcal{A}(S)$. We then set $\alpha_{n+1}(U) = \alpha_n(A)/2k^2$. Let t be fixed in T. With this notation, (6.10) means that for all $n \ge n_0$,

$$F\big(A_n(t)\big) \ge c\,2^{-1/2}q^{-n-1}\left(\log \frac{\alpha_n(A_n(t))}{2\alpha_{n+1}(A_{n+1}(t))}\right)^{1/2} + F\big(A_{n+2}(t)\big)$$

where we recall that we denote by $A_n(t)$ the element of $\mathcal{A}_n$ that contains t. Summing these inequalities separately on the even and odd integers, we get

$$2F(T) \ge c\,2^{-1/2} \sum_{n > n_0} q^{-n-1}\left(\log \frac{\alpha_n(A_n(t))}{2\alpha_{n+1}(A_{n+1}(t))}\right)^{1/2}$$

and thus

$$c(q-1)^{-1}q^{-n_0} + 2F(T) \ge c\,2^{-1/2}(1 - q^{-1}) \sum_{n > n_0} q^{-n}\left(\log \frac{1}{\alpha_n(A_n(t))}\right)^{1/2}.$$

Since $2q^{-n_0} \le D(T)$, and since

$$\begin{aligned} 2F(T) &= \sup\big\{\mathbb{E}\big(\sup_{s,t \in U} |X_s - X_t|\big); U \text{ finite in } T\big\} \\ &\ge \sup_{s,t \in T} \mathbb{E}|X_s - X_t| = \left(\frac{2}{\sqrt{\pi}}\right)^{1/2} D(T), \end{aligned}$$

it follows that, for some constant $C > 0$ only depending on q,

$$CF(T) \geq c \sum_{n > n_0} q^{-n} \left(\log \frac{1}{\alpha_n(A_n(t))} \right)^{1/2}.$$

Theorem 6.3 is therefore established. □

It may be shown that if the Gaussian process X in Theorem 6.3 is almost surely continuous on (T, d), then there is a majorizing measure satisfying (6.6). We refer to [Ta2] or [L-T2] for the details.

Theorem 6.3 thus solved the question of the regularity properties of any Gaussian process. Prior to this result however, X. Fernique showed [Fe4] that the convergence of Dudley's entropy integral was necessary for a stationary Gaussian process to be almost surely bounded or continuous. One can actually easily show (cf. [L-T2]) that, in this case, the entropy integral coincides with a majorizing measure integral with respect to the Haar measure on the underlying parameter set T endowed with a group structure. One may however also provide a direct and transparent proof of the stationary case on the basis of the above minoration principle (Lemma 6.4). We would like to conclude this chapter with a brief sketch of this proof.

Let thus T be a locally compact Abelian group. Let $X = (X_t)_{t \in T}$ be a stationary centered Gaussian process indexed by T, in the sense that the L^2-metric d induced by X is translation invariant on T. As announced, we aim to prove directly that for some numerical constant $C > 0$,

$$\int_0^\infty \big(\log N(T, d; \varepsilon) \big)^{1/2} d\varepsilon \leq CF(T).$$

(cf. [Fe4], [M-P], [L-T2] for more general statements along these lines.) Since d is translation invariant,

$$\mathbb{E}\big(\sup_{s \in B(t,\varepsilon)} X_s \big) \quad \text{and} \quad N\big(B(t,\varepsilon), d; \eta\big), \quad \varepsilon, \eta > 0,$$

are independent of the point t. They will therefore be simpler denoted as

$$\mathbb{E}\big(\sup_{s \in B(\varepsilon)} X_s \big) \quad \text{and} \quad N\big(B(\varepsilon), d; \eta\big).$$

Let $n \in \mathbb{Z}$. Choose in a ball $B(q^{-n})$ a maximal family $(t_1, \ldots, t_M)$ under the relations $d(t_k, t_\ell) \geq q^{-n-1}$, $k \neq \ell$. Then the balls $B(t_k, q^{-n-1})$, $1 \leq k \leq M$, cover $B(q^{-n})$ so that $M \geq N(B(q^{-n}), d; q^{-n-1})$. Apply then Lemma 6.4 with $\varepsilon = q^{-n-1}$, $q \geq q_0 = c^{-1}$ and $B_k = B(t_k, q^{-n-2})$. We thus get

$$\mathbb{E}\big(\sup_{t \in B(q^{-n})} X_t \big) \geq c q^{-n-1} \big(\log N(B(q^{-n}), d; q^{-n-1}) \big)^{1/2} + \mathbb{E}\big(\sup_{t \in B(q^{-n-2})} X_t \big).$$

Summing as before these inequalities along the even and the odd integers yields

$$F(T) \geq C^{-1} \sum_n q^{-n} \big(\log N(B(q^{-n}), d; q^{-n-1} \big)^{1/2}.$$

Since

$$N(T,d;q^{-n-1}) \leq N(T,d;q^{-n})N\big(B(q^{-n}),d;q^{-n-1})\big),$$

the proof is complete.

To conclude, let us mention the following challenging open problem. Let x_i, $i \in \mathbb{N}$, be real valued functions on a set T such that $\sum_i x_i(t)^2 < \infty$ for every $t \in T$. Let furthermore $(\varepsilon_i)_{i\in\mathbb{N}}$ be a sequence of independent symmetric Bernoulli random variables and set, for each $t \in T$, $X_t = \sum_i \varepsilon_i x_i(t)$ which converges almost surely. The question of characterizing those "Bernoulli" processes $(X_t)_{t\in T}$ which are almost surely bounded is almost completely open (cf. [L-T2], [Ta14]). The Gaussian study of this chapter of course corresponds to the choice for $(\varepsilon_i)_{i\in\mathbb{N}}$ of a standard Gaussian sequence.

Notes for further reading. On the history of entropy and majorizing measures, one may consult respectively [Du2], [Fe4] and [He], [Fe4], [Ta2], [Ta18]. The first proof of Theorem 6.3 by M. Talagrand [Ta2] was quite different from the proof presented here following [Ta7]. Another proof may be found in [L-T2]. These proofs are based on the fundamental principle, somewhat hidden here, that the size of a metric space with respect to the existence of a majorizing measure can be measured by the size of the well separated subsets it contains (see [Ta10], [Ta12] for more on this principle). More on majorizing measures and minoration of random processes may be found in [L-T2] and in the papers [Ta10], [Ta12], and in the recent survey [Ta18] where in particular new examples of applications are described. It is shown in [L-T2] how the upper bound techniques based on entropy or majorizing measures (Theorems 6.1 and 6.2) can yield deviation inequalities of the type (4.2), which are optimal by Theorem 6.3. Sharp bounds on the tail of the supremum of a Gaussian process can be obtained with these methods (see e.g. [Ta13], [Lif2], [Lif3] and the many references therein). On construction of majorizing measures, see [L-T2], [Ta14], [Ta18]. For the applications of the Dudley-Fernique theorem on stationary Gaussian processes to random Fourier series, see [M-P], [L-T2].

7. SMALL BALL PROBABILITIES FOR GAUSSIAN MEASURES AND RELATED INEQUALITIES AND APPLICATIONS

While, as we saw in Chapter 4, the behavior of (the complement of) large balls for Gaussian measures is relatively well described, small ball probabilities are much less known. This problem has gone recently a quick development and we intend to present in this chapter some significant recent results, although it seems that there is still a long way to the final word (if there is any). In the first part of this chapter, we describe a simple method to evaluate small Brownian balls and to establish various sharper concentration inequalities. Then, we present some more abstract and general results (due to J. Kuelbs and W. Li and M. Talagrand) which show in particular, on the basis of the isoperimetric tool, that small ball probabilities for Gaussian measures are closely related to some entropy numbers. In particular, we establish an important concentration inequality for enlarged balls due to M. Talagrand. To conclude this chapter, we briefly discuss some correlation inequalities which have been used recently to extend the support of a diffusion theorem, the large deviations of dynamical systems as well as the existence of Onsager-Machlup functionals for stronger norms or topologies on Wiener space.

We introduce the question of small ball probabilities for Gaussian measures with the example of Wiener measure. Let $W = (W(t))_{t \geq 0}$ be Brownian motion with values in $\mathbb{R}^d$. Denote by $\|x\|_\infty = \sup_{t\in[0,1]} |x(t)|$ the supnorm on the space of continuous functions $C_0([0,1];\mathbb{R}^d)$ (vanishing at the origin) where we equip, for example, $\mathbb{R}^d$ with its usual Euclidean norm $|\cdot|$. Let $\varepsilon > 0$. By the scaling property, for every $\lambda > 0$,

$$\mathbb{P}\big\{\|W\|_\infty \leq \varepsilon\big\} = \mathbb{P}\Big\{\sup_{0\leq t\leq\lambda} |W(t)| \leq \varepsilon\sqrt{\lambda}\Big\}.$$

Choosing $\lambda = \varepsilon^{-2}$, we see that

$$\mathbb{P}\big\{\|W\|_\infty \leq \varepsilon\big\} = \mathbb{P}\big\{\tau \geq \varepsilon^{-2}\big\}$$

where τ is the exit time of W from the unit ball B of $\mathbb{R}^d$. It is known (cf. [I-W]) that $u(t,x) = \mathbb{E}(f(W(t)+x)I_{\{\tau\geq t\}})$, $x \in B$, $t \geq 0$, is the solution of the initial value

Dirichlet problem

$$\frac{\partial u}{\partial t} = \frac{1}{2}\,\Delta u \text{ in } B, \quad u_{|\partial B} = 0, \quad u_{|t=0} = f.$$

Therefore

$$u(t,x) = \sum_{n=1}^{\infty} \mathrm{e}^{-\lambda_n t}\phi_n(x)\int_B \phi_n(y)f(y)dy,$$

where $0 < \lambda_1 \le \lambda_2 \le \ldots$ are eigenvalues and $\phi_1, \phi_2, \ldots$ are corresponding eigenfunctions of the eigenvalue problem

$$\frac{1}{2}\,\Delta\phi + \lambda\phi = 0 \text{ in } B, \quad \phi_{|\partial B} = 0.$$

In particular

$$\mathbb{P}\big\{\tau \ge \varepsilon^{-2}\big\} = \sum_{n=1}^{\infty} \mathrm{e}^{-\lambda_n/\varepsilon^2}\phi_n(0)\int_B \phi_n(y)dy$$

and thus

$$\mathbb{P}\big\{\|W\|_\infty \le \varepsilon\big\} \sim \mathrm{e}^{-\lambda_1/\varepsilon^2}\phi_1(0)\int_B \phi_1(y)dy. \tag{7.1}$$

In particular, it is known that $\lambda_1 = \pi^2/8$ for $d = 1$.

While the proof of (7.1) relies on some very specific properties of both the Brownian paths and the supnorm, one may wonder for the behavior of small ball probabilities for some other norms on Wiener space, such as for example L^p-norms or the classical Hölder norms of index α for every $0 < \alpha < \frac{1}{2}$. In what follows, we will describe some small ball Brownian probabilities, including the ones just mentioned, using some more abstract tools (which could eventually generalize to other Gaussian measures). We will however only work at the logarithmic scale. We use series expansions of Brownian motion in the Haar basis of $[0,1]$. We present the various results following the exposition of W. Stolz [St1], inspired by the works [B-R], [Ta9] and [Ta14] (among others). For simplicity, we work below with a one-dimensional Brownian motion and write $C_0([0,1])$ for $C_0([0,1];\mathbb{R})$.

We only concentrate here on the Brownian case. We mention at the end of the chapter references of extensions to some more general processes. Of course, a lot is known on small Hilbert balls for arbitrary Gaussian measures (cf. e.g. [Sy], [Zo], [K-L-L], [Li]...).

Let h_0, h_m, $m = 2^n + k - 1$, $n \ge 0$, $k = 1, \ldots, 2^n$ be the Haar functions on $[0,1]$. That is, $h_0 \equiv 1$,

$$h_1 = I_{[0,1/2)} - I_{[1/2,1]},$$

and, for every $m = 2^n + k - 1$, $n \ge 1$, $k = 1, \ldots, 2^n$,

$$h_m(t) = 2^{n/2}h_1(2^n t - k + 1), \quad 0 \le t \le 1.$$

Define then the Schauder functions φ_m, $m \in \mathbb{N}$, on $[0,1]$ by setting $\varphi_m(t) = \int_0^t h_m(s)ds$. The Schauder functions form a basis of the space of continous functions on $[0,1]$. In particular, Lévy's representation of Brownian motion may be expressed by saying that, almost surely,

$$W(t) = \sum_{m=0}^{\infty} g_m \varphi_m(t), \quad t \in [0,1], \tag{7.2}$$

where $(g_m)_{m\in\mathbb{N}}$ is a standard Gaussian sequence and where the convergence takes place uniformly on $[0,1]$ (cf. Proposition 4.2).

In what follows, $\|\cdot\|$ is a measurable norm on $C_0([0,1])$ for which the Wiener measure is a Radon measure, in other words for which the series (7.2) converges almost surely (cf. Proposition 4.2).

Theorem 7.1. *Let $0 \le \alpha < \frac{1}{2}$. If, for some constant $C > 0$,*

$$\bigg\| \sum_{k=1}^{2^n} a_k \varphi_{2^n+k-1} \bigg\| \le C 2^{-(\frac{1}{2}-\alpha)n} \max_{1\le k\le 2^n} |a_k|$$

for all real numbers $a_1, \ldots, a_{2^n}$ and every $n \ge 0$, then

$$\log \mathbb{P}\big\{ \|W\| \le \varepsilon \big\} \ge -C' \varepsilon^{-2/1-2\alpha}, \quad 0 < \varepsilon \le 1,$$

where $C' > 0$ only depends on α and C.

Proof. We simply replace the ball $\{\|W\| \le \varepsilon\}$ by an appropriate cube in $\mathbb{R}^{\mathbb{N}}$ through the representation (7.2). For q integer ≥ 1, define

$$b_n = b_n(q) = \begin{cases} 2^{(\frac{3}{4}-\frac{\alpha}{2})(n-q)} & \text{if } n \le q, \\ 2^{(\frac{1}{4}-\frac{\alpha}{2})(n-q)} & \text{if } n > q. \end{cases}$$

The choice of this sequence is not unique. If $|a_{2^n+k}| \le b_n$ for every $n \ge 0$ and $k = 1, \ldots, 2^n$, and $|a_0| \le b_0$, by the triangle inequality and the hypothesis,

$$\bigg\| \sum_{m=0}^{\infty} a_m \varphi_m \bigg\| \le C_1 2^{-(\frac{1}{2}-\alpha)q}$$

for some constant C_1 only depending on α. Therefore,

$$\begin{aligned} \mathbb{P}\big\{ \|W\| \le C_1 2^{-(\frac{1}{2}-\alpha)p} \big\} &\ge \mathbb{P}\big\{ |g_0| \le b_0,\, |g_{2^n+k-1}| \le b_n,\, n \ge 0,\, k = 1, \ldots, 2^n \big\} \\ &= \mathbb{P}\big\{ |g| \le b_0 \big\} \prod_{n=0}^{\infty} \mathbb{P}\big\{ |g| \le b_n \big\}^{2^n} \end{aligned}$$

where g is a standard normal variable. Now, we simply need evaluate this infinite product. To this aim, we use that

$$\mathbb{P}\big\{ |g| \le u \big\} \ge \frac{u}{3} \quad \text{if} \quad |u| \le 1 \tag{7.3}$$

and

$$\text{(7.4)} \qquad \mathbb{P}\{|g| \le u\} \ge 1 - \frac{1}{2}\, 2\mathrm{e}^{-u^2/2} \ge \exp\bigl(-2\mathrm{e}^{-u^2/2}\bigr) \quad \text{if} \quad |u| \ge 1.$$

It easily follows, after some elementary computations, that

$$\mathbb{P}\bigl\{\|W\| \le C_1 2^{-(\frac{1}{2}-\alpha)q}\bigr\} \ge \exp\left(-C_2 2^q\right)$$

from which Theorem 7.1 immediately follows. □

The next theorem gives an upper bound of the small ball probabilities under hypotheses dual to those of Theorem 7.1.

Theorem 7.2. *Let* $0 \le \alpha < \frac{1}{2}$. *If, for some constant* $C > 0$,

$$\left\| \sum_{k=1}^{2^n} a_k \varphi_{2^n+k-1} \right\| \ge \frac{1}{C}\, 2^{-(\frac{1}{2}-\alpha)n} \left(2^{-n} \sum_{k=1}^{2^n} |a_k| \right)$$

for all real numbers $a_1, \ldots, a_{2^n}$ *and every* $n \ge 0$, *then*

$$\log \mathbb{P}\bigl\{\|W\| \le \varepsilon\bigr\} \le -\frac{1}{C''}\, \varepsilon^{-2/1-2\alpha}, \quad 0 < \varepsilon \le 1,$$

where $C'' > 0$ *only depends on* α *and* C.

Proof. First recall Anderson's inequality [An]. Let μ be a centered Gaussian measure on a Banach space E as in Chapter 4. Then, for every convex symmetric subset A of E and every x in E,

$$\text{(7.5)} \qquad \mu(x + A) \le \mu(A).$$

Note that (7.5) is an easy consequence of (1.8) or the logconcavity of Gaussian measures (1.9) (cf. [Bo1], [Bo3], [D-HJ-S]...). Indeed, the set

$$Z = \bigl\{a \in E;\ \mu(a + A) \le \mu(x + A)\bigr\}$$

is symmetric and convex by (1.9). Now $x \in Z$, so that $-x \in Z$ by symmetry, and, by convexity, $0 = \frac{1}{2}\,x + \frac{1}{2}(-x) \in Z$ which is the result (7.5). By the series representation (7.2), independence and Fubini's theorem, it follows that

$$\mathbb{P}\bigl\{\|W\| \le \varepsilon\bigr\} \le \mathbb{P}\left\{ \left\| \sum_{k=1}^{2^n} g_k\, \varphi_{2^n+k-1} \right\| \le \varepsilon \right\}$$

for every $\varepsilon > 0$ and every $n \ge 0$. Therefore, by the hypothesis,

$$\mathbb{P}\bigl\{\|W\| \le \varepsilon\bigr\} \le \mathbb{P}\left\{ \sum_{k=1}^{2^n} |g_k| \le \varepsilon C 2^{(\frac{1}{2}-\alpha)n} 2^n \right\}.$$

By Chebyshev's exponential inequality, for every integer $N \geq 1$,

$$\mathbb{P}\left\{\sum_{k=1}^{N} |g_k| \leq cN\right\} \leq \mathrm{e}^{cN}\left(\mathbb{E}(\mathrm{e}^{-|g|})\right)^N \leq \mathrm{e}^{-cN}$$

where $c > 0$ is such that $\mathrm{e}^{-2c} = \mathbb{E}(\mathrm{e}^{-|g|}) < 1$. Take then n to be the largest integer such that $\varepsilon C 2^{(\frac{1}{2}-\alpha)n} \leq c$. The conclusion easily follows. The proof of Theorem 7.2 is complete. □

The main interest of Theorems 7.1 and 7.2 lies in the examples for which the hypotheses may easily be checked. Let us consider L^p-norms $\|\cdot\|_p$, $1 \leq p < \infty$, on $[0,1]$ for which

$$\left\|\sum_{k=1}^{2^n} a_k \varphi_{2^n+k-1}\right\|_p = \frac{1}{2}\,(p+1)^{-1/p} 2^{-n/2}\left(2^{-n}\sum_{k=1}^{2^n}|a_k|^p\right)^{1/p}$$

for all real numbers $a_1,\dots,a_{2^n}$. Since

$$2^{-n}\sum_{k=1}^{2^n}|a_k| \leq \left(2^{-n}\sum_{k=1}^{2^n}|a_k|^p\right)^{1/p} \leq \max_{1\leq k\leq 2^n}|a_k|,$$

we deduce from Theorem 7.1 and 7.2 that, for some constant $C > 0$ only depending on p and every $0 < \varepsilon \leq 1$,

$$-C\varepsilon^{-2} \leq \log \mathbb{P}\big\{\|W\|_p \leq \varepsilon\big\} \leq -C^{-1}\varepsilon^{-2}. \tag{7.6}$$

More precise estimates on the constant C are obtained in [B-M].

In the same way, let $\|\cdot\|_\alpha$ be the Hölder norm of index $0 < \alpha < \frac{1}{2}$ defined by

$$\|x\|_\alpha = \sup_{0\leq s\neq t\leq 1}\frac{|x(s)-x(t)|}{|s-t|^\alpha}.$$

Again, it is easily seen that for every $a_1,\dots,a_{2^n} \in \mathbb{R}$,

$$\frac{1}{2}\,2^{-(\frac{1}{2}-\alpha)n}\max_{1\leq k\leq 2^n}|a_k| \leq \left\|\sum_{k=1}^{2^n} a_k\varphi_{2^n+k-1}\right\|_\alpha \leq \sqrt{2}\,2^{-(\frac{1}{2}-\alpha)n}\max_{1\leq k\leq 2^n}|a_k|.$$

Hence, for some constant $C > 0$ only depending on α, for every $0 < \varepsilon \leq 1$,

$$-C\varepsilon^{-2/1-2\alpha} \leq \log \mathbb{P}\big\{\|W\|_\alpha \leq \varepsilon\big\} \leq -C^{-1}\varepsilon^{-2/1-2\alpha}. \tag{7.7}$$

This result is due to P. Baldi and B. Roynette [B-R].

Note that the supnorm may be included in either $p = \infty$ in (7.6) or $\alpha = 0$ in (7.7) so that we recover (7.1) with these elementary arguments, with however worse constants. More examples may be treated by these methods such as Besov's norm or Sobolev norms on the Wiener space. We refer to [St1] and [Li-S] for more details along these lines.

To try to investigate some further cases with these tools, let us consider, on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a sequence $(G_m)_{m\in\mathbb{N}}$ of independent standard one dimensional Brownian motions (on $[0,1]$). Replace then, in the series representation (7.2), the standard Gaussian sequence by this sequence of independent Brownian motions. We define in this way a Brownian motion S with values in $C_0([0,1])$ and with "reference measure" the Wiener measure itself. This is one way of defining the Wiener sheet which thus turns out to be a centered Gaussian process $S = (S(s,t))_{s,t\in[0,1]}$ with covariance

$$\mathbb{E}\big(S(s,t)S(s',t')\big) = \min(s,s')\min(t,t').$$

In this framework, we may thus ask for the behavior of $\mathbb{P}\{\|S\|_\infty \le \varepsilon\}$, $0 < \varepsilon \le 1$, for the supnorm on $[0,1]^2$. Since this norm is also the supremum norm of W considered as a one dimensional process with values in $C_0([0,1])$ (equipped with the supnorm), some of the preceding material may be used in this investigation. In particular, we can replace, in the proof of Theorem 7.1, (7.3) by the small ball behavior of Brownian motion (7.1). By Theorem 4.1, the large ball behavior (7.4) is unchanged at the exception of possibly different numerical constants. The argument of Theorem 7.3 then implies, exactly in the same way, that for some constant $C > 0$, and every $0 < \varepsilon \le 1$,

$$\log\mathbb{P}\big\{\|S\|_\infty \le \varepsilon\big\} \ge -C\varepsilon^{-2}\big(\log\varepsilon^{-1}\big)^3.$$

A similar vector valued extension of Theorem 7.2 however only yields that

$$\log\mathbb{P}\big\{\|S\|_\infty \le \varepsilon\big\} \le -C^{-1}\varepsilon^{-2}\big(\log\varepsilon^{-1}\big).$$

These estimates, which go back to M. A. Lifshits [Lif-T] (see also [Bas]), only rely on the small ball behavior (7.1) and are best possible among all Gaussian measures having this small ball behavior. For the special case of Wiener measure and the Wiener sheet, M. Talagrand [Ta15] however proved the striking following theorem. The proof is based on a new wavelet decomposition of the space $\mathrm{L}^2([0,1]^2)$ and various combinatorial arguments from Banach space theory. The method does not allow any precise information on constants. We refer to [Ta15] for the proof.

Theorem 7.3. *There is a numerical constant $C > 0$ such that, for every $0 < \varepsilon \le 1$,*

$$-C\varepsilon^{-2}\big(\log\varepsilon^{-1}\big)^3 \le \log\mathbb{P}\big\{\|S\|_\infty \le \varepsilon\big\} \le -C^{-1}\varepsilon^{-2}\big(\log\varepsilon^{-1}\big)^3.$$

In this framework of small ball probabilities for Gaussian measures, let us now come back to some of the concentration inequalities of Chapters 2 and 4. There, we studied inequalities for general sets A and their enlargements A_r. Now, we try to take advantage of some geometric structures on A, such as for example being a ball (convex and symmetric with respect to the origin). In a first step, we will notice how some of the preceding tools may be applied successfully to improve, for example, a statement such as Theorem 4.4. In particular, we aim to prove inequalities for subsets A with small measure and to be able to keep the dependence in this measure. The next statement (cf. [Ta8]) is a first example of what can be accomplished for various norms on the Wiener space. Recall the unit ball $\mathcal{K}$ of the Cameron-Martin Hilbert space $\mathcal{H}$ of absolutely continuous functions on $[0,1]$ whose almost everywhere derivative is in L^2.

Theorem 7.4. *Let $\|\cdot\|$ be a norm on $C_0([0,1])$ for which Wiener measure is a Radon measure. Denote by U the unit ball of $\|\cdot\|$. Let furthermore $0 \le \alpha < \frac{1}{2}$ and assume that, for some constant $C > 0$,*

$$\Big\| \sum_{k=1}^{2^n} a_k \varphi_{2^n+k-1} \Big\| \le C 2^{-(\frac{1}{2}-\alpha)n} \max_{1\le k\le 2^n} |a_k|$$

for all real numbers $a_1, \ldots, a_{2^n}$ and every $n \ge 0$. Then, for every $\varepsilon > 0$ and every $r \ge 0$,

$$\mathbb{P}\{W \in \varepsilon U + r\mathcal{K}\} \ge 1 - \exp\bigg(\frac{C'}{\varepsilon^{2/1-2\alpha}} - \frac{\varepsilon r}{2\sigma} - \frac{r^2}{2\sigma^2}\bigg)$$

where $C' > 0$ only depends on α and C where we recall that $\sigma = \sup_{x\in\mathcal{K}} \|x\|$.

Proof. We take again the notation of Theorem 7.1. First note that since $\mathcal{K} \subset \sigma U$,

$$\varepsilon U + r\mathcal{K} \supset \frac{\varepsilon}{2} U + \Big(r + \frac{\varepsilon}{2\sigma}\Big)\mathcal{K}.$$

(The choice of $\varepsilon/2$ is rather arbitrary here.) Set $r' = r + \frac{\varepsilon}{2\sigma}$. Recall the sequence $(b_n)_{n\in\mathbb{N}}$ of the proof of Theorem 7.1 which depends on some integer $q \ge 1$. Define a sequence of real numbers $(c_m)_{m\in\mathbb{N}}$ by setting

$$c_0 = b_0, \quad c_{2^n+k-1} = b_n \quad \text{for all} \quad k = 1, \ldots, 2^n,\ n \ge 0.$$

Consider the set $V = V_q$ of all functions φ on $[0,1]$ that can be written as $\varphi = \sum_{m=0}^{\infty} a_m \varphi_m$ where $|a_m| \le c_m$ for every m. By the hypothesis on the norm $\|\cdot\|$ and the triangle inequality, $V \subset C_1 2^{-(\frac{1}{2}-\alpha)q} U$ for some constant $C_1 > 0$. Therefore, if q is the smallest integer such that $2C_1 2^{-(\frac{1}{2}-\alpha)q} \le \varepsilon$, then $\varepsilon U + r\mathcal{K} \supset V + r'\mathcal{K}$. Hence, by the series representation (7.2),

$$\mathbb{P}\{W \in \varepsilon U + r\mathcal{K}\} \ge \mathbb{P}\{W \in V + r'\mathcal{K}\} = \gamma_\infty(Q + r'\sigma^{-1}B_2)$$

where γ_∞ is the canonical Gaussian measure on $\mathbb{R}^{\mathbb{N}}$, B_2 the unit ball of the reproducing kernel of γ_∞, that is the unit ball of ℓ^2, and Q the "cube" in $\mathbb{R}^{\mathbb{N}}$

$$Q = \big\{x = (x_m)_{m\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}; |x_m| \le c_m,\ m \in \mathbb{N}\big\}.$$

Consider the function on $\mathbb{R}^{\mathbb{N}}$ given by $d(x) = \inf\{u \ge 0; x \in Q + uB_2\}$. Note that $\gamma_\infty(Q + uB_2) = \gamma_\infty(d < u)$. By Chebyshev's inequality,

$$\gamma_\infty(d \ge u) \le \mathrm{e}^{-u^2/2} \int \mathrm{e}^{d^2/2} d\gamma_\infty.$$

For every $m \geq 0$, let $d_m(x) = (|x_m| - c_m)^+$. Then

$$\begin{aligned}\int \mathrm{e}^{d_m^2/2} d\gamma_\infty &= \frac{2}{\sqrt{2\pi}} \int_0^\infty \exp\bigg(\frac{1}{2}\big((|t| - c_m)^+\big)^2 - \frac{t^2}{2}\bigg) dt \\ &\leq 1 + \frac{2}{\sqrt{2\pi}} \int_{c_m}^\infty \exp\bigg(\frac{1}{2}(|t| - c_m)^2 - \frac{t^2}{2}\bigg) dt \\ &\leq 1 + \int_{c_m}^\infty \exp\bigg(-c_m t + \frac{c_m^2}{2}\bigg) dt = 1 + \frac{1}{c_m}\, \mathrm{e}^{-c_m^2/2}.\end{aligned}$$

Now, the nice feature of this geometric construction is that $d^2 = \sum_{m=0}^\infty d_m^2$. Therefore, it follows from the preceding that, for every $u \geq 0$,

$$\gamma_\infty(d \geq u) \leq \prod_{m=0}^\infty \bigg(1 + \frac{1}{c_m}\, \mathrm{e}^{-c_m^2/2}\bigg) \mathrm{e}^{-u^2/2}.$$

To conclude the proof, we need simply estimate this infinite product. By the definition of the sequence $(c_m)_{m \in \mathbb{N}}$, we see that

$$\prod_{m=0}^\infty \bigg(1 + \frac{1}{c_m}\, \mathrm{e}^{-c_m^2/2}\bigg) = \prod_{n=0}^\infty \bigg(1 + \frac{1}{b_n}\, \mathrm{e}^{-b_n^2/2}\bigg)^{2^n}.$$

Now, the very definition of the sequence $(b_n)_{n \in \mathbb{N}}$ implies, after elementary, though somewhat tedious, computations that the preceding infinite product is bounded above by $\exp(C_2 2^p)$ for some numerical constant $C_2 > 0$. By the choice of q, this completes the proof of Theorem 7.4 . □

With respect to the classical isoperimetric and concentration inequalities usually stated for sets with measure larger than $\frac{1}{2}$, we note here that the probability $\mathbb{P}\{W \in \varepsilon U\}$ can be very small as $\varepsilon \to 0$. Moreover, according to Theorem 7.1, the first term in the exponential extimate of Theorem 7.4 is precisely the order of $\mathbb{P}\{W \in \varepsilon U\}$. Theorem 7.4 applies to the supnorm and the Hölder norms and may be used in the study of rates of convergence in Strassen's law of the iterated logarithm. Let for example

$$Z_n(t) = \bigg(\frac{W(nt)}{\sqrt{2n\mathrm{LL}n}}\bigg)_{t \in [0,1]}, \quad n \geq 1,$$

where $\mathrm{LL}n = \log\log n$ if $n \geq 3$, $\mathrm{LL}n = 1$ if $n = 1, 2$. It is shown in [Ta8] using Theorem 7.4 that, almost surely,

$$0 < \limsup_{n \to \infty} (\mathrm{LL}n)^{2/3} d\big(Z_n, \mathcal{K}\big) < \infty$$

where $d(\cdot, \mathcal{K})$ is the uniform distance to the Strassen set (Cameron-Martin unit ball) on $C_0([0,1])$. See also [Gri] for an alternate proof and [Ta9] for further results.

Recently, M. Talagrand [Ta9] proved a deep extension of Theorem 7.4 in the abstract setting of enlarged balls. We now would like to present this statement. We will state and prove the main result for the canonical Gaussian measure γ_n on

$\mathbb{R}^n$. As in the preceding chapters, this is again the main inequality and standard tools may then be used to extend it to arbitrary Gaussian measures as in Chapter 4. The isoperimetric and concentration inequalities for γ_n yield powerful bounds of the measure of an enlarged set A_r, especially when r is large. However, the extremal sets of Gaussian isoperimetry are the half-spaces and it may well be that the concentration properties could be sharpened for sets with special geometrical structures such as for example convex symmetric bodies. The next theorem [Ta10] answers this problem.

Theorem 7.5. *Let C be a closed convex symmetric subset of $\mathbb{R}^n$. Assume that the polar C° of C may be covered by N sets $(T_i)_{1\le i\le N}$ such that $\int \sup_{y\in T_i}\langle x,y\rangle d\gamma_n(x) \le \frac{1}{2}$. Then, for every $r\ge 1$,*

$$\gamma_n(C_r) \ge 1 - 4N\log(\mathrm{e}r)\mathrm{e}^{-r^2/2}.$$

Proof. Denote for simplicity by B the closed Euclidean unit ball in $\mathbb{R}^n$. Since C is closed and B is compact, $C_r = C+rB$ is closed. By the bipolar theorem, $C+rB = U^\circ$ where $U = (C+rB)^\circ$. By definition,

$$U = \big\{x\in\mathbb{R}^n; \forall\, y\in C, \forall\, z\in B, \big|\langle x,y\rangle + r\langle x,z\rangle\big| \le 1\big\}.$$

Setting $\|x\|_C = \sup_{y\in C}\langle x,y\rangle = \sup_{y\in C}|\langle x,y\rangle|$, we see that

$$U = \big\{x\in\mathbb{R}^n; \|x\|_C + r|x| \le 1\big\}. \tag{7.8}$$

Observe also by the definition of the polar that $x\in \|x\|_C C^\circ$. If T is a subset of $\mathbb{R}^n$, we set

$$E(T) = \int \sup_{y\in T}\langle x,y\rangle d\gamma_n(x).$$

Let p_0 be the largest integer p such that $2^{p-1}\le r^2$. In particular, $p_0 \le 1+4\log r$. Now, set

$$U_0 = \big\{x\in U; |x| \ge r^{-1}(1-r^{-2})\big\}$$

and, for $1\le p\le p_0$,

$$U_p = \big\{x\in U; r^{-1}(1-r^{-2}2^p) \le |x| \le r^{-1}(1-r^{-2}2^{p-1})\big\}.$$

Thus we have $U\subset \bigcup_{0\le p\le p_0} U_p$. Moreover, for $x\in U_p$, by (7.8), $\|x\|_C \le r^{-2}2^p$. Therefore, $U_p\subset r^{-2}2^pC^\circ$. It thus follows from the hypothesis on C° that U_p can be covered by subsets $(T_{p,i})_{1\le i\le N}$ where $T_{p,i} = r^{-2}2^pT_i\cap U_p$. Hence $E(T_{p,i})\le r^{-2}2^{p-1}$.

The essential step of the proof is concentration. From (4.3) for example, we get that, for every subset T of $\mathbb{R}^n$ and every $t\ge E(T)$,

$$\gamma_n\big(x; \sup_{y\in T}\langle y,x\rangle \ge t\big) \le \exp\bigg(-\frac{(t-E(T))^2}{2\sigma^2}\bigg) \tag{7.9}$$

where $\sigma = \sigma(T) = \sup_{x \in T} |x|$. Note that for $p \le p_0$, $E(T_{p,i}) \le r^{-2}2^{p-1} \le 1$. Hence, using (7.9) for $t = 1$, we get, for every $0 \le p \le p_0$, $1 \le i \le N$,

$$\gamma_n\big(x; \sup_{y \in T_{p,i}} \langle y, x\rangle \ge 1\big) \le \exp\bigg(-\frac{(1 - r^{-2}2^{p-1})^2}{2\sigma(T_{p,i})^2}\bigg). \tag{7.10}$$

We first consider the case $p = 0$. We have $\sigma(T_{0,i}) \le r^{-1}$ so that, by summation over $1 \le i \le N$,

$$\gamma_n\big(x; \sup_{y \in U_0} \langle y, x\rangle \ge 1\big) \le N \exp\bigg(-\frac{r^2}{2}\big(1 - r^{-2}\big)^2\bigg) \le N\mathrm{e}\,\mathrm{e}^{-r^2/2}.$$

For $p \ge 1$, by definition of U_p, we have $\sigma(T_{p,i}) \le r^{-1}(1 - r^{-2}2^{p-1})$, so that, by summation of (7.10) over $1 \le i \le N$, we get

$$\gamma_n\big(x; \sup_{y \in U_p} \langle y, x\rangle \ge 1\big) \le N\,\mathrm{e}^{-r^2/2}.$$

Now, summation over $0 \le p \le p_0$ and the fact that $p_0 \le 1 + 4\log r$ yield

$$\begin{aligned}\gamma_n\big(x; \sup_{y \in U} \langle y, x\rangle \ge 1\big) &\le N(\mathrm{e} + 1 + 4\log r)\,\mathrm{e}^{-r^2/2}\\ &\le 4N\log(\mathrm{e}r)\,\mathrm{e}^{-r^2/2}.\end{aligned}$$

Since $C + rB = U^\circ$, the result follows. The proof of Theorem 7.5 is complete. □

Of course, Theorem 7.5 can be useful in applications only if the number N of the statement may be appropriately bounded. We will not go far in the technical details, but one of the main conclusions of the important paper [Ta9] is that N may actually be controlled by the behavior of $\gamma_n(\varepsilon A)$ for the small values of $\varepsilon > 0$. Actually, this observation is strongly related to a remarkable result of J. Kuelbs and W. Li [K-L2] connecting the small ball probabilities to some entropy numbers related to N. We now turn to this discussion. Related work of M. A. Lifshits [Lif1] deals with the geometric tool of Kolmogorov's widths.

Given two (convex) sets A and B in $\mathbb{R}^n$ (or more generaly in a linear vector space), denote by $N(A, B)$ the smallest number of translates of B which are needed to cover A. Now, let, as in Theorem 7.5, C be a closed convex symmetric set and let B be the Euclidean unit ball in $\mathbb{R}^n$. If $x \in \mathbb{R}^n$ is such that $(x + \varepsilon B) \cap C^\circ \ne \emptyset$, for $y \in (x + \varepsilon B) \cap C^\circ$ we have

$$(x + \varepsilon B) \cap C^\circ \subset (y + 2\varepsilon B) \cap C^\circ \subset y + (2\varepsilon B \cap 2C^\circ).$$

Hence, if ε is such that $E(C^\circ \cap \varepsilon B) \le \frac{1}{8}$, then the number N in Theorem 7.5 satisfies $N \le N(C^\circ, \varepsilon B)$. Now, it has been observed in local theory of Banach spaces [TJ] that the growth of the entropy numbers $N(C^\circ, \varepsilon B)$ is very similar to the growth of the dual entropy numbers $N(B, \varepsilon C)$ (cf. [L-T2], p. 82-83). The observation of J. Kuelbs and W. Li is precisely that the behavior of the small ball probabilities $\gamma_n(\varepsilon A)$ is related to these dual entropy numbers $N(B, \varepsilon C)$. They established namely the following theorem. One crucial argument in the proof is the Gaussian isoperimetric inequality. A prior partial result appeared in [Go2].

Theorem 7.6. *Under the preceding notation, if C is compact, convex and symmetric, and if $t = (2\log(\gamma_n(C)^{-1}))^{-1/2} > 0$, then*

$$\frac{1}{2\gamma_n(2C)} \le N(B, tC) \le \frac{1}{2\gamma_n(C/2)^2}\,.$$

Proof. First note, as a consequence of Cameron-Martin's formula (in finite dimension) and Anderson's inequality (7.5), that for every $x \in \mathbb{R}^n$,

$$\mathrm{e}^{-|x|^2/2}\gamma_n(C) \le \gamma_n(x + C) \le \gamma_n(C). \tag{7.11}$$

Consider now $u > 0$ and a finite subset F of uB such that for any two distinct points of F, the translates of C by these points are disjoints. By (7.11), for every x in F,

$$\gamma_n(x + C) \ge \mathrm{e}^{-u^2/2}\gamma_n(C).$$

It follows that $\mathrm{Card}(F) \le \gamma_n(C)^{-1}\mathrm{e}^{u^2/2}$. When F is maximal, the sets $(x + 2C)_{x\in F}$ cover uB so that

$$N(uB, 2C) \le \gamma_n(C)^{-1}\mathrm{e}^{u^2/2}.$$

(When $\gamma_n(C) \ge \frac{1}{2}$, this is how the dual Sudakov inequality is proved in [L-T2], p. 83.) If we choose $u = t^{-1} = (2\log(\gamma_n(C)^{-1}))^{1/2}$, we have $N(uB, 2C) \le \gamma_n(C)^{-2}$. Since $N(uB, C) = N(B, tC)$, the right hand side of Theorem 7.6 follows by replacing C by $\frac{1}{2}C$.

Conversely, by (7.11) again,

$$N(uB, C)\gamma_n(2C) \ge N(C + uB, 2C)\gamma_n(2C) \ge \gamma_n(C + uB).$$

Now, by the isoperimetric inequality (Theorem 1.3),

$$\Phi^{-1}\big(\gamma_n(C + uB)\big) \ge \Phi^{-1}\big(\gamma_n(C)\big) + u.$$

Let again $u = t^{-1} = (2\log(\gamma_n(C)^{-1}))^{1/2}$. Since $\Phi(-u) \le \mathrm{e}^{-u^2/2} = \gamma_n(C)$, the isoperimetric inequality implies that $\Phi^{-1}(\gamma_n(C+uB)) \ge 0$ that is, $\gamma_n(C+uB) \ge \frac{1}{2}$. The left hand side of the inequality of the theorem is thus also satisfied. The proof is complete. □

If we set

$$\varphi(\varepsilon) = \left(2\log\frac{1}{\gamma_n(\varepsilon C)}\right)^{1/2}, \quad \varepsilon > 0,$$

we see from Theorem 7.6 that

$$\frac{1}{2}\exp\left(\frac{\varphi(2\varepsilon)^2}{2}\right) \le N\Big(B, \frac{\varepsilon}{\varphi(\varepsilon)}\Big) \le \exp\Big(\varphi\Big(\frac{\varepsilon}{2}\Big)^2\Big).$$

Therefore, if φ is regularly varying, its behavior is essentially given by the behavior of the entropy numbers $N(B, \varepsilon C)$ (and conversely). Much more on the structure of C is thus involved with respect for example with the large ball behavior (cf. Chapter 4). While Theorem 7.6 and its proof are presented in finite dimension, the infinite dimensional extension yields a rather precise equivalence between small ball probabilities for a Gaussian measure μ on a Banach space E and the entropy numbers $N(\mathcal{K}, \varepsilon U)$ where $\mathcal{K}$ is the unit ball of the reproducing kernel Hilbert space $\mathcal{H}$ and U the unit ball of E. It has been shown by J. Kuelbs and W. Li [K-L2] to have striking consequences in approximation theory. For example, using the small ball behaviors for Wiener measure, we see that if $\mathcal{K}$ is the unit ball of the Cameron-Martin Hilbert space and C the L^p-ball on $([0,1], dt)$, $1 \leq p \leq \infty$, then

$$\log N(\mathcal{K}, \varepsilon C) \sim \frac{1}{\varepsilon} \quad \text{as } \varepsilon \to 0.$$

Theorem 7.3 similarly shows that for $\mathcal{K}$ the unit ball of the Cameron-Martin space associated to the Wiener sheet and for C the uniform unit ball on $C([0,1]^2; \mathbb{R})$,

$$\log N(\mathcal{K}, \varepsilon C) \sim \frac{1}{\varepsilon}\Big(\log \frac{1}{\varepsilon}\Big)^{3/2} \quad \text{as } \varepsilon \to 0.$$

This deep connection between entropy numbers and small ball probabilities is further investigated in [K-L2] and [Ta9]. In particular, in [Ta9], the author obtains very general rates for the variables $(2\log n)^{-1/2} X_n$ to cluster to $\mathcal{K}$, when $(X_n)_{n \in \mathbb{N}}$ is a sequence of independent identically distributed sequence with distribution μ. These rates depend only on the behavior of the small ball probabilities $\mu(\varepsilon U)$. These results have applications to rates of convergence in Strassen's law of the iterated logarithm for Brownian motion. Prior results on the convergence of $(2\log n)^{-1/2} X_n$ to $\mathcal{K}$ at the origin of this study are due to V. Goodman [Go1].

In [Ta9], M. Talagrand also established a general lower bound on supremum of Gaussian processes under entropy conditions. At the present time, it is one of the only few general results available in this subject of small ball probabilities. We briefly describe one simple statement. Let $(X_t)_{t \in T}$ be a (centered) Gaussian process as in Chapter 6. Recall also from this chapter the entropy numbers $N(T, d; \varepsilon)$, $\varepsilon > 0$, for the Dudley metric $d(s,t) = (\mathbb{E}|X_s - X_t|^2)^{1/2}$, $s, t \in T$. Assume that there is a nonnegative function ψ on $\mathbb{R}_+$ such that

$$N(T, d; \varepsilon) \leq \psi(\varepsilon), \quad \varepsilon > 0, \tag{7.11}$$

and such that for some constants $1 < c_1 \leq c_2 < \infty$ and all $\varepsilon > 0$

$$c_1 \psi(\varepsilon) \leq \psi\Big(\frac{\varepsilon}{2}\Big) \leq c_2 \psi(\varepsilon). \tag{7.12}$$

We thus have in mind a power type behavior $\psi(\varepsilon) = \varepsilon^{-a}$ of the entropy numbers. Then, for some $K > 0$ and every $\varepsilon > 0$,

$$\mathbb{P}\Big\{ \sup_{s,t \in T} |X_s - X_t| \leq \varepsilon \Big\} \geq \exp\big(-K\psi(\varepsilon)\big). \tag{7.13}$$

We prove this result following the notations introduced in the previous chapter. Let n_0 be the largest n in $\mathbb{Z}$ such that $2^{-n} \geq D(T)$ where $D(T)$ is the diameter of T, assumed to be finite (we may actually start with T finite). For every $n \geq n_0$, consider a subset T_n of T of cardinality $N(n) = N(T, d; 2^{-n})$ such that each point of T is within distance 2^{-n} of T_n. We let $T_{n_0} = \{t_0\}$ where t_0 is any fixed point in T. For $n > n_0$, choose $s_{n-1}(t) \in T_{n-1}$ such that $d(t, s_{n-1}(t)) \leq 2^{-n+1}$ and set

$$\mathcal{Y} = \{X_t - X_{s_{n-1}(t)}; t \in T_n\}.$$

Note that $\|Y\|_2 \leq 2^{-n+1}$ for every Y in $\mathcal{Y}$. Clearly, each X_t can be written as

$$X_t = X_{t_0} + \sum_{n > n_0} Y_n$$

where $Y_n \in \mathcal{Y}$, $n > n_0$. Therefore, if $(b_n)_{n > n_0}$ is a sequence of positive numbers with $\sum_{n > n_0} b_n \leq \frac{u}{2}$, $u > 0$,

$$\begin{aligned} \mathbb{P}\Big\{ \sup_{s,t \in T} |X_s - X_t| \leq u \Big\} &\geq \mathbb{P}\big\{ \forall\, n > n_0, \forall\, Y \in \mathcal{Y}, |Y_n| \leq b_n \big\} \\ &\geq \prod_{n > n_0} \big(\mathbb{P}\{ |g| \leq b_n 2^{n-1} \} \big)^{N(n)} \end{aligned} \tag{7.14}$$

where g is a standard normal variable. We used here the following consequence of the main inequality of [Kh], [Sc], [Si]... (see (7.16) below): if $(Z_1, \ldots, Z_n)$ is a (centered) Gaussian random vector, for every $\lambda_1, \ldots, \lambda_n \geq 0$,

$$\mathbb{P}\big\{ |Z_1| \leq \lambda_1, \ldots, |Z_n| \leq \lambda_n \big\} \geq \prod_{i=1}^{n} \mathbb{P}\big\{ |Z_i| \leq \lambda_i \big\}.$$

Let q be an integer with $q > n_0$ and set

$$b_n = b_n(q) = \begin{cases} 2^{-\frac{3q}{2} + \frac{n}{2} + 1} & \text{if } n_0 < n \leq q, \\ 2^{-\frac{q}{2} - \frac{n}{2} + 1} & \text{if } n > q. \end{cases}$$

Then $\sum_{n > n_0} b_n \leq 2^{-q+3}$. Apply then (7.14) with $u = 2^{-q+3}$. Using (7.3) and (7.4) and the hypothesis $N(n) \leq \psi(2^{-n})$, we get

$$\mathbb{P}\Big\{ \sup_{s,t \in T} |X_s - X_t| \leq u \Big\} \geq \prod_{n_0 < n \leq q} \big(3^{-1} 2^{-3(n-q)/2} \big)^{\psi(2^{-n})} \prod_{n > q} \exp\big(-2\psi(2^{-n}) \mathrm{e}^{-2^{n-q}} \big).$$

Now, by (7.12),

$$\begin{aligned} \sum_{n_0 < n \leq q} \psi(2^{-n}) \log\big(3^{-1} 2^{3(n-q)/2} \big) &\leq \psi(2^{-q}) \sum_{n_0 < n \leq q} c_1^{n-q} \log\big(3^{-1} 2^{3(n-q)/2} \big) \\ &\leq K(c_1) \psi(2^{-q}) \end{aligned}$$

while
$$\sum_{n>q}\psi(2^{-n})\,\mathrm{e}^{-2^{n-q}} \le \psi(2^{-q})\sum_{n>q} c_2^{n-q}\mathrm{e}^{-2^{n-q}} \le K(c_2)\psi(2^{-q})$$

where $K(c_1), K(c_2) > 0$ only depend on c_1 and c_2 respectively. It follows that

$$\mathbb{P}\Big\{\sup_{s,t\in T}|X_s - X_t| \le 2^{-q+3}\Big\} \ge \exp\big(-K\psi(2^{-q})\big)$$

where $q > n_0$. Let $\varepsilon \le 8D(T)$ and let $q > n_0$ be the largest integer such that $2^{-q+4} \ge \varepsilon$ ($\varepsilon \le 8D(T) \le 2^{-n_0+3}$, $2^{-q+3} \le \varepsilon$). Then

$$\mathbb{P}\Big\{\sup_{s,t\in T}|X_s - X_t| \le \varepsilon\Big\} \ge \exp\big(-K\psi(2^{-q})\big) \ge \exp\big(-K\psi(\varepsilon)\big).$$

When $\varepsilon \ge 8D(T)$, by concentration,

$$\mathbb{P}\Big\{\sup_{s,t\in T}|X_s - X_t| \le \varepsilon\Big\} \ge 1 - 2\exp\bigg(\frac{\varepsilon^2}{2D(T)^2}\bigg) \ge \frac{1}{2} \ge \exp\big(-\psi(\varepsilon)\big)$$

since $\psi(\varepsilon) \ge N(T,d;\varepsilon) \ge 1$. (7.13) thus is established.

To conclude this chapter, we mention some related correlation and conditional inequalities and their applications. These results have been used recently in various topological questions on Wiener space briefly discussed below.

The next inequality seems to mix small ball and large ball behaviors and might be of some interest in other contexts. It is related to conjecture (7.17) below. Let $W = (W(t))_{t\ge 0}$ be Brownian motion starting at the origin with values in $\mathbb{R}^d$. By Lévy's modulus of continuity of Brownian motion, one may consider some stronger topologies on the Wiener space $C_0([0,1];\mathbb{R}^d)$, such as Hölder topologies. For every function $x : [0,1] \to \mathbb{R}^d$, recall the Hölder norm of index $0 < \alpha < 1$ defined as

$$\|x\|_\alpha = \sup_{0\le s\ne t\le 1}\frac{|x(s) - x(t)|}{|s-t|^\alpha}\,.$$

It is known, and due to Z. Ciesielski [Ci1], that these Hölder norms are equivalent to sequence norms. More precisely, for every continuous function $x : [0,1] \to \mathbb{R}^d$ such that $x(0) = 0$, let, for $m = 2^n + k - 1$, $n \ge 0$, $k = 1,\ldots,2^n$,

$$\xi_m(x) = \xi_{2^n+k-1}(x) = 2^{n/2}\Big[2x\Big(\frac{2k-1}{2^{n+1}}\Big) - x\Big(\frac{k}{2^n}\Big) - x\Big(\frac{k-1}{2^n}\Big)\Big]$$

and $\xi_0(x) = x(1)$, be the evaluation of x in the Schauder basis on $C_0([0,1];\mathbb{R}^d)$. Set

$$\|x\|'_\alpha = \sup_{m\ge 0}(m+1)^{\alpha-\frac{1}{2}}\big|\xi_m(x)\big|.$$

Then, for every $0 < \alpha < 1$, there exists $C_\alpha^d > 0$ such that, for all $x \in C_0([0,1];\mathbb{R}^d)$,

$$(C_\alpha^d)^{-1}\|x\|'_\alpha \le \|x\|_\alpha \le C_\alpha^d \|x\|'_\alpha. \tag{7.15}$$

Note also that Wiener measure is a Radon measure on the subspace of the space of Hölder functions x such that

$$\lim_{\eta\to 0} \sup_{\substack{|s-t|\le\eta \\ 0\le s\ne t\le 1}} \frac{|x(s)-x(t)|}{|s-t|^\alpha} = 0.$$

The next proposition is the main conditional Gaussian inequality we would like to emphasize. It evaluates large oscillations of the Brownian paths conditionally on the fact that these paths are bounded [BA-G-L].

Proposition 7.7. *Let $0 < \alpha < \frac{1}{2}$. There exists a constant $C > 0$ only depending on d and α such that for every $u > 0$ and $v > 0$,*

$$\mathbb{P}\big\{\|W\|_\alpha \ge u \,\big|\, \|W\|_\infty \le v\big\} \le C \max\bigg(1, \Big(\frac{u}{v}\Big)^{1/\alpha}\bigg) \exp\bigg(-\frac{u^{1/\alpha}}{Cv^{(1/\alpha)-2}}\bigg).$$

Proof. We use (7.15) to write that, for $u, v > 0$,

$$\begin{aligned}\mathbb{P}\big\{\|W\|'_\alpha \ge u \,\big|\, \|W\|_\infty \le v\big\} &\le \sum_{m\ge 0} \mathbb{P}\big\{\big|\xi_m(W)\big| \ge u(m+1)^{\frac{1}{2}-\alpha} \,\big|\, \|W\|_\infty \le v\big\} \\ &\le \sum_{m\ge m_0} \mathbb{P}\big\{\big|\xi_m(W)\big| \ge u(m+1)^{\frac{1}{2}-\alpha} \,\big|\, \|W\|_\infty \le v\big\}\end{aligned}$$

where $m_0 = \max(0, (u/4v)^{1/\alpha} - 1)$ since, on $\{\|W\|_\infty \le v\}$, $|\xi_m(W)| \le 4v\sqrt{m+1}$. Now, if $a > 0$ and if A is a convex symmetric subset of $\mathbb{R}^n$, it has been shown in [Kh], [Si], [Sco]... (see also [DG-E-...]) that

$$\gamma_n(A \cap S) \ge \gamma_n(A)\gamma_n(S) \tag{7.16}$$

where, as usual, γ_n is the canonical Gaussian measure on $\mathbb{R}^n$ and where S is the strip $\{x \in \mathbb{R}^n; |x_1| \le a\}$. Since the ξ_m are continuous linear functionals on the Wiener space, a simple finite dimensional approximation on (7.16) (in the spirit, for example, of the approximation procedures described in Chapter 4) then shows that

$$\mathbb{P}\big\{\big|\xi_m(W)\big| \ge u(m+1)^{\frac{1}{2}-\alpha} \,\big|\, \|W\|_\infty \le v\big\} \le \mathbb{P}\big\{\big|\xi_m(W)\big| \ge u(m+1)^{\frac{1}{2}-\alpha}\big\}$$

for every m. Hence,

$$\mathbb{P}\big\{\|W\|'_\alpha \ge u \,\big|\, \|W\|_\infty \le v\big\} \le \sum_{m\ge m_0} \mathbb{P}\big\{\big|\xi_m(W)\big| \ge u(m+1)^{\frac{1}{2}-\alpha}\big\}.$$

Now, the variables $\xi_m(W)$ are distributed according to γ_d on $\mathbb{R}^d$. By the classical Gaussian exponential bound,

$$\mathbb{P}\big\{\|W\|'_\alpha \geq u \,\big|\, \|W\|_\infty \leq v\big\} \leq \sum_{m \geq m_0} \exp\biggl(-\frac{1}{C_d}\, u^2 (m+1)^{1-2\alpha}\biggr)$$

where $C_d > 0$ only depends on d. The conclusion then easily follows after some elementary computations. Proposition 7.7 is established. □

It might be worthwhile noting that we obtain a weaker, although already useful result, by replacing in Proposition 7.7 the conditional probability by the probability of the intersection, that is

$$\mathbb{P}\big\{\|W\|_\alpha \geq u, \|W\|_\infty \leq v\big\}.$$

The proof for this quantity is in fact easier since it does not use (7.16). M. A. Lifshits recently mentioned to me that the bound of Proposition 7.7 is actually two-sided at the logarithmic scale as the ratio $u^{1/\alpha}/v^{(1/\alpha)-2}$ is large. His argument is based on a delicate partitioning and clever use of the Markov property. One may ask for a general version of Proposition 7.7 dealing with some arbitrary norms on a Gaussian space.

In the proof of Proposition 7.7, we made crucial use of the correlation inequality (7.16). More generally than (7.16), one may ask whether, given two symmetric convex bodies A and B in $\mathbb{R}^n$,

$$\gamma_n(A \cap B) \geq \gamma_n(A)\gamma_n(B). \tag{7.17}$$

This was established when $n = 2$ by L. Pitt [Pit], and thus for arbitrary n when B is a symmetric strip in [Kh], [Si], [Sco] (see also [DG-E-...], [Bo7]...). The general case is so far open.

Proposition 7.7 was used recently in [BA-G-L] to extend the Stroock-Varadhan support of a diffusion theorem (cf. [I-W]) to the stronger Hölder topology of index $0 < \alpha < \frac{1}{2}$ on Wiener space. This result was obtained independently in [A-K-S] and [M-SS] by other methods. It was further used in [BA-L2] to extend to this topology the Freidlin-Wentzell large deviation principle for small perturbations of dynamical systems. These results may appear as attempts to understand the role of the topolgy in these classical statements. In this direction, the support theorem is established in [G-N-SS] (see also [Me]) for fairly general modulus norms (related to the description of the natural functional norms on the Brownian paths given in [Ci2]). In the context of large deviations, one may wonder for example whether some analogue of Theorem 4.5 holds for diffusion processes. An even more precise result would be a concentration inequality for diffusions.

The next theorem is due to C. Borell [Bo4] in 1977 with a proof using the logconcavity (1.9) of Gaussian measures. We follow here the alternate proof of [S-Z1] based on the correlation inequality (7.16). This result may be used to establish conditional exponential inequalities that allow one to prove existence of Onsager-Machlup functionals for tubes around every element in the Cameron-Martin space.

Theorem 7.8. *Let $(E, \mathcal{H}, \mu)$ be an abstract Wiener space and let $h \in \mathcal{H}$. Then*

$$\lim_{\varepsilon \to 0} \frac{\mu(h + B(0, \varepsilon))}{\mu(B(0, \varepsilon))} = \mathrm{e}^{-|h|^2/2}$$

where $B(0, \varepsilon)$ is the (closed) ball with center the origin and radius $\varepsilon > 0$ for the norm on E. Equivalently, by Cameron-Martin's formula (4.11),

$$\lim_{\varepsilon \to 0} \frac{1}{\mu(B(0, \varepsilon))} \int_{B(0,\varepsilon)} \mathrm{e}^{\tilde{h}} d\mu = 1$$

where we recall that $\widetilde{h} = (j^*{}_{|E_2^*})^{-1}(h)$.

On the Wiener space $C_0([0, 1])$ (with the supnorm), if h is an element of the Cameron-Martin Hilbert space, we know that $\widetilde{h} = \int_0^1 h'(t) dW(t)$. As we have seen in Chapter 4, this is still the case for a norm $\|\cdot\|$ on $C_0([0, 1]$ such that, for example, $\|x\| \geq C \int_0^1 |x(t)| dt$ for every x in $C_0([0, 1]$ and some constant $C > 0$.

Proof. By symmetry and Jensen's inequality, for each $\varepsilon > 0$,

$$\frac{1}{\mu(B(0, \varepsilon))} \int_{B(0,\varepsilon)} \mathrm{e}^{\tilde{h}} d\mu \geq 1.$$

Therefore, it suffices to show that

$$\limsup_{\varepsilon \to 0} \frac{1}{\mu(B(0, \varepsilon))} \int_{B(0,\varepsilon)} \mathrm{e}^{\tilde{h}} d\mu \leq 1. \tag{7.18}$$

It is plain that (7.18) holds when $h = j^* j(\xi)$ for some $\xi \in E^*$, in other words, $\widetilde{h}(\cdot) = j(\xi)(\cdot) = \langle \xi, \cdot \rangle$ (considered as an element of $\mathrm{L}^2(\mu)$). Now, since $\mathcal{H} = j^*(E_2^*)$, where we recall that E_2^* is the closure of E^* in $\mathrm{L}^2(\mu)$ (cf. Chapter 4), there is a sequence $(\xi_n)_{n \in \mathbb{N}}$ in E^* such that $\lim_{n \to \infty} \|\widetilde{h} - j(\xi_n)\|_{\mathrm{L}^2(\mu)} = 0$. By the Cauchy-Schwarz inequality, for every $\varepsilon > 0$ and every n,

$$\int_{B(0,\varepsilon)} \mathrm{e}^{\tilde{h}} d\mu \leq \left(\int_{B(0,\varepsilon)} \mathrm{e}^{2j(\xi_n)} d\mu \right)^{1/2} \left(\int_{B(0,\varepsilon)} \mathrm{e}^{2(\tilde{h} - j(\xi_n))} d\mu \right)^{1/2}.$$

The result will therefore be established if we show that, for every $\varepsilon > 0$ and every k in $\mathcal{H}$,

$$\frac{1}{\mu(B(0, \varepsilon))} \int_{B(0,\varepsilon)} \mathrm{e}^{\tilde{k}} d\mu \leq \int \mathrm{e}^{|\tilde{k}|} d\mu. \tag{7.19}$$

Indeed, if this is the case, let $k = k_n = 2(h - j^* j(\xi_n))$. Then $\widetilde{k}_n$ is a Gaussian random variable on the probability space $(E, \mathcal{B}, \mu)$ with variance $4\|\widetilde{h} - j(\xi_n)\|^2_{\mathrm{L}^2(\mu)} \to 0$. Therefore, by the dominated convergence theorem,

$$\lim_{n \to \infty} \int \mathrm{e}^{|\tilde{k}_n|} d\mu = 1.$$

Hence, letting ε tend to zero and then n tend to infinity yields the result. We are left with the proof of (7.19). We will actually establish that, for every $t \geq 0$,

$$\mu\big(|\widetilde{k}| \geq t \,\big|\, B(0,\varepsilon)\big) \leq \mu\big(|\widetilde{k}| \geq t\big) \tag{7.20}$$

from which (7.19) immediately follows by integration by parts. To this purpose, assume that $|k| = 1$. We may choose an orthonormal basis $(e_i)_{i\geq 1}$ of $\mathcal{H}$ such that $e_1 = k$. Recall from Proposition 4.2 that the $g_i = ({j^*}_{|E_2^*})^{-1}(e_i) = \widetilde{e}_i$, $i \geq 1$, are independent standard Gaussian random variables. By (7.16), for every convex symmetric set B in $\mathbb{R}^n$, and every $t \geq 0$,

$$\mathbb{P}\big\{|g_1| < t, (g_1, \ldots, g_n) \in B\big\} \geq \mathbb{P}\big\{|g_1| < t\big\}\mathbb{P}\big\{(g_1, \ldots, g_n) \in B\big\}.$$

If we let then $B = \{x \in \mathbb{R}^n; \|\sum_{i=1}^n x_i e_i\| \leq \varepsilon\}$, (7.20) immediately follows from this inequality by Proposition 4.2. The proof of Theorem 7.8 is complete. □

Note that the proof of Theorem 7.8 also applies to $|\widetilde{h}|$ and $c\widetilde{h}^2$ (with $c < 1/2|h|^2$) instead of $\widetilde{h}$. With this tool, L. A. Shepp and O. Zeitouni initiated in [S-Z2] the study of Onsager-Machlup functionals for some completely symmetric norms on Wiener space. In [Ca], a general result in this direction is proved for rotational invariant norms with a known small ball behavior (including in particular Hölder norms and various Sobolev type norms).

Notes for further reading. More on small ball probabilities for Gaussian measures may be found in [D-HJ-S] and in the more recent papers [Gri], [K-L2], [K-L-L], [K-L-S], [K-L-T], [Li], [M-R], [Sh], [S-W], [St2]... In particular, in the latter papers, the small ball behaviors are used in the study of rates of convergence in both Strassen's and Chung's law of the iterated logarithm. Some general statements towards this goal are stated in [Ta9]. Recall also the paper [D-L] on Strassen's law of the iterated logarithm for Brownian motion for arbitrary seminorms. See also the recent reference [Lif3]. More on the support of a diffusion theorem, small perturbations of dynamical systems and Onsager-Machlup functionals in stronger topologies on Wiener space can be found in the afore mentioned papers [A-K-S], [B-R], [BA-G-L], [BA-L2], [Ca], [Ci2], [G-N-SS], [Me], [M-SS], [S-Z1], [S-Z2]...

8. ISOPERIMETRY AND HYPERCONTRACTIVITY

In this last chapter, we further investigate the tight relationships between isoperimetry and semigroup techniques as started in Chapter 2. More precisely, we present some of the semigroup tools which may be used to investigate the isoperimetric inequality in Euclidean and Gauss space. In particular, we will concentrate on the isoperimetric and concentration inequalities for Gaussian measures and show how these relate to hypercontractivity of the Ornstein-Uhlenbeck semigroup. The overwhole approach is inspired by the work of N. Varopoulos in his functional approach to isoperimetric inequalities on groups and manifolds. To better illustrate the scheme of proofs, we start with the classical isoperimetry in $\mathbb{R}^n$ and observe, in particular, that the isoperimetric inequality in $\mathbb{R}^n$ is equivalent to saying that the L^2-norm of the heat semigroup acting on characteristic functions of sets increases under isoperimetric rearrangement. Then, we investigate the analogous situation with respect to the canonical Gaussian measure γ_n. As for the concentration of measure phenomenon, we will discover how the various properties of the Ornstein-Uhlenbeck semigroup such as the commutation property or hypercontractivity can yield in a simple way (a form of) the isoperimetric inequality for Gaussian measures.

Recall from Chapter 1 that the classical isoperimetric inequality in $\mathbb{R}^n$ states that among all compact subsets A with fixed volume $\mathrm{vol}_n(A)$ and smooth boundary ∂A, Euclidean balls minimize the surface measure of the boundary. In other words, whenever $\mathrm{vol}_n(A) = \mathrm{vol}_n(B)$ where B is a ball with some radius r (and $n > 1$),

$$\mathrm{vol}_{n-1}(\partial A) \geq \mathrm{vol}_{n-1}(\partial B). \tag{8.1}$$

Now, $\mathrm{vol}_{n-1}(\partial B) = nr^{n-1}\omega_n$ where ω_n is the volume of the ball of radius 1 so that (8.1) is equivalent to saying that

$$\mathrm{vol}_{n-1}(\partial A) \geq n\omega_n^{1/n}\mathrm{vol}_n(A)^{(n-1)/n}. \tag{8.2}$$

The function $n\omega_n^{1/n}x^{(n-1)/n}$ on $\mathbb{R}^+$ is the isoperimetric function of the classical isoperimetric problem on $\mathbb{R}^n$. Euclidean balls are the extremal sets and achieve equality in (8.2).

It is well-known that (8.2) may be expressed equivalently on functions by means of the coarea formula [Fed], [Maz2], [Os]. After integration by parts (see e.g. [Maz2], p. ...), it yields

$$n\omega^{1/n}\|f\|_{n/n-1} \le \big\||\nabla f|\big\|_1 \tag{8.3}$$

for every C^∞ compactly supported function f on $\mathbb{R}^n$. This inequality is equivalent to (8.2) by letting f approximate the characterisitic function I_A of a set A whose boundary ∂A is smooth enough so that $\int |\nabla f|\,dx$ approaches $\mathrm{vol}_{n-1}(\partial A)$. For simplicity, smoothness properties will be understood in this way here. Inequality (8.3) is due independently to E. Gagliardo [Ga] and L. Nirenberg [Ni] with a nice inductive proof on the dimension. This proof, however, does not seem to yield the optimal constant, and therefore the extremal character of balls. The connection between (8.2) and (8.3) through the coarea formula seems to be due to H. Federer and W. H. Fleming [F-F] and V. G. Maz'ja [Maz1] (cf. [Os]).

Inequality (8.3) of course belongs to the family of Sobolev inequalities. Replacing f (positive) by f^α for some appropriate α easily yields after an application of Hölder's inequality that, for every C^∞ compactly supported function f on $\mathbb{R}^n$,

$$\|f\|_q \le C(n,p,q)\big\||\nabla f|\big\|_p \tag{8.4}$$

with $\frac{1}{q} = \frac{1}{p} - \frac{1}{n}$ and $C(n,p,q) > 0$ a constant only depending on n, p, q, $1 \le p < n$. The family of inequalities (8.4) with $1 < p < n$ goes back to S. Sobolev [So], the inequality for $p = 1$ (which implies the others) having thus been established later on. Of particular interest is the value $p = 2$ which may be expressed equivalently by integration by parts as $(n > 2)$

$$\|f\|^2_{2n/n-2} \le C\int |\nabla f|^2 dx = C\int f(-\Delta f)dx \tag{8.5}$$

where Δ is the usual Laplacian on $\mathbb{R}^n$. As developed in an abstract setting by N. Varopoulos [Va2] (cf. [Va5], [V-SC-C]), this Dirichlet type inequality (8.5) is closely related to the behavior of the heat semigroup $T_t = \mathrm{e}^{t\Delta}$, $t \ge 0$, as $\|T_t f\|_\infty \le Ct^{-n/2}\|f\|_1$, $t > 0$. We will come back to this below.

Our first task will be to describe, in this concrete setting, some aspects of the semigroup techniques of [Va2], [Va3], and to show how these can yield, in a very simple way, (a form of) the isoperimetric inequality. We will work with the integral representation of the heat semigroup $T_t = \mathrm{e}^{t\Delta}$, $t \ge 0$, as

$$T_t f(x) = \int_{\mathbb{R}^n} f\big(x + \sqrt{2t}\,y\big)d\gamma_n(y), \quad x \in \mathbb{R}^n, \quad f \in \mathrm{L}^1(dx),$$

where γ_n is the canonical Gaussian measure on $\mathbb{R}^n$.

The following proposition is crucial for the understanding of the general principle. Set, for Borel subsets A, B in $\mathbb{R}^n$, and $t \ge 0$,

$$K_t^T(A,B) = \int_B T_t(I_A)dx.$$

A^c denotes below the complement of A.

Proposition 8.1. *For every compact set A in $\mathbb{R}^n$ with smooth boundary ∂A and every $t \geq 0$,*

$$K_t^T(A, A^c) \leq \left(\frac{t}{\pi}\right)^{1/2} \mathrm{vol}_{n-1}(\partial A).$$

Proof. Let f, g be smooth functions on $\mathbb{R}^n$. For every $t \geq 0$, we can write

$$\begin{aligned}\int g\,(T_t f - f)dx &= \int_0^t \left(\int g\, \Delta T_s f\, dx\right) ds \\ &= -\int_0^t \left(\int \langle \nabla T_s g, \nabla f\rangle\, dx\right) ds.\end{aligned}$$

Now, by integration by parts,

$$\nabla T_s g = \frac{1}{\sqrt{2s}} \int_{\mathbb{R}^n} y\, g\big(x + \sqrt{2s}\, y\big) d\gamma_n(y)\,.$$

Hence

$$\int g(T_t f - f) dx = -\int_0^t \frac{1}{\sqrt{2s}} \iint \langle \nabla f(x), y\rangle\, g\big(x + \sqrt{2s}\, y\big) dx d\gamma_n(y) ds.$$

This inequality of course extends to $g = I_{A^c}$. Since

$$\iint \langle \nabla f(x), y\rangle dx d\gamma_n(y) = 0,$$

we see that, for every s,

$$\begin{aligned}-\iint \langle \nabla f(x), y\rangle\, I_{A^c}\big(x + \sqrt{2s}\, y\big) dx d\gamma_n(y) &\leq \iint \big(\langle \nabla f(x), y\rangle\big)^- dx d\gamma_n(y) \\ &= \frac{1}{2} \iint \big|\langle \nabla f(x), y\rangle\big| dx d\gamma_n(y) \\ &= \frac{1}{\sqrt{2\pi}} \int |\nabla f| dx\end{aligned}$$

by partial integration with respect with respect to $d\gamma_n(y)$. The conclusion follows since, by letting f approximate I_A, $\int |\nabla f|\, dx$ approaches $\mathrm{vol}_{n-1}(\partial A)$. The proof of Proposition 8.1 is complete. □

Proposition 8.1 is sharp since it may be tested on balls. Namely, if B is an Euclidean ball, one may check that

$$\lim_{t \to 0} \left(\frac{\pi}{t}\right)^{1/2} K_t^T(B, B^c) = \mathrm{vol}_{n-1}(\partial B). \tag{8.6}$$

By translation invariance and homogeneity, one may assume that B is the unit ball with center the origin and radius 1. Then, for $t > 0$,

$$K_t^T(B, B^c) = \int_{\{|x|>1\}} \gamma_n\big(y \in \mathbb{R}^n; \big|x + \sqrt{2t}\, y\big| \leq 1\big) dx.$$

Using polar coordinates and the rotational invariance of γ_n,

$$\begin{aligned} K_t^T(B, B^c) &= \int_1^\infty \int_{\omega \in \partial B} \rho^{n-1} \gamma_n\big(y; \big|\rho\omega + \sqrt{2t}\, y\big| \leq 1\big) d\rho d\omega \\ &= \mathrm{vol}_{n-1}(\partial B) \int_1^\infty \rho^{n-1} \gamma_1 \otimes \gamma_{n-1}\big((y_1, \widetilde{y}); \big|\rho + \sqrt{2t} y_1\big|^2 + 2t|\widetilde{y}|^2 \leq 1\big) d\rho \end{aligned}$$

where $y = (y_1, \widetilde{y})$, $y_1 \in \mathbb{R}$, $\widetilde{y} \in \mathbb{R}^{n-1}$. We then use Fubini's theorem to write

$$K_t^T(B, B^c) = \mathrm{vol}_{n-1}(\partial B) \int J_t(y_1, \widetilde{y}) d\gamma_1(y_1) d\gamma_{n-1}(\widetilde{y})$$

where

$$J_t(y_1, \widetilde{y}) = I_{\{2t|\widetilde{y}|^2 \leq 1; \sqrt{2t} y_1 \leq \sqrt{1-2t|\widetilde{y}|^2} - 1\}} \int_1^\infty \rho^{n-1} I_{\{|\rho + \sqrt{2t} y_1|^2 \leq 1 - 2t|\widetilde{y}|^2\}} d\rho.$$

By a simple integration of the preceding, it is easily seen that

$$\lim_{t \to 0} \frac{1}{\sqrt{t}}\, J_t(y_1, \widetilde{y}) = -\sqrt{2}\, y_1 I_{\{y_1 \leq 0\}}$$

so that, by dominated convergence,

$$\lim_{t \to 0} \frac{1}{\sqrt{t}}\, K_t^T(B, B^c) = -\,\mathrm{vol}_{n-1}(\partial B) \int_{-\infty}^0 \sqrt{2}\, y_1 d\gamma_1(y_1) = \frac{1}{\sqrt{\pi}}\, \mathrm{vol}_{n-1}(\partial B)$$

which is the claim (8.6).

As a consequence of (8.6), the isoperimetric inequality (8.2) is equivalent to saying that, for every $t \geq 0$ and every compact subset A with smooth boundary, $K_t^T(A, A) \leq K_t^T(B, B)$ whenever B is a ball with the same volume as A. In other words, since $K_t^T(A, A) = \|T_{t/2}(I_A)\|_2^2$,

$$\big\|T_t(I_A)\big\|_2 \leq \big\|T_t(I_B)\big\|_2, \quad t \geq 0. \tag{8.7}$$

Indeed, under such a property, by Proposition 8.1, for every $t > 0$,

$$\mathrm{vol}_{n-1}(\partial A) \geq \Big(\frac{\pi}{t}\Big)^{1/2} K_t^T(A, A^c) \geq \Big(\frac{\pi}{t}\Big)^{1/2} K_t^T(B, B^c)$$

and, when $t \to 0$, $\mathrm{vol}_{n-1}(\partial A) \geq \mathrm{vol}_{n-1}(\partial B)$ by (8.6).

Inequality (8.7) was actually established by A. Baernstein and B. A. Taylor [B-T] through delicate symmetrization arguments of isoperimetric nature (see also

[Ba]). While we noticed its equivalence with isoperimetry, one may wonder for an independent simpler proof of (8.7).

If one does not mind bad constants, one can actually deduce (a form of) isoperimetry from Proposition 8.1 in an elementary way. We will use below this simpler argument in the context of Riemannian manifolds. Note from the uniform estimate $\|T_t f\|_\infty \leq Ct^{-n/2}\|f\|_1$, $t > 0$, that, by interpolation, $\|T_t f\|_2 \leq \sqrt{C}t^{-n/4}\|f\|_1$, $t > 0$, for every f in $\mathrm{L}^1(dx)$. Hence, by Proposition 8.1, for every compact subset A in $\mathbb{R}^n$ with smooth boundary ∂A, and every $t > 0$,

$$\begin{aligned}\mathrm{vol}_{n-1}(\partial A) &\geq \left(\frac{\pi}{t}\right)^{1/2} K_t^T(A, A^c)\\ &\geq \left(\frac{\pi}{t}\right)^{1/2}\Big[\mathrm{vol}_n(A) - \|T_{t/2}(I_A)\|_2^2\Big]\\ &\geq \left(\frac{\pi}{t}\right)^{1/2}\Big[\mathrm{vol}_n(A) - C\Big(\frac{t}{2}\Big)^{-n/2}\mathrm{vol}_n(A)^2\Big].\end{aligned}$$

Optimizing over $t > 0$ then yields

$$\mathrm{vol}_{n-1}(\partial A) \geq C'\mathrm{vol}_n(A)^{(n-1)/n}$$

hence (8.2), with however a worse constant. This easy proof could appear even simpler than the one by E. Gagliardo and L. Nirenberg.

These elementary arguments may be used in the same way in greater generality, for example in Riemannian manifolds. Following [Va2], [Va3], we briefly describe how the arguments should be developed in this case.

It is known ([C-L-Y], [Va1]) that an isoperimetric inequality on a Riemannian manifold M, for example, always forces some control on the heat kernel of M. More precisely, let M be a complete connected Riemannian manifold of dimension N, and, say, noncompact and of infinite volume. Let furthermore Δ be the Laplace-Beltrami operator on M and denote by $(P_t)_{t\geq 0}$ the heat semigroup with kernel $p_t(x,y)$.

Theorem 8.2. *Assume that there exist $n > 1$ and $C > 0$ such that for all compacts subsets A of M with smooth boundary ∂A,*

$$\mathrm{vol}(A)^{(n-1)/n} \leq C\,\mathrm{vol}(\partial A). \tag{8.8}$$

Then, for some constant $C' > 0$,

$$p_t(x,y) \leq \frac{C'}{t^{n/2}} \tag{8.9}$$

for every $t > 0$ and every $x, y \in M$. Furthermore, for each $\delta > 0$, there exists $C_\delta > 0$ such that

$$p_t(x,y) \leq \frac{C_\delta}{t^{n/2}}\,\exp\left(-\frac{d(x,y)^2}{4(1+\delta)t^2}\right) \tag{8.10}$$

for every $t > 0$ and every $x, y \in M$.

The proof of this theorem is entirely similar to the Euclidean case. We compare both (8.8) and (8.9) on the scale of Sobolev inequalities. One of the main points is the formal equivalence, due to N. Varopoulos [Va2] (cf. [Va5], [V-SC-C] and the references therein), of the L^2-Sobolev inequality (8.5) and the uniform control of the heat semigroup or kernel

$$\|P_t f\|_\infty \leq \frac{C}{t^{n/2}} \|f\|_1, \quad t > 0, \quad f \in C_o^\infty(M). \tag{8.11}$$

This result, inspired from the work of J. Nash [Na] and J. Moser [Mo] on the regularity of solutions of parabolic differential equations, is the main link between analysis and geometry. Various techniques then allow one to deduce from the uniform control (8.11) of the kernel the Gaussian off-diagonal estimates (8.10) (cf. [Da], [L-Y], [Va4]...). Theorem 8.2 may be localized in small time (from an isoperimetric inequality on sets of small volume), or in large time [C-F] (sets of large volume).

As we have seen in the classical case, it is sometimes possible to reverse the preceding procedure and to deduce some isoperimetric property from a (uniform) control of the heat kernel. To emphasize the methods rather than the result itself, let us consider only, for simplicity, Riemannian manifolds with nonnegative Ricci curvature. Owing to the Euclidean example, we need to understand how we should complement a Sobolev inequality at the level L^2 (8.5) in order to reach the level L^1 (8.3) and therefore isoperimetry. In this Riemannian setting, this step may be performed with a fundamental inequality due to P. Li et S.-T. Yau [L-Y] in their study of parabolic Harnack inequalities. This inequality is a functional translation of curvature and its proof (see e.g. [Da]) is only based, as in Chapter 2, on Bochner formula and the related curvature-dimension inequalities (cf. Proposition 2.2). We only state it in manifolds with nonnegative Ricci curvature.

Proposition 8.3. *Let M be a Riemannian manifold of dimension N and nonnegative Ricci curvature and let $(P_t)_{t\geq 0}$ be the heat semigroup on M. For every strictly positive function f in $C_o^\infty(V)$ and every $t > 0$,*

$$\frac{|\nabla P_t f|^2}{(P_t f)^2} - \frac{\Delta P_t f}{P_t f} \leq \frac{N}{2t}\,. \tag{8.12}$$

As shown by N. Varopoulos [Va4], one easily deduces from the pointwise inequality (8.12) that, for every f smooth enough and every $t > 0$,

$$\big\||\nabla P_t f|\big\|_\infty \leq \frac{C}{\sqrt{t}} \|f\|_\infty \tag{8.13}$$

for some C only depending on the dimension N, that is a control of the spatial derivatives of the heat kernel. Indeed, according to (8.12), $(\Delta P_t f)^- \leq N(2t)^{-1} P_t f$ so that $\|\Delta P_t f\|_1 \leq N t^{-1} \|f\|_1$. By duality, $\|\Delta P_t f\|_\infty \leq N t^{-1} \|f\|_\infty$. This estimate, used in (8.12) again, then immediately yields (8.13).

The control (8.13) of the gradient of the semigroup in $\sqrt{t}$ (similar to Proposition 8.1) is then the crucial information which, together with (8.11), allows us to reach isoperimetry. Note that the dimension only comes into (8.11) and that (8.13) is in a sense independent of this dimension (besides the constant). We will come back to this comment in the Gaussian setting next. Inequality (8.13) shows that, by duality, for every f in $C_o^\infty(M)$ and every $t > 0$,

$$\|f - P_t f\|_1 \leq C\sqrt{t}\, \big\| |\nabla f| \big\|_1. \tag{8.14}$$

Indeed, for every smooth function g such that $\|g\|_\infty \leq 1$,

$$\begin{aligned} \int g\,(f - P_t f)\,dx &= -\int_0^t \bigg(\int g\,\Delta P_s f\,dx \bigg) ds \\ &= -\int_0^t \bigg(\int \Delta P_s g\, f\,dx \bigg) ds \\ &= \int_0^t \bigg(\int \langle \nabla P_s g, \nabla f\rangle dx \bigg) ds \leq \int_0^t \|\nabla P_s g\|_\infty \big\| |\nabla f| \big\|_1\, ds. \end{aligned}$$

Now (8.14) together with (8.11) imply, exactly as in the Euclidean setting, that for some constant $C > 0$ and every compact subset A of M with smooth boundary ∂A,

$$\mathrm{vol}(A)^{(n-1)/n} \leq C\,\mathrm{vol}(\partial A),$$

that is the announced isoperimetry. We thus established the following theorem [Va4].

Theorem 8.4. *Let M be a Riemannian manifold with nonnegative Ricci curvature. If for some $n > 1$ and some $C > 0$,*

$$p_t(x, y) \leq \frac{C}{t^{n/2}}$$

uniformly in $t > 0$ and $x, y \in M$, then, for some constant $C' > 0$ and every compact subset A of M with smooth boundary ∂A

$$\mathrm{vol}(A)^{(n-1)/n} \leq C'\,\mathrm{vol}(\partial A).$$

When the Ricci curvature is only bounded below, the preceding result can only hold locally. In general, the geometry at infinity of the manifold is such that a heat kernel estimate of the type (8.11) (for large t's) only yields isoperimetry for half of the dimension (cf. [C-L] for further details).

A third most important part of the theory concerns the relation of the preceding isoperimetric and Sobolev inequalities with minorations of volumes of balls. We refer to the works of P. Li and S.-T. Yau [L-Y] and N. Varopoulos [Va4], [Va5] and to the monographs [Da], [V-C-SC].

Now, we turn to the Gaussian isoperimetric inequality and the Ornstein-Uhlenbeck semigroup. We already saw in Chapter 2 how this semigroup may be used in order to describe the concentration properties of Gaussian measures. We use it here to try to reach the full isoperimetric statement and base our approach on hypercontractivity. As we will see indeed, hypercontractivity and logarithmic Sobolev inequalities may indeed be considered as analogues of heat kernel bounds and L^2-Sobolev inequalities in this context.

Recall the canonical Gaussian measure γ_n on $\mathbb{R}^n$ with density with respect to Lebesgue measure $\varphi_n(x) = (2\pi)^{-n/2}\exp(-|x|^2/2)$. Recall also that the isoperimetric property for γ_n indicates that if A is a Borel set in $\mathbb{R}^n$ and H is a half-space

$$H = \{x \in \mathbb{R}^n; \langle x, u\rangle \le a\}, \quad |u| = 1, \quad a \in \mathbb{R},$$

such that $\gamma_n(A) = \gamma_n(H) = \Phi(a)$, then, for any real number $r \ge 0$,

$$\gamma_n(A_r) \ge \gamma_n(H_r) = \Phi(a + r).$$

In the applications to hypercontractivity and logarithmic Sobolev inequalities, we will use the Gaussian isoperimetric inequality in its infinitesimal formulation connecting the "Gaussian volume" of a set to the "Gaussian length" of its boundary (which is really isoperimetry). More precisely, given a Borel subset A of $\mathbb{R}^n$, define ([Eh3], [Fed]) the Gaussian Minkowski content of its boundary ∂A as

$$\mathcal{O}_{n-1}(\partial A) = \liminf_{r\to 0} \frac{1}{r}\big[\gamma_n(A_r) - \gamma_n(A)\big].$$

If ∂A is smooth, $\mathcal{O}_{n-1}(\partial A)$ may be obtained as the integral of the Gaussian density along ∂A (see below). In this langague, the isoperimetric inequality then expresses that if H is a half-space with the same measure as A, then

$$\mathcal{O}_{n-1}(\partial A) \ge \mathcal{O}_{n-1}(\partial H).$$

Now, one may easily compute (in dimension one) the Minkowski content of a half-space as

$$\mathcal{O}_{n-1}(\partial H) = \liminf_{r\to 0} \frac{1}{r}\big[\Phi(a + r) - \Phi(a)\big] = \varphi_1(a)$$

where $\Phi(a) = \gamma_n(H) = \gamma_n(A)$ and where $\varphi_1(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, $x \in \mathbb{R}$. Hence, denoting by Φ^{-1} the inverse function of Φ, we get that for every Borel set A in $\mathbb{R}^n$,

$$\mathcal{O}_{n-1}(\partial A) \ge \varphi_1 \circ \Phi^{-1}\big(\gamma_n(A)\big). \tag{8.15}$$

The function $\varphi_1 \circ \Phi^{-1}$ is the isoperimetric function of the Gauss space $(\mathbb{R}^n, \gamma_n)$. It may be compared to the function $n\omega^{1/n}x^{(n-1)/n}$ of the classical isoperimetric inequality in $\mathbb{R}^n$. The function $\varphi_1 \circ \Phi^{-1}$ is still concave; it is defined on $[0, 1]$, is symmetric with respect to the vertical line going through $\frac{1}{2}$ with a maximum there equal to $(2\pi)^{-1/2}$, and its behavior at the origin (or at 1 by symmetry) is governed by the equivalence

$$\lim_{x\to 0} \frac{\varphi_1 \circ \Phi^{-1}(x)}{x(2\log(1/x))^{1/2}} = 1. \tag{8.16}$$

This can easily be established by noting that the derivative of $\varphi_1 \circ \Phi^{-1}$ is $-\Phi^{-1}$ and by comparing $\Phi^{-1}(x)$ to $(2\log(1/x))^{1/2}$.

As in the classical case, (8.15) may be expressed equivalently on functions by means, again, of the coarea formula (see [Fed], [Maz2], [Eh3]). Writing for a smooth function f on $\mathbb{R}^n$ with gradient ∇f that

$$\int |\nabla f| d\gamma_n = \int_0^\infty \bigg(\int_{C_s} \varphi_n(x)\, d\mathcal{H}_{n-1}(x) \bigg) ds$$

where $C_s = \{x \in \mathbb{R}^n; |f(x)| = s\}$ and where $d\mathcal{H}_{n-1}$ is the Hausdorff measure of dimension $n-1$ on C_s, we deduce from (8.15) that

$$\int |\nabla f| d\gamma_n \geq \int_0^\infty \varphi_1 \circ \Phi^{-1}\big(\gamma_n\big(|f| \geq s\big)\big) ds. \tag{8.17}$$

When f is a smooth function approximating the indicator function of a set A, we of course recover (8.15) from (8.17), at least for subsets A with smooth boundary. Due to the equivalence (8.16), one sees in particular on (8.17) that a smooth function f satisfying $\int |\nabla f| d\gamma_n < \infty$ is such that $\int |f| (\log(1 + |f|))^{1/2} d\gamma_n < \infty$. Indeed, we first see from (8.17) and (8.16) that for every s_0 large enough

$$\int |\nabla f| d\gamma_n \geq \int_{s_0}^\infty \gamma_n\big(|f| \geq s\big) ds$$

from which $\int |f| d\gamma_n \leq C < \infty$ by the classical integration by parts formula. For every $s \geq 0$, $\gamma_n(|f| \geq s) \leq C/s$ so that, by (8.17) and (8.16) again, for every large s_0,

$$\int |\nabla f| d\gamma_n \geq \int_{s_0}^\infty \gamma_n\big(|f| \geq s\big) \big(\log(s/C)\big)^{1/2} ds$$

from which the claim immediately follows. In analogy with (8.3), such an inequality belongs to the family of Sobolev inequalities, but here of logarithmic type.

It is plain that inequalities (8.15) and (8.17) have analogues in infinite dimension for the appropriate notions of surface measure and gradient (as we did for example with concentration in Chapter 4). Again, the crucial inequalities are the ones in finite dimension.

We showed in Chapter 2 how the Ornstein-Uhlenbeck semigroup $(P_t)_{t \geq 0}$ (and for the large values of the time t) may be used to investigate the concentration phenomenon of Gaussian measures. Our purpose here will be to show, in the same spirit as what we presented in the classical case, that the behavior of $(P_t)_{t \geq 0}$ for the small values of t together with its hypercontractivity property may properly be combined to yield (a version of) the infinitesimal version (8.15) of the isoperimetric inequality. More precisely, we will show, with these tools, that there exists a small enough numerical constant $0 < c < 1$ such that for every A with smooth boundary,

$$\mathcal{O}_{n-1}(\partial A) \geq c\, \varphi_1 \circ \Phi^{-1}\big(\gamma_n(A)\big).$$

We doubt that this approach can lead to the exact constant $c = 1$. The line of reasoning will follow the one of the classical case, simply replacing actually the classical heat semigroup estimates and Sobolev inequalities on $\mathbb{R}^n$ by the hypercontractivity property and logarithmic Sobolev inequalities of the Ornstein-Uhlenbeck semigroup. We follow [Led5] and now turn to hypercontractivity and logarithmic Sobolev inequalities.

Let $(W(t))_{t\geq 0}$ be a standard Brownian motion starting at the origin with values in $\mathbb{R}^n$. Consider the stochastic differential equation

$$dX(t) = \sqrt{2}\, dW(t) - X(t)dt$$

with initial condition $X(0) = x$, whose solution simply is

$$X(t) = \mathrm{e}^{-t}\left(x + \sqrt{2}\int_0^t \mathrm{e}^s dW(s)\right), \quad t \geq 0.$$

Since $\sqrt{2}\int_0^t \mathrm{e}^s dW(s)$ has the same distribution as $W(\mathrm{e}^{2t}-1)$, the Markov semigroup $(P_t)_{t\geq 0}$ of $(X(t))_{t\geq 0}$ is given by

$$(8.18)\quad P_t f(x) = \mathbb{E}\big(f\big(\mathrm{e}^{-t}x+\mathrm{e}^{-t}W\big(\mathrm{e}^{2t}-1\big)\big)\big) = \int_{\mathbb{R}^n} f\big(\mathrm{e}^{-t}x+(1-\mathrm{e}^{-2t})^{1/2}y\big)\, d\gamma_n(y)$$

for any f in $\mathrm{L}^1(\gamma_n)$ (for example), thus defining the Ornstein-Uhlenbeck or Hermite semigroup with respect to the Gaussian measure γ_n. As we have seen in Chapter 2, $(P_t)_{t\geq 0}$ is a Markovian semigroup of contractions on all $\mathrm{L}^p(\gamma_n)$-spaces, $1 \leq p \leq \infty$, symmetric and invariant with respect to γ_n, and with generator L which acts on each smooth function f on $\mathbb{R}^n$ as $Lf(x) = \Delta f(x) - \langle x, \nabla f(x)\rangle$. The generator L satisfies the integration by parts formula with respect to γ_n

$$\int f(-Lg)\, d\gamma_n = \int \langle \nabla f, \nabla g\rangle\, d\gamma_n$$

for every smooth functions f, g on $\mathbb{R}^n$.

One of the remarkable properties of the Ornstein-Uhlenbeck semigroup is the hypercontractivity property discovered by E. Nelson [Nel]: whenever $1 < p < q < \infty$ and $t > 0$ satisfy $\mathrm{e}^t \geq [(q-1)/(p-1)]^{1/2}$, then, for all functions f in $\mathrm{L}^p(\gamma_n)$,

$$(8.19)\qquad \|P_t f\|_q \leq \|f\|_p$$

where (now) $\|\cdot\|_p$ is the norm in $\mathrm{L}^p(\gamma_n)$. In other words, P_t maps $L^p(\gamma_n)$ in $L^q(\gamma_n)$ $(q > p)$ with norm one. Many simple proofs of (8.19) have been given in the literature (see [Gr4]), mainly based on its equivalent formulation as logarithmic Sobolev inequalities due to L. Gross [Gr3]. Fix $p = 2$ and take $q(t) = 1 + \mathrm{e}^{2t}$, $t \geq 0$. Given a smooth function f, set $\Psi(t) = \|P_t f\|_{q(t)}$ where $q(t) = 1 + \mathrm{e}^{2t}$. Under the hypercontractivity property (2.2), $\Psi(t) \leq \Psi(0)$ for every $t \geq 0$ and thus $\Psi'(0) \leq 0$. Performing this differentiation, we see that

$$(8.20)\qquad \int f^2 \log|f|\, d\gamma_n - \int f^2 d\gamma_n \,\log\left(\int f^2 d\gamma_n\right)^{1/2} \leq \int |\nabla f|^2\, d\gamma_n \quad \left(= \int f(-\mathrm{L}f) d\gamma_n\right)$$

which in turn implies (8.19) by applying it to $(P_t f)^p$ instead of $f \geq 0$ for every t and every $p \geq 1$ (cf. [B-É]). The inequality (8.20) is called a logarithmic Sobolev inequality. One may note, with respect to the classical Sobolev inequalities on $\mathbb{R}^n$, that it is only of logarithmic type, with however constants independent of the dimension, a characteristic feature of Gaussian measures.

Simple proofs of (8.20) may be found in e.g. [Ne3], [A-C], [B-É], [Bak]... (cf. [Gr4]). The one which we present now for completeness already appeared in [Led3] and only relies (see also [B-É]) on the commutation property (2.6)

$$\nabla P_t f = \mathrm{e}^{-t} P_t(\nabla f).$$

That is, the proof we will give of hypercontractivity relies on exactly the same argument which allowed us to describe the concentration of γ_n in the form of (2.7) through Proposition 2.1 and is actually very similar. We will come back to this important point. In order to establish (8.20), replacing f (positive, or better such that $0 < a \leq f \leq b$ for constants a, b) by $\sqrt{f}$, it is enough to show that

$$\int f \log f \, d\gamma_n - \int f \, d\gamma_n \, \log\left(\int f \, d\gamma_n \right) \leq \frac{1}{2} \int \frac{1}{f} \, |\nabla f|^2 \, d\gamma_n. \tag{8.21}$$

To this aim, we can write by the semigroup properties and integration by parts that

$$\begin{aligned}
\int f \log f \, d\gamma_n - \int f \, d\gamma_n \, \log\left(\int f \, d\gamma_n \right) &= - \int_0^\infty \left(\frac{d}{dt} \int P_t f \, \log P_t f \, d\gamma_n \right) dt \\
&= - \int_0^\infty \left(\int L P_t f \, \log P_t f \, d\gamma_n \right) dt \\
&= \int_0^\infty \left(\int \langle \nabla P_t f, \nabla (\log P_t f) \rangle \, d\gamma_n \right) dt \\
&= \int_0^\infty \left(\int \frac{1}{P_t f} \, |\nabla P_t f|^2 \, d\gamma_n \right) dt.
\end{aligned}$$

Now, setting

$$F(t) = \int \frac{1}{P_t f} \, |\nabla P_t f|^2 \, d\gamma_n \quad t \geq 0,$$

the commutation property $\nabla P_t f = \mathrm{e}^{-t} P_t(\nabla f)$ and Cauchy-Schwarz inequality on the integral representation of P_t show that, for every $t \geq 0$,

$$\begin{aligned}
F(t) &= \mathrm{e}^{-2t} \sum_{i=1}^n \int \frac{1}{P_t f} \left(P_t \frac{\partial f}{\partial x_i} \right)^2 d\gamma_n \\
&\leq \mathrm{e}^{-2t} \sum_{i=1}^n \int P_t \left(\frac{1}{f} \left(\frac{\partial f}{\partial x_i} \right)^2 \right) d\gamma_n = \mathrm{e}^{-2t} \int \frac{1}{f} \, |\nabla f|^2 \, d\gamma_n
\end{aligned}$$

which immediately yields (8.21). Therefore, hypercontractivity is established in this way.

While our aim is to investigate isoperimetric inequalities via semigroup techniques, it is of interest however to notice that the Gaussian isoperimetric inequality (8.15) or (8.17) may be used to establish the logarithmic Sobolev inequality (8.20) and therefore hypercontractivity. This was observed in [Led1] in analogy with the classical case discussed in the first part of this chapter. Let f be a smooth positive function on $\mathbb{R}^n$ with $\|f\|_2 = 1$. Apply then (8.17) to $g = f^2(\log(1+f^2))^{1/2}$. Using (8.16), one obtains after some elementary, although cumbersome, computations that for every $\varepsilon > 0$ there exists $C(\varepsilon) > 0$ only depending on ε such that

$$\int f^2 \log(1+f^2)\, d\gamma_n \leq (1+\varepsilon)\Big(2\int |\nabla f|^2 d\gamma_n\Big)^{1/2}\Big(\int f^2\log(1+f^2)\, d\gamma_n + 2\Big)^{1/2} + C(\varepsilon).$$

It follows that

$$\begin{aligned} 2\int f^2 \log f\, d\gamma_n &\leq \int f^2 \log(1+f^2)\, d\gamma_n \\ &\leq 2(1+\varepsilon)^4 \int |\nabla f|^2 d\gamma_n + 2(1+\varepsilon)^2\Big(\int |\nabla f|^2 d\gamma_n\Big)^{1/2} + C'(\varepsilon) \end{aligned}$$

where $C'(\varepsilon) = (1+\varepsilon)C(\varepsilon)/\varepsilon$. To get rid of the extra terms on the left of this inequality, we use a tensorization argument of A. Ehrhard [Eh4]: this inequality namely holds with constants independent of the dimension n; therefore, applying it to $f^{\otimes k}$ in $(\mathbb{R}^n)^k = \mathbb{R}^{nk}$ yields

$$k\int f^2 \log f\, d\gamma_n \leq k(1+\varepsilon)^4 \int |\nabla f|^2 d\gamma_n + \sqrt{k}\,(1+\varepsilon)^2\Big(\int |\nabla f|^2 d\gamma_n\Big)^{1/2} + C'(\varepsilon).$$

Divide then by k, let k tend to infinity and then ε to zero and we obtain (8.20).

Now, we would like to try to understand how hypercontractivity and logarithmic Sobolev inequalities may be used in order to reach isoperimetry in this Gaussian setting. Of course, our approach to known results and theorems is only formal, but it could be of some help in more abstract frameworks.

Before turning to the main argument, let us briefly discuss, on two specific questions, why hypercontractiviy should be of potential interest to isoperimetry and concentration. The following comments are not presented in the greatest generality.

Recall the Hermite polynomials $\{\sqrt{k!}\,h_k; k \in \mathbb{N}\}$ which forms an orthonormal basis of $\mathrm{L}^2(\gamma_1)$ (cf. the introduction of Chapter 5). In the same way, for any fixed $n \geq 1$, set, for every $\underline{k} = (k_1, \ldots, k_n) \in \mathbb{N}^n$ and every $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$H_{\underline{k}}(x) = \prod_{i=1}^n \sqrt{k_i!}\, h_{k_i}(x_i).$$

Then, $\{H_{\underline{k}}; \underline{k} \in \mathbb{N}^n\}$ is an orthonormal basis of $\mathrm{L}^2(\gamma_n)$. Therefore, as in Chapter 5 in greater generality, a function f in $\mathrm{L}^2(\gamma_n)$ can be written as

$$f = \sum_{\underline{k}\in\mathbb{N}^n} f_{\underline{k}} H_{\underline{k}}$$

where $f_{\underline{k}} = \int f H_{\underline{k}} d\gamma_n$. This sum may also be written as

$$f = \sum_{d=0}^{\infty} \bigg(\sum_{|\underline{k}|=d} f_{\underline{k}} H_{\underline{k}} \bigg) = \sum_{d=0}^{\infty} \Psi^{(d)}(f)$$

where $|\underline{k}| = k_1 + \cdots + k_n$. $\Psi^{(d)}(f)$ is known as the chaos of degree d of f. Since $h_0 \equiv 1$, $\Psi^{(0)}(f)$ is simply the mean of f; $h_1(x) = x$, so chaos of order or degree 1 are (in probabilistic notation) Gaussian sums $\sum_{i=1}^{n} g_i a_i$ (where $(g_1, \ldots, g_n)$ are independent standard Gaussian random variables and a_i real numbers) etc (cf. Chapter 5).

Now, it is easily seen that, for every $t \geq 0$,

$$P_t \Psi^{(d)}(f) = \mathrm{e}^{-dt} \Psi^{(d)}(f). \tag{8.22}$$

But we can then apply the hypercontractivity property of $(P_t)_{t \geq 0}$. Fix for example $p = 2$ and let $q = q(t) = 1 + \mathrm{e}^{2t}$. Then, combining (8.22) and (8.19) we get that, for every $q \geq 2$ or $t \geq 0$,

$$(q-1)^{-d/2} \big\| \Psi^{(d)}(f) \big\|_q = \mathrm{e}^{-dt} \big\| \Psi^{(d)}(f) \big\|_q = \big\| P_t \Psi^{(d)}(f) \big\|_q \leq \big\| \Psi^{(d)}(f) \big\|_2 \tag{8.23}$$

The next step in this development is that (8.23) applies in the same way to vector valued functions. Let E be a Banach space with norm $\|\cdot\|$. Given f on $\mathbb{R}^n$ with values in E, the previous chaotic decomposition is entirely similar. We need then simply apply hypercontractivity to the real valued function $\|f\|$ and Jensen's inequality immediately shows that (8.19) also holds for E-valued functions, with the $\mathrm{L}^p(\gamma_n)$-norms replaced by $\mathrm{L}^p(\gamma_n; E)$-norms. In particular, if $e_1, \ldots, e_n$ are elements of E, the vector valued version of (8.23) for $d = 1$ for example implies that, for every $q \geq 2$,

$$\bigg\| \sum_{i=1}^{n} g_i e_i \bigg\|_q \leq (q-1)^{1/2} \bigg\| \sum_{i=1}^{n} g_i e_i \bigg\|_2 . \tag{8.24}$$

These inequalities are exactly the moment equivalences (4.5) which we obtain next to Theorem 4.1, with the same behavior of the constant as q increases to infinity (and with even a better numerical value). Since this constant is independent of n, it is not difficult to see (although we will not go into these details) that (8.24) essentially allows us to recover the integrability properties and tail behaviors of norms of Gaussian random vectors (Theorem 4.1) as well as of Wiener chaos (cf. Chapter 5). This very interesting and powerful line of reasoning was extensively developed by C. Borell to which we refer the interested reader ([Bo8], [Bo9]). Note that these hypercontractivity ideas may also be used in the context of the two point space to recover, for example, inequalities (3.6) [Bon], [Bo6].

Recently, a parallel approach was developed by S. Aida, T. Masuda and I. Shigekawa [A-M-S], but on the basis of logarithmic Sobolev inequalities rather than hypercontractivity. As we already noticed it, we established both the concentration of measure phenomenon for γ_n (Proposition 2.1) and the logarithmic Sobolev inequality (8.20) on the basis of the same commutation property $\nabla P_t f = \mathrm{e}^{-t} P_t(\nabla f)$ of the Ornstein-Uhlenbeck semigroup. In [A-M-S], it is actually shown that concentration follows from a logarithmic Sobolev inequality and hypercontractivity. Although the paper [A-M-S] is concerned with logarithmic Sobolev inequalities in an abstract Dirichlet space setting, let us restrict again to the Gaussian case to sketch the idea and show how (2.7) may be deduced from the logarithmic Sobolev inequality. (The implication is thus only formal, as this whole chapter actually.) Let thus f be a Lipschitz map on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \leq 1$ and mean zero. Let us apply the logarithmic Sobolev inequality (8.20) to $\mathrm{e}^{\lambda f/2}$ for every $\lambda \in \mathbb{R}$. Setting

$$\varphi(\lambda) = \int \mathrm{e}^{\lambda f} d\gamma_n, \quad \lambda \in \mathbb{R},$$

we see that

$$\lambda\varphi'(\lambda) - \varphi(\lambda)\log\varphi(\lambda) \leq \tfrac{1}{2}\,\lambda^2\varphi(\lambda), \quad \lambda \in \mathbb{R}.$$

We need then simply integrate this differential inequality (this was first done in [Da-S], originally by I. Herbst). Set $\psi(\lambda) = \frac{1}{\lambda}\log\varphi(\lambda)$, $\lambda > 0$. Hence, for every $\lambda > 0$, $\psi'(\lambda) \leq \frac{1}{2}$. Since $\psi(0) = \varphi'(0)/\varphi(0) = \int f d\gamma_n = 0$, it follows that

$$\psi(\lambda) \leq \frac{\lambda}{2}$$

for every $\lambda \geq 0$. Therefore, we have obtained (2.7), that is

$$\int \mathrm{e}^{\lambda f} d\gamma_n \leq \mathrm{e}^{\lambda^2/2}$$

for every $\lambda \geq 0$ and, replacing f by $-f$, also for all $\lambda \in \mathbb{R}$.

As we discussed it in Chapter 2, there is however a long way from concentration to true isoperimetry. To complete this chapter, we turn to the isoperimetric inequality (8.15) itself which we would like to analyze with the Ornstein-Uhlenbeck semigroup as we did in the classical case in the first part of this chapter. The next proposition, implicit in [Pi1, p. 180], is the first step towards our goal and is the analogue of Proposition 8.1. Given Borel sets A, B in $\mathbb{R}^n$ and $t \geq 0$, we set

$$K_t(A,B) = \int_A P_t(I_B)\,d\gamma_n.$$

Note that $K_t(A,A) = \|P_{t/2}(I_A)\|_2^2$. The notation K_t is used in analogy with that of a kernel. Large deviation estimates of the kernel $K_t(A,B)$ for the Wiener measure when $d(A,B) > 0$ are developed at the end of Chapter 4.

Proposition 8.5. *For every Borel set A in $\mathbb{R}^n$ with smooth boundary ∂A and every $t \geq 0$,*

$$K_t(A, A^c) \leq (2\pi)^{-1/2}\arccos(\mathrm{e}^{-t})\,\mathcal{O}_{n-1}(\partial A).$$

Proof. It is similar to the proof of Proposition 8.1. Let f, g be smooth functions on $\mathbb{R}^n$. For every $t \geq 0$, we can write

$$\begin{aligned}\int g\,(P_t f - f)\,d\gamma_n &= \int_0^t \left(\int g\,LP_s f\,d\gamma_n\right) ds \\ &= -\int_0^t \left(\int \langle \nabla P_s g, \nabla f\rangle d\gamma_n\right) ds.\end{aligned}$$

Now, by integration by parts on the representation of P_s using the Gaussian density,

$$\nabla P_s f = \frac{\mathrm{e}^{-s}}{(1-\mathrm{e}^{-2s})^{1/2}} \int_{\mathbb{R}^n} y\, g\big(\mathrm{e}^{-s}x + (1-\mathrm{e}^{-2s})^{1/2}y\big)\,d\gamma_n(y)\,.$$

Hence

$$\begin{aligned}&\int g(P_t f - f)\,d\gamma_n \\ &\quad = -\int_0^t \frac{\mathrm{e}^{-s}}{(1-\mathrm{e}^{-2s})^{1/2}} \iint \langle \nabla f(x), y\rangle\, g\big(\mathrm{e}^{-s}x + (1-\mathrm{e}^{-2s})^{1/2}y\big)\,d\gamma_n(x)d\gamma_n(y)ds.\end{aligned}$$

This identity of course extends to $g = I_{A^c}$. Since

$$\iint \langle \nabla f(x), y\rangle d\gamma_n(x)d\gamma_n(y) = 0\,,$$

we see that, for every s,

$$\begin{aligned}-\iint \langle \nabla f(x), y\rangle I_{A^c}\big(\mathrm{e}^{-s}x &+ (1-\mathrm{e}^{-2s})^{1/2}y\big)d\gamma_n(x)d\gamma_n(y) \\ &\leq \iint \big(\langle \nabla f(x), y\rangle\big)^- d\gamma_n(x)d\gamma_n(y) \\ &= \frac{1}{2}\iint \big|\langle \nabla f(x), y\rangle\big| d\gamma_n(x)d\gamma_n(y) \\ &= \frac{1}{\sqrt{2\pi}}\int |\nabla f| d\gamma_n\,.\end{aligned}$$

The conclusion follows by letting f approximate I_A since then $\int |\nabla f| d\gamma_n$ will approach $\mathcal{O}_{n-1}(\partial A)$ when ∂A is smooth enough. Proposition 8.5 is established. □

The inequality of the proposition is sharp in many respects. When $t \to \infty$, it reads

$$\mathcal{O}_{n-1}(\partial A) \geq 2\left(\frac{2}{\pi}\right)^{1/2} \gamma_n(A)\big(1-\gamma_n(A)\big), \tag{8.25}$$

that is, when $\gamma_n(A) = \frac{1}{2}$, the maximum of the function $\varphi_1 \circ \Phi^{-1}(x)$ at $x = \frac{1}{2}$. Inequality (8.25) may actually be interpreted as Cheeger's isoperimetric constant [Ch] of the Gauss space $(\mathbb{R}^n, \gamma_n)$. It is responsible for the optimal factor $\pi/2$ which appears in the vector valued inequalities (4.10). Indeed, one may integrate (8.25) by the coarea formula (see [Ya]) to get that for every smooth function f with mean zero,

$$\int |f|\, d\gamma_n \le \Big(\frac{\pi}{2}\Big)^{1/2} \int |\nabla f|\, d\gamma_n,$$

an inequality which is easily seen to be best possible (take $n = 1$ and f on $\mathbb{R}$ be defined by $f(x) = x/\varepsilon$ for $|x| \le \varepsilon$, $f(x) = x/|x|$ elsewhere, and let $\varepsilon \to 0$).

Proposition 8.5 may also be tested on half-spaces, as we did on balls in the classical case. Namely, if we let $H = \{x \in \mathbb{R}^n; \langle x, u\rangle \le a\}$, $|u| = 1$, $a \in \mathbb{R}$, it is easily checked (start in dimension one and use polar coordinates) that

$$\begin{aligned} K_t(H, H^c) &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} \mathrm{e}^{-x^2/2} \mathrm{e}^{-y^2/2} I_{\{x \le |a|,\, \mathrm{e}^{-t}x + (1-\mathrm{e}^{-2t})^{1/2} y > |a|\}} dx dy \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty \rho\, \mathrm{e}^{-\rho^2/2} I_{\{\rho \sin(\varphi) \le |a|,\, \rho \sin(\varphi + \theta) > |a|\}} d\varphi d\rho \\ &= \frac{\theta}{2\pi}\, \mathrm{e}^{-a^2/2} - \frac{1}{2\pi} \int_{|a|}^{|a|/\sin((\pi-\theta)/2)} \big(2 \arcsin\big(\rho^{-1}|a|\big) + \theta - \pi\big) \rho\, \mathrm{e}^{-\rho^2/2} d\rho \end{aligned}$$

where $\theta = \arccos(\mathrm{e}^{-t})$. The absolute value of the second term of the latter may be bounded by

$$\frac{\theta}{2\pi} \big(\mathrm{e}^{-a^2/2} - \mathrm{e}^{-a^2/2\cos^2(\theta/2)}\big) \le \frac{\theta}{2\pi} \cdot \frac{a^2}{2} \tan^2\Big(\frac{\theta}{2}\Big) \mathrm{e}^{-a^2/2} \le \frac{\theta^3}{2\pi}\, a^2\, \mathrm{e}^{-a^2/2}$$

at least for all θ small enough. In particular, since $\theta = \arccos(\mathrm{e}^{-t})$ and thus $\theta \sim \sqrt{2t}$ when $t \to 0$, it follows that

$$\lim_{t\to 0} (2\pi)^{1/2} \big[\arccos(\mathrm{e}^{-t})\big]^{-1} K_t(H, H^c) = \mathcal{O}_{n-1}(\partial H). \tag{8.26}$$

On the basis of Proposition 8.5, we now would need lower estimates of the functional $K_t(A, A^c)$ for the small values of t. The typical isoperimetric approach would be to use a symmetrization result of C. Borell [Bo10], analogous to (8.7), asserting that if H is a half-space with the same measure as A, then for every $t \ge 0$,

$$K_t(A, A) \le K_t(H, H). \tag{8.27}$$

Hence $K_t(A, A^c) \ge K_t(H, H^c)$ and we would conclude from Proposition 8.5 and (8.26) that

$$\mathcal{O}_{n-1}(\partial A) \ge \mathcal{O}_{n-1}(\partial H).$$

In particular, and as in the classical case, isoperimetry is therefore equivalent to saying that

$$\big\|P_t(I_A)\big\|_2 \le \big\|P_t(I_H)\big\|_2, \quad t \ge 0\,, \tag{8.28}$$

for H a half-space with the same measure as A. This inequality is established in [Bo10], extending ideas of [Eh2] and based on techniques developed in the classical case in [Ba], via the Gaussian isoperimetric inequality. It might be that a simple direct approach to (8.28) (and also (8.7) as we mentioned it) is possible.

Our approach to bound $K_t(A,A)$ will be to use hypercontractivity as the corresponding semigroup estimate in this Gaussian setting. Namely, we simply write for A a Borel set in $\mathbb{R}^n$ and $p(t) = 1 + \mathrm{e}^{-t}$ that

$$K_t(A,A) = \big\|P_{t/2}(I_A)\big\|_2^2 \le \big\|I_A\big\|_{p(t)}^2, \quad t \ge 0. \tag{8.29}$$

Hence

$$K_t(A,A^c) \ge \gamma_n(A)\big[1 - \gamma_n(A)^{(2/p(t))-1}\big].$$

Therefore, combined with Proposition 8.5,

$$\mathcal{O}_{n-1}(\partial A) \ge (2\pi)^{1/2}\gamma_n(A) \sup_{t>0}\big[\big(\arccos(\mathrm{e}^{-t})\big)^{-1}\big(1 - \gamma_n(A)^{(2/p(t))-1}\big)\big].$$

Setting $\theta = \arccos(\mathrm{e}^{-t}) \in (0, \frac{\pi}{2}]$ we need to evaluate

$$\sup_{0<\theta\le\frac{\pi}{2}} \frac{1}{\theta}\left[1 - \exp\left(-\frac{1-\cos\theta}{1+\cos\theta}\,\log\frac{1}{\gamma_n(A)}\right)\right].$$

To this aim, we can note for example that

$$\frac{1-\cos\theta}{1+\cos\theta} \ge \frac{\theta^2}{2\pi}\,,$$

and choosing thus θ of the form

$$\theta = (2\pi)^{1/2}\left(\log\frac{1}{\gamma_n(A)}\right)^{-1/2},$$

provided that $\gamma_n(A) \le \mathrm{e}^{-8/\pi}$, we find that

$$\mathcal{O}_{n-1}(\partial A) \ge \left(1 - \frac{1}{\mathrm{e}}\right)\gamma_n(A)\left(\log\frac{1}{\gamma_n(A)}\right)^{1/2}.$$

Due to the equivalence (8.16), there exists $\delta > 0$ such that when $\gamma_n(A) \le \delta$,

$$\mathcal{O}_{n-1}(\partial A) \ge \frac{1}{3}\,\varphi_1 \circ \Phi^{-1}\big(\gamma_n(A)\big).$$

When $\delta < \gamma_n(A) \le 1/2$, we can always use (8.25) to get

$$\mathcal{O}_{n-1}(\partial A) \ge \left(\frac{\pi}{2}\right)\gamma_n(A) \ge c(\delta)\,\varphi_1 \circ \Phi^{-1}\big(\gamma_n(A)\big)$$

for some $c(\delta) > 0$. These two inequalities, together with symmetry, yield that, for some numerical constant $0 < c < 1$ and all subsets A in $\mathbb{R}^n$ with smooth boundary,

$$\mathcal{O}_{n-1}(\partial A) \ge c\,\varphi_1 \circ \Phi^{-1}\big(\gamma_n(A)\big). \tag{8.30}$$

One may try to tighten the preceding computations to reach the value $c = 1$ in (8.30). This however does not seem likely and it is certainly in the hypercontractive estimate (8.29) that a good deal of the best constant is lost. One may wonder why this is the case. It seems that hypercontractivity, while an equality on exponential functions, is perhaps not that sharp on indicator functions. This would have to be understood in connection with (8.28). Note finally that one may integrate back (8.30) to obtain, with these functional tools, the following analogue of the Gaussian isoperimetric inequality: if $\gamma_n(A) \geq \Phi(a)$, for every $r \geq 0$,

$$\gamma_n(A_r) \geq \Phi(a + cr).$$

(As noticed by S. Bobkov [Bob] in a related context, this may be shown to hold for every Borel set.) It is likely that the preceding approach has some interesting consequences in more abstract settings.

It might be worthwhile noting finally that Ehrhard's tensorization argument together with symmetrization may also be used to establish directly hypercontractivity, a comment we learned from C. Borell. One approach through logarithmic Sobolev inequalities is developed in [Eh4]. Alternatively, by the result of [Bo10],

$$\int gP_t f d\gamma_n \leq \int g^* P_t f^* d\gamma_1$$

for every $t \geq 0$ and every f, g say in $\mathrm{L}^2(\gamma_n)$ where f^* denotes the (one-dimensional) nonincreasing rearrangement of f with respect to the Gaussian measure γ_n (see [Eh3], [Bo10]). If $1 < p < q < \infty$ and $q < 1 + (p-1)\mathrm{e}^{2t}$, a trivial application of Hölder's inequality shows that, for every φ in $\mathrm{L}^p(\gamma_1)$,

$$\|P_t \varphi\|_q \leq C \|\varphi\|_p$$

for some numerical $C > 0$. Now, if q' is the conjugate of q,

$$\begin{aligned} \int gP_t f d\gamma_n &\leq \int g^* P_t f^* d\gamma_1 \\ &\leq \|g^*\|_{q'} \|P_t f^*\|_q \\ &\leq C \|g^*\|_{q'} \|f^*\|_p \leq C \|g\|_{q'} \|f\|_p \end{aligned}$$

so that, by duality,

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

Applying this inequality to $f^{\otimes k}$ on $(\mathbb{R}^n)^k = \mathbb{R}^{nk}$ yields

$$\|P_t f\|_q \leq C^{1/k} \|f\|_p.$$

Letting k tend to infinity, and q to its optimal value $1 + (p-1)\mathrm{e}^{2t}$ concludes the proof of the claim.

References

[A-C] R. A. Adams, F. H. Clarke. Gross's logarithmic Sobolev inequality: a simple proof. Amer. J. Math. 101, 1265–1269 (1979).

[A-K-S] S. Aida, S. Kusuoka, D. Stroock. On the support of Wiener functionals. Asymptotic problems in probability theory: Wiener functionals and asymptotics. Pitman Research Notes in Math. Series 284, 1–34 (1993). Longman.

[A-M-S] S. Aida, T. Masuda, I. Shigekawa. Logarithmic Sobolev inequalities and exponential integrability. J. Funct. Anal. 126, 83–101 (1994).

[A-L-R] M. Aizenman, J. L. Lebowitz, D. Ruelle. Some rigorous results on the Sherrington-Kirkpatrick spin glass model. Comm. Math. Phys. 112, 3–20 (1987).

[An] T. W. Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. 6, 170–176 (1955).

[A-G] M. Arcones, E. Giné. On decoupling, series expansions and tail behavior of chaos processes. J. Theoretical Prob. 6, 101–122 (1993).

[Az] R. Azencott. Grandes déviations et applications. École d'Été de Probabilités de St-Flour 1978. Lecture Notes in Math. 774, 1–176 (1978). Springer-Verlag.

[Azu] K. Azuma. Weighted sums of certain dependent random variables. Tohoku Math. J. 19, 357–367 (1967).

[B-C] A. Badrikian, S. Chevet. Mesures cylindriques, espaces de Wiener et fonctions aléatoires gaussiennes. Lecture Notes in Math. 379, (1974). Springer-Verlag.

[Ba] A. Baernstein II. Integral means, univalent functions and circular symmetrization. Acta Math. 133, 139–169 (1974).

[B-T] A. Baernstein II, B. A. Taylor. Spherical rearrangements, subharmonic functions and *-functions in n-space. Duke Math. J. 43, 245–268 (1976).

[Bak] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. École d'Été de Probabilités de St-Flour. Lecture Notes in Math. 1581, 1–114 (1994). Springer-Verlag.

[B-É] D. Bakry, M. Émery. Diffusions hypercontractives. Séminaire de Probabilités XIX. Lecture Notes in Math. 1123, 175–206 (1985). Springer-Verlag.

[B-R] P. Baldi, B. Roynette. Some exact equivalents for Brownian motion in Hölder norm. Prob. Th. Rel. Fields 93, 457–484 (1992).

[B-BA-K] P. Baldi, G. Ben Arous, G. Kerkyacharian. Large deviations and the Strassen theorem in Hölder norm. Stochastic Processes and Appl. 42, 171–180 (1992).

[Bas] R. Bass. Probability estimates for multiparameter Brownian processes. Ann. Probability 16, 251–264 (1988).

[Be1] W. Beckner. Inequalities in Fourier analysis. Ann. Math. 102, 159–182 (1975).

[Be2] W. Beckner. Unpublished (1982).

[Bel] D. R. Bell. The Malliavin calculus. Pitman Monographs 34. Longman (1987).

[BA-L1] 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.

[BA-L2] G. Ben Arous, M. Ledoux. Grandes déviations de Freidlin-Wentzell en norme hölderienne. Séminaire de Probabilités XXVIII. Lecture Notes in Math. 1583, 293–299 (194). Springer-Verlag.

[BA-G-L] G. Ben Arous, M. Gradinaru, M. Ledoux. Hölder norms and the support theorem for diffusions. Ann. Inst. H. Poincaré 30, 415–436 (1994).

[Bob] S. Bobkov. A functional form of the isoperimetric inequality for the Gaussian measure. Preprint (1993).

[Bon] A. Bonami. Etude des coefficients de Fourier des fonctions de $L^p(G)$. Ann. Inst. Fourier 20, 335–402 (1970).

[Bo1] C. Borell. Convex measures on locally convex spaces. Ark. Mat. 12, 239–252 (1974).

[Bo2] C. Borell. The Brunn-Minskowski inequality in Gauss space. Invent. Math. 30, 207–216 (1975).

[Bo3] C. Borell. Gaussian Radon measures on locally convex spaces. Math. Scand. 38, 265–284 (1976).

[Bo4] C. Borell. A note on Gauss measures which agree on small balls. Ann. Inst. H. Poincaré 13, 231–238 (1977).

[Bo5] C. Borell. Tail probabilities in Gauss space. Vector Space Measures and Applications, Dublin 1977. Lecture Notes in Math. 644, 71–82 (1978). Springer-Verlag.

[Bo6] C. Borell. On the integrability of Banach space valued Walsh polynomials. Séminaire de Probabilités XIII. Lecture Notes in Math. 721, 1–3 (1979). Springer-Verlag.

[Bo7] C. Borell. A Gaussian correlation inequality for certain bodies in $\mathbb{R}^n$. Math. Ann. 256, 569–573 (1981).

[Bo8] C. Borell. On polynomials chaos and integrability. Prob. Math. Statist. 3, 191–203 (1984).

[Bo9] C. Borell. On the Taylor series of a Wiener polynomial. Seminar Notes on multiple stochastic integration, polynomial chaos and their integration. Case Western Reserve University, Cleveland (1984).

[Bo10] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck process. Z. Wahrscheinlichkeitstheor. verw. Gebiete 70, 1–13 (1985).

[Bo11] C. Borell. Analytic and empirical evidences of isoperimetric processes. Probability in Banach spaces 6. Progress in Probability 20, 13–40 (1990). Birkhäuser.

[B-M] A. Borovkov, A. Mogulskii. On probabilities of small deviations for stochastic processes. Siberian Adv. Math. 1, 39–63 (1991).

[B-Z] Y. D. Burago, V. A. Zalgaller. Geometric inequalities. Springer-Verlag (1988). First Edition (russian): Nauka (1980).

[C-M] R. H. Cameron, W. T. Martin. Transformations of Wiener integrals under translations. Ann. Math. 45, 386–396 (1944).

[Ca] M. Capitaine. Onsager-Machlup functional for some smooth norms on Wiener space (1994). To appear in Prob. Th. Rel. Fields.

[C-F] I. Chavel, E. Feldman. Modified isoperimetric constants, and large time heat diffusion in Riemannian manifold. Duke Math. J. 64, 473–499 (1991).

[Ch] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. Problems in Analysis, Symposium in honor of S. Bochner, 195–199, Princeton Univ. Press, Princeton (1970).

[C-L-Y] S. Cheng, P. Li, S.-T. Yau. On the upper estimate of the heat kernel on a complete Riemannian manifold. Amer. J. Math. 156, 153–201 (1986).

[Che] S. Chevet. Gaussian measures and large deviations. Probability in Banach spaces IV. Lecture Notes in Math. 990, 30–46 (1983). Springer-Verlag.

[Ci1] Z. Ciesielski. On the isomorphisms of the spaces H_α and m. Bull. Acad. Pol. Sc. 8, 217–222 (1960).

[Ci2] Z. Ciesielski. Orlicz spaces, spline systems and brownian motion. Constr. Approx. 9, 191–208 (1993).

[Co] F Comets. A spherical bound for the Sherrington-Kirkpatrick model. Preprint (1994).

[C-N] F. Comets, J. Neveu. The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. Preprint (1993).

[C-L] T. Coulhon, M. Ledoux. Isopérimétrie, décroissance du noyau de la chaleur et transformations de Riesz: un contre-exemple. Ark. Mat. 32, 63–77 (1994).

[DG-E-...] S. Das Gupta, M. L. Eaton, I. Olkin, M. Perlman, L. J. Savage, M. Sobel. Inequalities on the probability content of convex regions for elliptically contoured distributions. Proc. Sixth Berkeley Symp. Math. Statist. Prob. 2, 241–264 (1972). Univ. of California Press.

[Da] E. B. Davies. Heat kernels and spectral theory. Cambridge Univ. Press (1989).

[Da-S] E. B. Davies, B. Simon. Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. J. Funct. Anal. 59, 335–395 (1984).

[D-L] P. Deheuvels, M. A. Lifshits. Strassen-type functional laws for strong topologies. Prob. Th. Rel. Fields 97, 151–167 (1993).

[De] J. Delporte. Fonctions aléatoires presque sûrement continues sur un intervalle fermé. Ann. Inst. H. Poincaré 1, 111–215 (1964).

[D-S] J.-D. Deuschel, D. Stroock. Large deviations. Academic Press (1989).

[D-F] P. Diaconis, D. Freedman. A dozen de Finetti-style results in search of a theory. Ann. Inst. H. Poincaré 23, 397–423 (1987).

[D-V] M. D. Donsker, S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time III. Comm. Pure Appl. Math. 29, 389–461 (1976).

[Du1] R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J. Funct. Anal. 1, 290–330 (1967).

[Du2] R. M. Dudley. Sample functions of the Gaussian process. Ann. Probability 1, 66–103 (1973).

[D-HJ-S] R. M. Dudley, J. Hoffmann-Jorgensen, L. A. Shepp. On the lower tail of Gaussian seminorms. Ann. Probability 7, 319–342 (1979).

[Dv] A. Dvoretzky. Some results on convex bodies and Banach spaces. Proc. Symp. on Linear Spaces, Jerusalem, 123–160 (1961).

[Eh1] A. Ehrhard. Une démonstration de l'inégalité de Borell. Ann. Scientifiques de l'Université de Clermont-Ferrand 69, 165–184 (1981).

[Eh2] A. Ehrhard. Symétrisation dans l'espace de Gauss. Math. Scand. 53, 281–301 (1983).

[Eh3] A. Ehrhard. Inégalités isopérimétriques et intégrales de Dirichlet gaussiennes. Ann. scient. Éc. Norm. Sup. 17, 317–332 (1984).

[Eh4] A. Ehrhard. Sur l'inégalité de Sobolev logarithmique de Gross. Séminaire de Probabilités XVIII. Lecture Notes in Math. 1059, 194–196 (1984). Springer-Verlag.

[Eh5] A. Ehrhard. Eléments extrémaux pour les inégalités de Brunn-Minskowski gaussiennes. Ann. Inst. H. Poincaré 22, 149–168 (1986).

[E-S] O. Enchev, D. Stroock. Rademacher's theorem for Wiener functionals. Ann. Probability 21, 25–33 (1993).

[Fa] S. Fang. On the Ornstein-Uhlenbeck process. Stochastics and Stochastic Reports 46, 141–159 (1994).

[Fed] H. Federer. Geometric measure theory. Springer-Verlag (1969).

[F-F] H. Federer, W. H. Fleming. Normal and integral current. Ann. Math. 72, 458–520 (1960).

[Fe1] X. Fernique. Continuité des processus gaussiens. C. R. Acad. Sci. Paris 258, 6058–6060 (1964).

[Fe2] X. Fernique. Intégrabilité des vecteurs gaussiens. C. R. Acad. Sci. Paris 270, 1698–1699 (1970).

[Fe3] X. Fernique. Régularité des processus gaussiens. Invent. Math. 12, 304–320 (1971).

[Fe4] X. Fernique. Régularité des trajectoires des fonctions aléatoires gaussiennes. École d'Été de Probabilités de St-Flour 1974. Lecture Notes in Math. 480, 1–96 (1975). Springer-Verlag.

[Fe5] X. Fernique. Gaussian random vectors and their reproducing kernel Hilbert spaces. Technical report, University of Ottawa (1985).

[F-L-M] T. Figiel, J. Lindenstrauss, V. D. Milman. The dimensions of almost spherical sections of convex bodies. Acta Math. 139, 52–94 (1977).

[F-W1] M. Freidlin, A. Wentzell. On small random perturbations of dynamical systems. Russian Math. Surveys 25, 1–55 (1970).

[F-W2] M. Freidlin, A. Wentzell. Random perturbations of dynamical systems. Springer-Verlag (1984).

[Ga] E. Gagliardo. Proprieta di alcune classi di funzioni in piu variabili. Ricerche Mat. 7, 102–137 (1958).

[G-R-R] A. M. Garsia, E. Rodemich, H. Rumsey Jr.. A real variable lemma and the continuity of paths of some Gaussian processes. Indiana Math. J. 20, 565–578 (1978).

[Gal] L. Gallardo. Au sujet du contenu probabiliste d'un lemme d'Henri Poincaré. Ann. Scientifiques de l'Université de Clermont-Ferrand 69, 185–190 (1981).

[G-H-L] S. Gallot, D. Hulin, J. Lafontaine. Riemannian Geometry. Second Edition. Springer-Verlag (1990).

[Go1] V. Goodman. Characteristics of normal samples. Ann. Probability 16, 1281–1290 (1988).

[Go2] V. Goodman. Some probability and entropy estimates for Gaussian measures. Probability in Banach spaces 6. Progress in Probability 20, 150–156 (1990). Birkhäuser.

[G-K1] V. Goodman, J. Kuelbs. Cramér functional estimates for Gaussian measures. Diffusion processes and related topics in Analysis. Progress in Probability 22, 473–495 (1990). Birkhäuser.

[G-K2] V. Goodman, J. Kuelbs. Gaussian chaos and functional laws of the iterated logarithm for Ito-Wiener integrals. Ann. Inst. H. Poincaré 29, 485–512 (1993).

[Gri] K. Grill. Exact convergence rate in Strassen's law of the iterated logarithm. J. Theoretical Prob. 5, 197–204 (1991).

[Gro] M. Gromov. Paul Lévy's isoperimetric inequality. Preprint I.H.E.S. (1980).

[G-M] M. Gromov, V. D. Milman. A topological application of the isoperimetric inequality. Amer. J. Math. 105, 843–854 (1983).

[Gr1] L. Gross. Abstract Wiener spaces. Proc. 5th Berkeley Symp. Math. Stat. Prob. 2, 31–42 (1965).

[Gr2] L. Gross. Potential theory on Hilbert space. J. Funct. Anal. 1, 123–181 (1967).

[Gr3] L. Gross. Logarithmic Sobolev inequalities. Amer. J. Math. 97, 1061–1083 (1975).

[Gr4] L. Gross. Logarithmic Sobolev inequalities and contractive properties of semigroups. Dirichlet forms, Varenna (Italy) 1992. Lecture Notes in Math. 1563, 54–88 (1993). Springer-Verlag.

[G-N-SS] I. Gyöngy, D. Nualart, M. Sanz-Solé. Approximation and support theorems in modulus spaces. Preprint (1994).

[Ha] H. Hadwiger. Vorlesungen über Inhalt, Oberfläche und Isoperimetrie. Springer-Verlag (1957).

[Har] L. H. Harper. Optimal numbering and isoperimetric problems on graphs. J. Comb. Th. 1, 385–393 (1966).

[He] B. Heinkel. Mesures majorantes et régularité de fonctions aléatoires. Aspects Statistiques et Aspects Physiques des Processus Gaussiens, St-Flour 1980. Colloque C.N.R.S. 307, 407–434 (1980).

[I-S-T] I. A. Ibragimov, V. N. Sudakov, B. S. Tsirel'son. Norms of Gaussian sample functions. Proceedings of the third Japan-USSR Symposium on Probability Theory. Lecture Notes in Math. 550, 20–41 (1976). Springer-Verlag.

[I-W] N. Ikeda, S. Watanabe. Stochastic differential equations and diffusion processes. North-Holland (1989).

[It] K. Itô. Multiple Wiener integrals. J. Math. Soc. Japan 3, 157–164 (1951).

[Ka1] J.-P. Kahane. Sur les sommes vectorielles $\sum \pm u_n$. C. R. Acad. Sci. Paris 259, 2577–2580 (1964).

[Ka2] J.-P. Kahane. Some random series of functions. Heath Math. Monographs (1968). Second Edition: Cambridge Univ. Press (1985).

[Ke] H. Kesten. On the speed of convergence in first-passage percolation. Ann. Appl. Probability 3, 296–338 (1993).

[Kh] C. Khatri. On certain inequalities for normal distributions and their applications to simultaneous confidence bounds. Ann. Math. Statist. 38, 1853–1867 (1967).

[K-L1] J. Kuelbs, W. Li. Small ball probabilities for Brownian motion and the Brownian sheet. J. Theoretical Prob. 6, 547–577 (1993).

[K-L2] J. Kuelbs, W. Li. Metric entropy and the small ball problem for Gaussian measures J. Funct. Anal. 116, 133-157 (1993).

[K-L-L] J. Kuelbs, W. Li, W. Linde. The Gaussian measure of shifted balls. Prob. Th. Rel. Fields 98, 143–162 (1994).

[K-L-S] J. Kuelbs, W. Li, Q.-M. Shao. Small ball probabilities for Gaussian processes with stationary increments under Hölder norms. Preprint (1993).

[K-L-T] J. Kuelbs, W. Li, M. Talagrand. Liminf results for Gaussian samples and Chung's functional LIL (1992). Ann. Probability, to appear.

[Ku] H.-H. Kuo. Gaussian measures in Banach spaces. Lecture Notes in Math. 436 (1975). Springer-Verlag.

[Kus] S. Kusuoka. A diffusion process on a fractal. Probabilistic methods in mathematical physics. Proc. of Taniguchi International Symp. 1985, 251–274. Kinokuniga, Tokyo (1987).

[Kw] S. Kwapień. A theorem on the Rademacher series with vector valued coefficients. Probability in Banach Spaces, Oberwolfach 1975. Lecture Notes in Math. 526, 157–158 (1976). Springer-Verlag.

[K-S] S. Kwapień, J. Sawa. On some conjecture concerning Gaussian measures of dilatations of convex symmetric sets. Studia Math. 105, 173–187 (1993).

[L-S] H. J. Landau, L. A. Shepp. On the supremum of a Gaussian process. Sankhyà A32, 369–378 (1970).

[La] R. Latała. A note on the Ehrhard inequality. Preprint (1994).

[L-O] R. Latała, K. Oleszkiewicz. On the best constant in the Khintchine-Kahane inequality. Studia Math. 109, 101–104 (1994).

[Led1] M. Ledoux. Isopérimétrie et inégalités de Sobolev logarithmiques gaussiennes. C. R. Acad. Sci. Paris 306, 79–82 (1988).

[Led2] M. Ledoux. A note on large deviations for Wiener chaos. Séminaire de Probabilités XXIV, Lecture Notes in Math. 1426, 1–14 (1990). Springer-Verlag.

[Led3] M. Ledoux. On an integral criterion for hypercontractivity of diffusion semigroups and extremal functions. J. Funct. Anal. 105, 444–465 (1992).

[Led4] M. Ledoux. A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold. Geom. and Funct. Anal. 2, 221–224 (1992).

[Led5] M. Ledoux. Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space (1992). Bull. Sc. math., to appear.

[L-T1] M. Ledoux, M. Talagrand. Characterization of the law of the iterated logarithm in Banach spaces. Ann. Probability 16, 1242–1264 (1988).

[L-T2] M. Ledoux, M. Talagrand. Probability in Banach spaces (Isoperimetry and processes). Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag (1991).

[Lé] P. Lévy. Problèmes concrets d'analyse fonctionnelle. Gauthier-Villars (1951).

[Li] W. Li. Comparison results for the lower tail of Gaussian semi-norms. J. Theoretical Prob. 5, 1–31 (1992).

[Li-S] W. Li, Q.-M. Shao. Small ball estimates for Gaussian processes under Sobolev type norms. Preprint (1994).

[L-Y] P. Li, S.-T. Yau. On the parabolic kernel of the Schrödinger operator. Acta Math. 156, 153–201 (1986).

[Lif1] M. A. Lifshits. On the distribution of the maximum of a Gaussian process. Probability Theory and its Appl. 31, 125-132 (1987).

[Lif2] M. A. Lifshits. Tail probabilities of Gaussian suprema and Laplace transform. Ann. Inst. H. Poincaré 30, 163–180 (1994).

[Lif3] M. A. Lifshits. Gaussian random functions. Kluwer (1994).

[Lif-T] M. A. Lifshits, B. S. Tsirel'son. Small deviations of Gaussian fields. Probability Theory and its Appl. 31, 557-558 (1987).

[L-Z] T. Lyons, W. Zheng. A crossing estimate for the canonical process on a Dirichlet space and tightness result. Colloque Paul Lévy, Astérisque 157-158, 249–272 (1988).

[MD] C. J. H. McDiarmid. On the method of bounded differences. Twelfth British Combinatorial Conference. Surveys in Combinatorics, 148–188 (1989). Cambrige Univ. Press.

[MK] H. P. McKean. Geometry of differential space. Ann. Probability 1, 197–206 (1973).

[M-P] M. B. Marcus, G. Pisier. Random Fourier series with applications to harmonic analysis. Ann. Math. Studies, vol. 101 (1981). Princeton Univ. Press.

[M-S] M. B. Marcus, L. A. Shepp. Sample behavior of Gaussian processes. Proc. of the Sixth Berkeley Symposium on Math. Statist. and Prob. 2, 423–441 (1972).

[Ma1] B. Maurey. Constructions de suites symétriques. C. R. Acad. Sci. Paris 288, 679–681 (1979).

[Ma2] B. Maurey. Sous-espaces ℓ^p des espaces de Banach. Séminaire Bourbaki, exp. 608. Astérisque 105-106, 199–216 (1983).

[Ma3] B. Maurey. Some deviations inequalities. Geometric and Funct. Anal. 1, 188–197 (1991).

[MW-N-PA] E. Mayer-Wolf, D. Nualart, V. Perez-Abreu. Large deviations for multiple Wiener-Itô integrals. Séminaire de Probabilités XXVI. Lecture Notes in Math. 1526, 11–31 (1992). Springer-Verlag.

[Maz1] V. G. Maz'ja. Classes of domains and imbedding theorems for function spaces. Soviet Math. Dokl. 1, 882–885 (1960).

[Maz2] V. G. Maz'ja. Sobolev spaces. Springer-Verlag (1985).

[Me] M. Mellouk. Support des diffusions dans les espaces de Besov-Orlicz. C. R. Acad. Sci. Paris 319, 261–266 (1994).

[M-SS] A. Millet, M. Sanz-Solé. A simple proof of the support theorem for diffusion processes. Séminaire de Probabilités XXVIII, Lecture Notes in Math. 1583, 36–48 (1994). Springer-Verlag.

[Mi1] V. D. Milman. New proof of the theorem of Dvoretzky on sections of convex bodies. Funct. Anal. Appl. 5, 28–37 (1971).

[Mi2] V. D. Milman. The heritage of P. Lévy in geometrical functional analysis. Colloque Paul Lévy sur les processus stochastiques. Astérisque 157-158, 273–302 (1988).

[Mi3] V. D. Milman. Dvoretzky's theorem - Thirty years later (Survey). Geometric and Funct. Anal. 2, 455–479 (1992).

[Mi-S] V. D. Milman, G. Schechtman. Asymptotic theory of finite dimensional normed spaces. Lecture Notes in Math. 1200 (1986). Springer-Verlag.

[M-R] D. Monrad, H. Rootzén. Small values of fractional Brownian motion and locally non-deterministic Gaussian processes. Preprint (1993).

[Mo] J. Moser. On Harnack's theorem for elliptic differential equations. Comm. Pure Appl. Math. 14, 557–591 (1961).

[Na] J. Nash. Continuity of solutions of parabolic and elliptic equations. Amer. J. Math. 80, 931–954 (1958).

[Nel] E. Nelson. The free Markov field. J. Funct. Anal. 12, 211–227 (1973).

[Ne1] J. Neveu. Processus aléatoires gaussiens. Presses de l'Université de Montréal (1968).

[Ne2] J. Neveu. Martingales à temps discret. Masson (1972).

[Ne3] J. Neveu. Sur l'espérance conditionnelle par rapport à un mouvement brownien. Ann. Inst. H. Poincaré 2, 105–109 (1976).

[Ni] L. Nirenberg. On elliptic partial differential equations. Ann. Sc. Norm. Sup. Pisa 13, 116–162 (1959).

[Nu] D. Nualart. The Malliavin calculs and related topics (1994). To appear.

[Os] R. Osserman. The isoperimetric inequality. Bull. Amer. Math. Soc. 84, 1182–1238 (1978).

[Pi1] G. Pisier. Probabilistic methods in the geometry of Banach spaces. Probability and Analysis, Varenna (Italy) 1985. Lecture Notes in Math. 1206, 167–241 (1986). Springer-Verlag.

[Pi2] G. Pisier. Riesz transforms : a simpler analytic proof of P. A. Meyer inequality. Séminaire de Probabilités XXII. Lecture Notes in Math. 1321, 485–501, Springer-Verlag (1988).

[Pi3] G. Pisier. The volume of convex bodies and Banach space geometry. Cambridge Univ. Press (1989).

[Pit] L. Pitt. A Gaussian correlation inequality for symmetric convex sets. Ann. Probability 5, 470–474 (1977).

[Pr1] C. Preston. Banach spaces arising from some integral inequalities. Indiana Math. J. 20, 997–1015 (1971).

[Pr2] C. Preston. Continuity properties of some Gaussian processes. Ann. Math. Statist. 43, 285–292 (1972).

[Sc] M. Schilder. Asymptotic formulas for Wiener integrals. Trans. Amer. Math. Soc. 125, 63–85 (1966).

[Sch] E. Schmidt. Die Brunn-Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie. Math. Nach. 1, 81–157 (1948).

[Sco] A. Scott. A note on conservative confidence regions for the mean value of multivariate normal. Ann. Math. Statist. 38, 278–280 (1967).

[S-Z1] L. A. Shepp, O. Zeitouni. A note on conditional exponential moments and Onsager-Machlup functionals. Ann. Probability 20, 652–654 (1992).

[S-Z2] L. A. Shepp, O. Zeitouni. Exponential estimates for convex norms and some applications. Barcelona seminar on Stochastic Analysis, St Feliu de Guixols 1991. Progress in Probability 32, 203–215 (1993). Birkhäuser.

[Sh] Q.-M. Shao. A note on small ball probability of a Gaussian process with stationary increments. J. Theoretical Prob. 6, 595–602 (1993).

[S-W] Q.-M. Shao, D. Wang. Small ball probabilities of Gaussian fields (1994). Preprint.

[Si] Z. Sidak. Rectangular confidence regions for the means of multivariate normal distributions. J. Amer. Statist. Assoc. 62, 626–633 (1967).

[Sk] A. V. Skorohod. A note on Gaussian measures in a Banach space. Theor. Probability Appl. 15, 519–520 (1970).

[Sl] D. Slepian. The one-sided barrier problem for Gaussian noise. Bell. System Tech. J. 41, 463–501 (1962).

[So] S. L. Sobolev. On a theorem in functional analysis. Amer. Math. Soc. Translations (2) 34, 39–68 (1963); translated from Mat. Sb. (N.S.) 4 (46), 471–497 (1938).

[St1] W. Stolz. Une méthode élémentaire pour l'évaluation de petites boules browniennes. C. R. Acad. Sci. Paris, 316, 1217–1220 (1993).

[St2] W. Stolz. Some small ball probabilities for Gaussian processes under non-uniform norms. Preprint (1994).

[Str] D. Stroock. Homogeneous chaos revisited. Séminaire de Probabilités XXI. Lecture Notes in Math. 1247, 1–7 (1987). Springer-Verlag.

[Su1] V. N. Sudakov. Gaussian measures, Cauchy measures and ε-entropy. Soviet Math. Dokl. 10, 310–313 (1969).

[Su2] V. N. Sudakov. Gaussian random processes and measures of solid angles in Hilbert spaces. Soviet Math. Dokl. 12, 412–415 (1971).

[Su3] V. N. Sudakov. A remark on the criterion of continuity of Gaussian sample functions. Proceedings of the Second Japan-USSR Symposium on Probability Theory . Lecture Notes in Math. 330, 444–454 (1973). Springer-Verlag.

[Su4] V. N. Sudakov. Geometric problems of the theory of infinite-dimensional probability distributions. Trudy Mat. Inst. Steklov 141 (1976).

[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).

[Sy] G. N. Sytaya. On some asymptotic representation of the Gaussian measure in a Hilbert space. Theory of Stochastic Processes (Kiev) 2, 94-104 (1974).

[Sz] S. Szarek. On the best constant in the Khintchine inequality. Studia Math. 58, 197–208 (1976).

[Tak] M. Takeda. On a martingale method for symmetric diffusion processes and its applications. Osaka J. Math. 26, 605–623 (1989).

[Ta1] M. Talagrand. Sur l'intégrabilité des vecteurs gaussiens. Z. Wahrscheinlichkeitstheor. verw. Gebiete 68, 1–8 (1984).

[Ta2] M. Talagrand. Regularity of Gaussian processes. Acta Math. 159, 99–149 (1987).

[Ta3] M. Talagrand. An isoperimetric theorem on the cube and the Khintchine-Kahane inequalities. Proc. Amer. Math. Soc. 104, 905–909 (1988).

[Ta4] M. Talagrand. Small tails for the supremum of a Gaussian process. Ann. Inst. H. Poincaré 24, 307–315 (1988).

[Ta5] M. Talagrand. Isoperimetry and integrability of the sum of independent Banach space valued random variables. Ann. Probability 17, 1546–1570 (1989).

[Ta6] M. Talagrand. A new isoperimetric inequality for product measure and the tails of sums of independent random variables. Geometric and Funct. Anal. 1, 211-223 (1991).

[Ta7] M. Talagrand. Simple proof of the majorizing measure theorem. Geometric and Funct. Anal. 2, 118–125 (1992).

[Ta8] M. Talagrand. On the rate of clustering in Strassen's law of the iterated logarithm. Probability in Banach spaces 8. Progress in Probability 30, 339–351 (1992). Birkhäuser.

[Ta9] M. Talagrand. New Gaussian estimates for enlarged balls. Geometric and Funct. Anal. 3, 502–526 (1993).

[Ta10] M. Talagrand. Regularity of infinitely divisible processes. Ann. Probability 21, 362–432 (1993).

[Ta11] M. Talagrand. Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem. Geometric and Funct. Anal. 3, 295-314 (1993).

[Ta12] M. Talagrand. The supremum of some canonical processes. Amer. Math. J. 116, 283–325 (1994).

[Ta13] M. Talagrand. Sharper bounds for Gaussian and empirical processes. Ann. Probability 22, 28–76 (1994).

[Ta14] M. Talagrand. Constructions of majorizing measures. Bernoulli processes and cotype (1992). Geometric and Funct. Anal., to appear.

[Ta15] M. Talagrand. The small ball problem for the Brownian sheet (1992). Ann. Probability, to appear.

[Ta16] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Preprint (1994).

[Ta17] M. Talagrand. Isoperimetry in product spaces: higher level, large sets. Preprint (1994).

[Ta18] M. Talagrand. Majorizing measures: the generic chaining. Preprint (1994).

[TJ] N. Tomczak-Jaegermann. Dualité des nombres d'entropie pour des opérateurs à valeurs dans un espace de Hilbert. C. R. Acad. Sci. Paris 305, 299–301 (1987).

[Var] S. R. S. Varadhan. Large deviations and applications. S. I. A. M. Philadelphia (1984).

[Va1] N. Varopoulos. Une généralisation du théorème de Hardy-Littlewood-Sobolev pour les espaces de Dirichlet. C. R. Acad. Sci. Paris 299, 651–654 (1984).

[Va2] N. Varopoulos. Hardy-Littlewood theory for semigroups. J. Funct. Anal. 63, 240–260 (1985).

[Va3] N. Varopoulos. Isoperimetric inequalities and Markov chains. J. Funct. Anal. 63, 215–239 (1985).

[Va4] N. Varopoulos. Small time Gaussian estimates of heat diffusion kernels. Part I: The semigroup technique. Bull. Sc. math. 113, 253–277 (1989).

[Va5] N. Varopoulos. Analysis and geometry on groups. Proceedings of the International Congress of Mathematicians, Kyoto (1990), vol. II, 951–957 (1991). Springer-Verlag.

[V-SC-C] N. Varopoulos, L. Saloff-Coste, T. Coulhon. Analysis and geometry on groups. Cambridge Univ. Press (1992).

[Wa] S. Watanabe. Lectures on stochastic differential equations and Malliavin calculus. Tata Institute of Fundamental Research Lecture Notes. Springer-Verlag (1984).

[W-W] D. L. Wang, P. Wang. Extremal configurations on a discrete torus and a generalization of the generalized Macaulay theorem. Siam J. Appl. Math. 33, 55–59 (1977).

[Wi] N. Wiener. The homogeneous chaos. Amer. Math. J. 60, 897–936 (1930).

[Ya] S.-T. Yau. Isoperimetric constants and the first eigenvalue of a compact Riemannian manifold. Ann. scient. Éc. Norm. Sup. 8, 487–507 (1975).

[Zo] V. M. Zolotarev. Asymptotic behavior of the Gaussian measure in ℓ^2. J. Sov. Math. 24, 2330-2334 (1986).