

# Some Isoperimetric Inequalities and Their Applications

*Michel Talagrand*

Equipe d'Analyse, E.R.A. au C.N.R.S. no. 294, Université Paris VI, 4 Place Jussieu  
F-75230 Paris Cedex 05, France

## 0. Introduction

Consider a product of measure spaces, provided with the product measure. Consider a subset  $A$  of this product, of measure at least one half. An important fact (the so-called concentration of measure phenomenon) is that even a small “enlargement” of  $A$  has measure very close to one. The inequalities we present describe this phenomenon for several notions of “enlargement”.

## 1. The Isoperimetric Inequality for Gaussian Measure

We denote by  $S_n$  the Euclidean sphere of  $\mathbb{R}^{n+1}$ , equipped with the geodesic distance  $\varrho$  and a rotation invariant probability  $\mu_n$ . For a (measurable) subset  $A$  of  $S_n$ , consider the set  $A_u$  of points of  $S_n$  within geodesic distance  $u$  of  $A$ . The isoperimetric inequality for the sphere, discovered by P. Lévy, is of fundamental importance. It states that  $\mu_n(A_u) \geq \mu_n(C_u)$ , where  $C$  is a cap (intersection of the sphere and of a half space) of the same measure as  $A$ .

We denote by  $\gamma_n$  the canonical Gaussian measure on  $\mathbb{R}^n$ , of density  $(2\pi)^{-n/2} e^{-\|x\|^2/2}$  with respect to Lebesgue measure. Observe the simple, but essential fact that  $\gamma_n$  is the product measure on  $\mathbb{R}^n$  when each factor is endowed with  $\gamma_1$ . It is an old observation, known as Poincaré lemma (although it does seem to be due to Maxwell) that, as  $N \rightarrow \infty$ , the projection of the normalized measure on  $\sqrt{N}S_N$  onto  $\mathbb{R}^n$  has  $\gamma_n$  as a limit. Therefore, it should not come as a surprise that Lévy's isoperimetric inequality on the sphere implies an isoperimetric inequality for  $\gamma_n$ . This was discovered independently by C. Borell [B1], and V. N. Sudakov and B. S. Tsirelson [S-T]. If we denote by  $A_u$  the set of points within Euclidean distance  $u$  of  $A$ , then  $\gamma_n(A_u) \geq \gamma_n(H_u)$ , where  $H_u$  is a half space with  $\gamma_n(H) = \gamma_n(A)$ . Taking this half space to be orthogonal to a coordinate axis, and remembering that  $\gamma_n$  is a product measure shows that if  $\gamma_n(H) = \gamma_1((-\infty, a])$ , then  $\gamma_n(H_u) = \gamma_1((-\infty, a + u])$ .

For simplicity, we set  $\Phi(u) = \gamma_1((-\infty, u]) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-x^2/2} dx$ . Thus we have

$$\text{if } \gamma_n(A) = \Phi(a), \text{ then } \gamma_n(A_u) \geq \Phi(a + u). \quad (1.1)$$

It is important to state this inequality, not only for the measure  $\gamma_n$ , but also for its infinite dimensional version  $\gamma$ , the product measure on  $\mathbb{R}^{\mathbb{N}}$  when each factor is endowed with  $\gamma_1$  (the result for  $\gamma$  follows from the result for  $\gamma_n$  and an obvious approximation). We denote by  $B$  the unit ball of  $\ell^2$ , i.e.  $B = \{x \in \mathbb{R}^{\mathbb{N}}, \sum_{k \geq 1} x_k^2 \leq 1\}$ . The Gaussian isoperimetric inequality can then be stated as follows

$$\text{If } \gamma(A) = \Phi(a) \text{ then } \gamma_*(A + uB) \geq \Phi(a + u). \quad (1.2)$$

There  $A + uB = \{x + uy; x \in A, y \in B\}$ ; the inner measure is needed as  $A + uB$  might not be measurable. As became customary, we call (1.1) and (1.2) Borell's inequality. Lévy's inequality is usually proved using symmetrization (see e.g. the appendix of [F-L-M]). A. Ehrhard [E1] has developed a symmetrization method adapted to the measures  $\gamma_n$  that yields a direct proof of (1.2) as well as of the following remarkable inequality of Brunn-Minkowski's type: For two convex sets  $A, B$  of  $\mathbb{R}^n$ , and  $0 \leq \lambda \leq 1$ ,

$$\Phi^{-1}(\gamma_n(\lambda A + (1 - \lambda)B)) \geq \lambda \Phi^{-1}(\gamma_n(A)) + (1 - \lambda) \Phi^{-1}(\gamma_n(B)). \quad (1.3)$$

(It is still open whether this inequality holds for non convex sets.)

Borell's inequality is a principle of remarkable power. It can be argued that, concerning applications, this inequality is used in two different forms.

The first type of use consist of rewriting (1.1) as  $u^{-1} \gamma_n(A_u \setminus A) \geq u^{-1} \gamma_1([a, u + a])$  so that

$$\liminf_{u \rightarrow 0} u^{-1} \gamma_n(A_u \setminus A) \geq \frac{1}{\sqrt{2\pi}} \exp - \frac{a^2}{2}. \quad (1.4)$$

thereby recovering what is the more classical formulation of the isoperimetric inequality [O]. In this spirit (and using his symmetrization methods) A. Ehrhard has proved a number of interesting inequalities, that are versions for the Gauss measure of classical results [E2].

Inequality (1.4) for functions rather than sets [L] yields in particular that a function on  $\mathbb{R}^n$  whose gradient belongs to  $L^1(\gamma_n)$ , belongs to the Orlicz space  $L^1(\log L)^{1/2}$  of this measure, connecting with logarithmic Sobolev inequalities and hypercontractivity.

The second type of use of Borell's inequality is for "large" values of  $u$  (while Borell's inequality for large values of  $u$  follows from (1.4), the spirit of application is very different). It is mostly used in the following forms

$$\text{If } \gamma_n(A) \geq 1/2, \text{ then } \gamma_n(A_u) \geq \gamma_1((-\infty, u]) \quad (1.5)$$

$$\text{If } \gamma(A) \geq 1/2, \text{ then } \gamma_*(A + uB) \geq \gamma_1((-\infty, u]) \geq 1 - \frac{1}{2} \exp(-u^2/2). \quad (1.6)$$

In the terminology of V. Milman [M2] (1.5) is a “concentration of measure phenomenon”. An immediate consequence of (1.5) is that if  $f$  is a Lipschitz function on  $\mathbb{R}^n$ , we have

$$\gamma_n(\{|f - M_f| \geq u\}) \leq 2\gamma_1\left(\left[\frac{u}{\|f\|_{\text{Lip}}}, \infty\right)\right) \leq \exp\left(-\frac{u^2}{2\|f\|_{\text{Lip}}^2}\right) \quad (1.7)$$

where  $M_f$  is a median of  $f$ , i.e.  $\gamma_n(\{f \geq M_f\}) = \gamma_n(\{f \leq M_f\}) = 1/2$ , and where  $\|f\|_{\text{Lip}} = \sup_{x \neq y} |f(x) - f(y)|/\|x - y\|$ .

It has been discovered by V. Milman [M1] that (1.7) (or, equivalently, a corresponding inequality on the sphere  $S_n$ ) is at the root of the celebrated Dvoretzky's theorem. Actually, the following inequality is sufficient to prove Dvoretzky's theorem: There is a numerical constant  $K$  such that if  $f$  is a Lipschitz function on  $\mathbb{R}^n$ , then

$$\gamma_n(\{|f - \int f d\gamma_n| \geq u\}) \leq 2\exp\left(-\frac{u^2}{K\|f\|_{\text{Lip}}^2}\right). \quad (1.8)$$

A very simple proof of this inequality (1.8) was discovered by B. Maurey and G. Pisier (cf. [P]; in that same reference is included a different proof due to Maurey using stochastic integrals which yields  $K = 2$ ).

To understand better the relationship between (1.7) and (1.8) one should note that either of these inequalities imply the fact that  $|M_f - \int f d\gamma_n| \leq K\|f\|_{\text{Lip}}$ . Here, as in the sequel,  $K$  denotes a universal constant, not necessarily the same at each occurrence. It is not our purpose here to enter the topic of local theory of Banach spaces, that was covered by Milman's paper [M2], and we turn towards the application of (1.6) to probability theory. The importance of (1.6) stems from the fact that  $\gamma$  is the prototype for all Gaussian measures. To stress the point, we now outline the “canonical” way to look at Gaussian processes, that was put forward in [D] and that turned out to be of crucial importance. Given a point  $t$  in  $\ell^2$ , the series  $\sum_{k \geq 1} t_k x_k$  converges  $\gamma$  a.e. (since  $(x_k)$  is a sequence of independent r.v.) and thereby defines an element  $X_t$  of  $L^2(\gamma)$ , of law  $N(0, \|t\|_2)$ . Any subset  $T$  of  $\ell^2$  thus defines a Gaussian process  $(X_t)_{t \in T}$ . For many purposes all Gaussian processes can be reduced to this type. We say that the process is bounded if  $\sup_{t \in T} X_t < \infty$   $\gamma$  a.e. (to avoid technicalities, we assume from now on that  $T$  is countable).

A problem of historical importance was, given a Gaussian process (that is, in our setting a subset  $T$  of  $\ell^2$ ), to understand, under the conditions that  $G$  is bounded, what are the tails of  $Y = \sup_{t \in T} |X_t|$ , i.e. the behavior of the function  $\gamma(\{Y \geq u\})$  as  $u \rightarrow \infty$ . It was proved by Landau and Shepp [L-S] and, independently by Fernique [F], that  $E(e^{\alpha Y^2}) < \infty$  for some  $\alpha > 0$ , where for simplicity, we write  $Ef$  for  $\int f d\gamma$ . Interestingly, the proof of Landau and Shepp is isoperimetric in nature. In [B1], C. Borell use (1.5) as follows. Set  $\sigma = \sup_{t \in T} \|t\|_2$ . It is then clear that  $Y(x) \leq Y(y) + \sigma u$  if  $x \in y + uB$ . Thus by (1.5)

$$\gamma(\{Y \geq M_Y + \sigma u\}) \leq \gamma_1([u, \infty)).$$

This implies that

$$\alpha > \sigma \Rightarrow E \exp \frac{Y^2}{2\alpha^2} < \infty. \quad (1.9)$$

C. Borell also used the same approach to obtain sharp integrability results for homogeneous chaos [B2, B3].

It turns out that, when more information is available on  $T$  (e.g. information about entropy numbers) results sharper than (1.10) can be obtained by specific methods. This has unfortunately lead some researchers to doubt the power of (1.6); the issue is that the usefulness of (1.6) is greatly enhanced by an appropriate use of  $A$ . This point was brought in particular to light in [T1], where the following is proved. Given a bounded process  $T \subset \ell^2$ , set

$$\tau = \inf\{u > 0; \gamma(\{\sup_{t \in T} |X_t| < u\}) > 0\}.$$

Then

$$\tau' > \tau \Rightarrow E \exp \frac{1}{2\sigma^2} (Y - \tau')^2 < \infty. \quad (1.10)$$

This result should be compared to (1.9). It can be interpreted as a tail estimate. It means that the function  $f(u) = \Phi^{-1}(\gamma(\{y \leq u\}))$  (that is concave by (1.3)) satisfies

$$0 \geq \lim_{u \rightarrow \infty} \left( f(u) - \frac{u}{\sigma} \right) \geq -\frac{\tau}{\sigma}. \quad (1.11)$$

Thus,  $f(u)$  has an asymptote  $(u/\sigma) + \ell$  with  $-\tau/\sigma \leq \ell \leq 0$ . This result is optimal in the sense that  $f$  can approach this asymptote arbitrarily slowly. We refer to [L-T2], Chapter 3, for an extension of this result to homogeneous chaos, and to [G-K] for further developments of the same idea.

While (1.10) is optimal for general processes, it can be improved when one has more information about  $T$ . In [T3] a method was introduced relying on (1.6) to improve the tail estimate (1.10) in the specific case where  $T$  is compact and there is a unique  $t \in T$  with  $\|t\| = \sigma$ . The method has been developed further in [D-M-W]. It could also be used in many other situations, e.g. to simplify the results of [B-K].

While (1.10) uses in a rather precise form the information provided by (1.6), it is often sufficient (e.g. for the proof of Dvoretzky's theorem) to have a weaker information of the type

$$\gamma(A) \geq 1/2 \Rightarrow \gamma(A + uB) \geq 1 - K \exp \left( -\frac{u^2}{K} \right) \quad (1.12)$$

without precise information on the constant  $K$ . It is this principle, rather than (1.5) that we now on call the concentration of measure phenomenon (for the Gauss measure).

## 2. The Concentration of Measure Phenomenon

It seems rather unlikely that (1.6) could at all be improved, but it could come as a surprise that on the other hand (1.12) can be improved, in the sense that a similar inequality holds when the set  $A + uB$  is replaced by a smaller (and, in some cases, much smaller) set. The central result of this section is that, in the class of product measures, the natural setting for the concentration of measure phenomenon is not the Gaussian measure  $\gamma$  but rather the product measure  $\nu$  on  $\mathbb{R}^N$  obtained by providing each factor with the measure  $\nu_1$  of density  $\frac{1}{2}e^{-|x|}$  with respect to Lebesgue measure. We set

$$B_1 = \left\{ x \in \mathbb{R}^N; \sum |x_k| \leq 1 \right\}; B_2 = \left\{ x \in \mathbb{R}^N; \sum x_k^2 \leq 1 \right\}.$$

**Theorem 2.1** [T6]. *There exists a universal constant  $K$  such that for all subsets  $A$  of  $\mathbb{R}^N$ , all  $u \geq 0$ , we have*

$$\nu(A) = \nu_1((-\infty, a]) \Rightarrow \nu_*(A + \sqrt{u}B_2 + uB_1) \geq \nu_1\left(\left(-\infty, a + \frac{u}{K}\right]\right). \quad (2.1)$$

*In particular*

$$\nu(A) \geq 1/2 \Rightarrow \nu_*(A + \sqrt{u}B_2 + uB_1) \geq \nu_1\left(\left(\infty, \frac{u}{K}\right]\right) = 1 - \frac{1}{2} \exp\left(-\frac{u}{K}\right). \quad (2.2)$$

A striking difference between this inequality and (1.6) is that the set  $A$  is enlarged by the mixture  $\sqrt{u}B_2 + uB_1$  of the  $\ell^2$  and  $\ell^1$  balls, whose shape changes with the value of  $u$ . To understand the reason for the strange set  $\sqrt{u}B_2 + uB_1$ , it is instructive to derive from (2.1) the size of the tails  $\nu(\{X_t \geq u\})$ , where  $X_t(x) = \sum t_k x_k$  and  $t \in \ell^2$  (these can of course be obtained by a simple direct argument). The set  $A = \{X_t \leq 0\}$  satisfies  $\nu(A) \geq 1/2$  by symmetry. Thus by (2.1), we have  $\nu_*(A + \sqrt{u}B_2 + uB_1) \geq 1 - \frac{1}{2}e^{-u/K}$ . But obviously  $X_t \leq \sqrt{u}\|t\|_2 + u\|t\|_\infty$  on  $A + \sqrt{u}B_2 + uB_1$  (where  $\|t\|_\infty = \sup_{k \geq 1} |t_k|$ ). Thus we get

$$\nu(\{X_t \geq \sqrt{u}\|t\|_2 + u\|t\|_\infty\}) \leq \frac{1}{2}e^{-u/K}$$

which can be formulated as

$$\nu(\{X_t \geq u\}) \leq \frac{1}{2} \exp\left(-\min\left(\frac{u^2}{K\|t\|_2^2}, \frac{u}{K\|t\|_\infty}\right)\right)$$

(and gives the correct order for  $-\log \nu(\{X_t \geq u\})$ ).

Another difference between (2.1) and (1.6) is the unspecified constant  $K$  on the left side, that actually makes (2.1) closer to (1.12) than to (1.6). An interesting problem would be to find an “exact” version of (2.1). One could ask for example if there is a natural “smallest” set  $W(u)$  (whose shape would depend on  $u$  in a possibly complex way) that could be used instead of  $\sqrt{u}B_2 + uB_1$  in (2.1). The resulting inequality should give sharp estimates for  $\nu(\{X_t \geq u\})$ ; the variety of

competing estimates for this quantity [Hoe] might indicate the difficulty of the task.

The proof of Theorem 2.1 is made complicated by the fact that, in contrast with the Gauss measure or Lebesgue measure, the measure  $\nu$  has less symmetries (in particular is not invariant by rotations) and thus that this restricts the use of rearrangements. The method of proof is to consider a statement similar to (2.1) (the set  $\sqrt{u}B_2 + uB_1$  being replaced by a more amenable set  $C(u)$  of comparable size) and prove it by induction over  $n$ , when the set  $A$  is assumed to depend on  $n$  coordinates only. The key observation is that the proof of the induction step can be deduced from a two-dimensional statement. While the proof in (2.1) is not simple, it is beyond doubt that the important part of (2.1) is (2.2) for large values of  $u$  ( $u \geq K$ ). Fortunately, this is much simpler to prove. The idea is to prove, again by induction over the number of coordinates of which  $A$  depends, that, if one sets

$$h_A(x) = \inf\{u \geq 0; x \in A + C(u)\}$$

then  $E \exp(h_A(x)/K) \leq 1/P(A)$ , so that, by Chebyshev inequality,

$$\nu(A + C(u)) \geq 1 - \frac{1}{P(A)} \exp\left(-\frac{u}{K}\right),$$

that recovers (2.2) for  $u$  large enough.

We now explain why (2.2) is an improvement over (1.12). The argument that we will present will be referred to in the sequel as the “contraction argument”. The precise form we use was introduced by G. Pisier [P, Ch. 2] and played an essential role in the discovery of the correct formulation of Theorem 2.1. (A similar idea occurs earlier in [G-M], Section 2-1).

Consider the increasing map  $\psi$  from  $\mathbb{R}$  to  $\mathbb{R}$  that transforms  $\nu_1$  into  $\gamma_1$ . It is a simple matter to see that

$$|\psi(x) - \psi(y)| \leq K \min(|x - y|, |x - y|^{1/2}). \quad (2.3)$$

Consider the map  $\Psi$  from  $\mathbb{R}^N$  to  $\mathbb{R}^N$ , such that  $\Psi((x_k)_{k \geq 1}) = (\psi(x_k))_{k \geq 1}$ . Thus  $\Psi$  transforms  $\nu$  into  $\gamma$ .

Consider a Borel set  $A$  of  $\mathbb{R}^N$  such that  $\gamma(A) \geq 1/2$ . Then

$$\begin{aligned} \gamma(\Psi(\Psi^{-1}(A) + \sqrt{u}B_2 + uB_1)) &= \nu(\Psi^{-1}(A) + \sqrt{u}B_2 + uB_1) \\ &\geq 1 - \frac{1}{2} \exp\left(-\frac{u}{K}\right). \end{aligned} \quad (2.4)$$

However, it follows from (2.3) that

$$A_u = \Psi(\Psi^{-1}(A) + \sqrt{u}B_2 + uB_1) \subset A + K\sqrt{u}B_2 \quad (2.5)$$

and thus (2.4) improves over (1.12). To illustrate the improvement of (2.4) over (1.12), consider the case where  $A = \{x; \forall k \leq n, |x_k| \leq a_n\}$ , where  $a_n$  is chosen so that  $\gamma(A) = 1/2$  (and hence is of order  $\sqrt{\log n}$ ). Then, for  $u \ll \log n$ , the set  $A_u$  is easily seen to be contained in

$$A + K \left( \sqrt{\frac{u}{\log n}} B_2 + \frac{u B_1}{\sqrt{\log n}} \right) \leq A + \left( \frac{Ku}{\log n} \right)^{1/2} \sqrt{u} B_2,$$

where  $u/\log n \ll 1$ .

One intriguing aspect of Theorem 2.1, when used as an improvement over (1.12), is that it breaks the rotational invariance of the Gauss measure. Indeed, it not only tells us that  $\gamma_n(A_u) \geq 1 - \exp(-u/K)$  (where  $A_u$  is defined in (2.5)) but also that  $\gamma_n((RA)_u) \geq 1 - \exp(-u/K)$  for any rotation  $R$  of  $\mathbb{R}^n$ .

A natural question is whether (2.2) is the correct formulation of the concentration of measure phenomenon. This seems to be the case, at least in the setting of product measures. Indeed, consider a probability measure  $\theta_1$ , on  $\mathbb{R}$ , and its product  $\theta$  on  $\mathbb{R}^N$ . Suppose that the following holds (that is much weaker than (1.2)). There exists  $K > 0$ , such that

$$\theta(A) \geq 1/2 \Rightarrow \theta(A + KB_\infty) \geq 3/4$$

where  $B_\infty = \{x \in \mathbb{R}^N, \forall k \geq 1, |x_k| \leq 1\}$ . Then the tails  $f(u) = \theta(\{|x| \geq u\})$  must decay exponentially [T6]. Note that, if these tails decay exponentially in a smooth enough way,  $\theta_1$  is the image of  $\nu_1$  by a contraction, and Pisier's contraction argument presented before shows that (2.1) will also hold for  $\theta$ .

Consider now  $1 \leq \alpha < \infty$  and the measure  $\nu^\alpha$  on  $\mathbb{R}^N$ , obtained as the product measure when each factor is endowed with the probability measure  $a_\alpha e^{-|x|^\alpha} dx$  (where  $a_\alpha$  is a normalizing constant). The contraction argument presented above shows that

$$\nu^\alpha(A) \geq 1/2 \Rightarrow \nu_*^\alpha(A + U_\alpha(u)) \geq 1 - \exp\left(-\frac{u}{K(\alpha)}\right). \quad (2.6)_\alpha$$

where  $U_\alpha(u) = u^{1/2} B_2 + u^{1/\alpha} B_\alpha$  for  $\alpha \leq 2$ ,  $U_\alpha(u) = u^{1/2} B_2 \cap u^{1/\alpha} B_\alpha$  for  $\alpha \geq 2$ , and  $B_\alpha = \{x \in \mathbb{R}^N; \sum |x_k|^\alpha \leq 1\}$ . For  $\alpha > \beta$ ,  $(2.6)_\alpha$  is a consequence of  $(2.6)_\beta$  (by the contraction argument).

As in the Gaussian case, to each point  $t \in \ell^2$  one can associate the random variable  $X_t = \sum_{k \geq 1} t_k x_k$  on  $(\mathbb{R}^N, \nu^\alpha)$ ; and each subset  $T$  of  $\ell^2$  thus defines a stochastic process. The main motivation for proving  $(2.6)_\alpha$  was the discovery [T7, T8] of a new approach to the problem of finding lower bounds for  $E \sup_{t \in T} X_t$  that makes  $(2.6)_\alpha$  an essential step. This new approach eliminates the use of Slepian's lemma [S], which is a specific property of Gaussian processes. It replaces it by the use of  $(2.6)_\alpha$ , combined with a Sudakov-type minoration [Su]. It enables to describe  $E \sup_{t \in T} X_t$  in terms of the geometry of  $T$ , thereby extending the results of [T2] for the Gaussian case  $\alpha = 2$ . But due to limitations of space we cannot discuss this point further.

### 3. Concentration of Measure for Bernoulli Random Variables

Pisier used his contraction argument mentioned above to conclude from (1.2) that if  $\lambda_n$  denotes the product measure on  $\mathbb{R}^n$  when  $\mathbb{R}$  is equipped with the uniform measure  $\lambda_1$  on  $[-1, 1]$ , then

$$\lambda_n(A) = \Phi(a) \Rightarrow \lambda_n(A_u) \geq \Phi\left(a + \frac{u}{K}\right). \quad (3.1)$$

Closely related to  $\lambda_n$ , but of somewhat greater importance in Probability (since it corresponds to random signs) is the probability  $\mu_n$  on  $\{-1, 1\}^n$  that gives mass  $2^{-n}$  to each point. The problem arises whether a concentration of measure principle as strong as (3.1) holds for  $\mu_n$ . This is not the case (as follows from the example given after (3.3)). The appropriate formulation for a substitute to (3.1) requires to think to  $\{-1, 1\}^n$  as a subset of  $\mathbb{R}^n$ . For a non-empty subset  $A$  of  $\{-1, 1\}^n$ , we set  $\varphi_A(x) = \inf\{\|x - y\|_2; y \in \text{conv } A\}$ , where  $\text{conv } A$  denotes the convex hull of  $A$  in  $\mathbb{R}^n$ .

**Theorem 3.1** [T4].  $E \exp(\varphi_A^2/8) \leq 1/\mu_n(A)$ .

Using Chebyshev inequality gives

$$\text{For } u \geq 0, \mu_n(\{\varphi_A \geq u\}) \leq \frac{1}{\mu_n(A)} e^{-u^2/8}. \quad (3.2)$$

We first explore the consequences of this result. Consider a *convex* function for  $\mathbb{R}^n$ . Then one can deduce from (3.2) that if  $M_f$  is median of  $f$  (for  $\mu_n$ ), we have

$$\mu_n(\{|f - M_f| > u\}) \leq 4 \exp\left(-\frac{u^2}{8\|f\|_{\text{Lip}}^2}\right). \quad (3.3)$$

This inequality should be compared to (1.7). A major difference with (1.7) is however that this result is really specific to *convex* functions. To see it, consider  $n$  even, and let  $A = \{x \in \{-1, 1\}^n; \sum_{i \leq n} x_i \leq 0\}$ , so that  $\mu_n(A) \geq 1/2$ . Define  $f(x) = \inf\{\|x - y\|_2; y \in A\}$ , so that  $\|f\|_{\text{Lip}} = 1$ , and  $M_f = 0$ . It is easy to see that for  $y \in \{-1, 1\}^n$ , we have  $f(y) = \sqrt{2}((\sum_{i \leq n} y_i)^+)^{1/2}$ . But the central limit theorem shows that  $\mu_n(\{f \geq cn^{1/4}\}) \geq 1/4$  for some constant  $c$  independent of  $n$ . (Note then that  $\mu_n(A_{cn^{1/4}}) \leq 3/4$  and that (3.1) fails.)

Consider now a Banach space  $E$  and vectors  $(x_k)_{k \leq n}$  in  $E$ .

Set

$$\sigma = \sup \left\{ \sum_{k \leq n} x^*(x_k)^2; x^* \in E^*, \|x^*\| \leq 1 \right\}.$$

The function on  $\mathbb{R}^n$  given by  $f(y) = \|\sum_{k \leq n} y_k x_k\|_E$  is convex and satisfies  $\|f\|_{\text{Lip}} = \sigma$ . Consider a sequence  $(\varepsilon_k)_{k \leq n}$  of (symmetric) Bernoulli random variables; that is, the sequence is independent identically distributed and  $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$ . Then (3.3) implies



$$P\left(\left|\left\|\sum_{k \leq n} \varepsilon_k x_k\right\| - M\right| \geq u\right) \leq 4e^{-u^2/8\sigma^2} \quad (3.4)$$

where  $M$  is a median of  $\left\|\sum_{k \leq n} \varepsilon_k x_k\right\|$ . From (3.4) and elementary computations follows that

$$\left\|\sum_{k \leq n} \varepsilon_k x_k\right\|_p \leq \left\|\sum_{k \leq n} \varepsilon_k x_k\right\|_1 + K\sigma p^{1/2},$$

a precise form of the so called Khintchine-Kahane inequalities.

It is not known whether the exponent  $1/8$  in (3.2) can be improved; the best possible exponent would be  $1/2$ . Another problem of interest would be the determination of  $\min\{\mu_n(\{\varphi_A \geq u\}); \mu_n(A) = u\}$ . It is likely that the sets which achieve this minimum depend on  $A, u$ ; thus the problem might be difficult.

It is of interest to compare Theorem 3.1 with the classical results concerning Hamming distance. The Hamming distance  $d(s, t)$  of two points  $s, t$  of a product of sets is the number of coordinates where  $s, t$  differ. For a subset  $A$  of  $\{-1, 1\}^n$ , we set  $d_A(x) = \inf\{d(x, y), y \in A\}$ . It follows from an isoperimetric inequality of Harper [Ha] that for  $\mu_n(A) \geq 1/2$ , we have

$$\mu_n(\{d_A \geq u\sqrt{n}\}) \leq \exp(-2u^2). \quad (3.5)$$

On the other hand, it is simple to see that  $2d_A \leq \sqrt{n}\varphi_A$ . Thus  $\{d_A \geq u\sqrt{n}\} \subset \{\varphi_A \geq 2u\}$ . Now (3.2) provides the estimate

$$\mu_n(\{\varphi_A \geq 2u\}) \leq 2\exp(-u^2/2).$$

Compared with (3.5), this provides a weaker bound (but of the same essential strength) for a larger set. The most important difference is however that (3.2), in contrast with (3.5), is independent of the dimension.

We now present an "abstract" extension of Theorem 3.1. Consider a sequence  $(\Omega_k, \mu_k)_{k \leq n}$  of probability spaces and denote by  $P$  the product measure on  $\Omega = \prod_{k \leq n} \Omega_k$ . Consider a subset  $A$  of  $\Omega$ . For  $x \in \Omega$ , consider the set

$$U_A(x) = \{t \in \{0, 1\}^n; \exists y \in A, t_k = 0 \Rightarrow x_k = y_k\}.$$

We consider  $U_A(x)$  as a subset of  $\mathbb{R}^n$ ; we denote by  $V_A(x)$  the convex hull of  $U_A(x)$ .

For  $\alpha \geq 1$ ,  $0 \leq u \leq 1$ , we consider the function

$$\xi(\alpha, u) = \alpha(1-u) \log(1-u) - (\alpha+1-\alpha u) \log\left(1 - \frac{\alpha u}{1+\alpha}\right).$$

Elementary calculus show that this function increases in  $\alpha, u$ , and is convex in  $u$ .

We set

$$\psi_{\alpha, A}(x) = \inf \left\{ \sum_{i \leq n} \xi(\alpha, y_i); y = (y_i)_{i \leq n} \in V_A(x) \right\}.$$

**Theorem 3.2** [T9].  $E \exp \psi_{\alpha, A} \leq (1/P_*(A))^\alpha$ .

Calculus shows that  $\xi(1, u) \geq u^2/4$ ; thus Theorem 3.2 implies Theorem 3.1, but only with the worse exponent  $1/16$  instead of  $1/8$ . An essential improvement of Theorem 3.2 over Theorem 3.1 is that for  $\alpha$  large and  $u$  close to 1,  $\xi(\alpha, u)$  is of order  $\xi(\alpha, 1) = \log(1 + \alpha)$ . The following bound seems to be of particular interest. For  $t \geq 1$ ,

$$P_x(A) \geq 1/2 \Rightarrow P(\{\psi_{t,A} \geq t\}) \leq (2/e)^t.$$

It has been observed in [J-S] (using the method of [T4]) that if  $0 < \eta < 1$ , and if  $\mu_n$  denotes now the measure  $((1 - \eta)\delta_0 + \eta\delta_1)^n$  on  $\{0, 1\}^n$ , then for a set  $A \subset \{0, 1\}^n$ , we have  $E \exp\{\varphi_A^2/4\} \leq 1/\mu_n(A)$  (this also follows from Theorem 3.2). An interesting fact in that direction is that the tails of  $\varphi_A$  do not improve when  $\eta$  is small. This is somewhat unexpected. To see it, consider the case where  $A = \{x \in \{0, 1\}^n; \sum_{k \leq n} x_k \leq \eta n\}$ , so that  $\mu_n(A)$  is of order  $1/2$  by the law of large numbers. On the other hand, it is simple to see that (for  $\eta n$  integer)  $\varphi_A(y) \leq u$  if and only if  $\sum y_k = p$  where  $\sqrt{p}(1 - \eta n/p) \leq u$ , so that  $p \leq \eta n + u\sqrt{p}$ . For  $u \leq (\eta n)^{1/2}$ , this implies  $p \leq 2\eta n$ , so that  $p \leq \eta n + u\sqrt{2\eta n}$ . Thus for  $p > \eta n + u\sqrt{2\eta n}$  we have  $\varphi_A(y) > u$ . It follows from the central limit theorem that if  $0 < \eta \leq 1/2$ , then for  $n$  large enough, we have  $\mu_n(\{\varphi_A(y) > u\}) \geq \exp(-cu^2)$  for some  $c$  independent of  $n, \eta$ .

#### 4. An Isoperimetric Inequality for Product Measure

An important concentration of measure phenomenon for product measures is as follows. Consider a sequence  $(\Omega_k, \mu_k)_{k \leq n}$  of probability spaces. Denote by  $P$  the product measure on  $\Omega = \prod_{k \leq n} \Omega_k$ . Then

$$P(A) \geq 1/2 \Rightarrow P(\{d_A \geq u\}) \leq 2 \exp(-u^2/Kn). \quad (4.1)$$

where the Hamming distance  $d_A$  has been introduced in Section 3. This is an extension of (3.5) (with worse constants). It is easy to prove using the martingale approach introduced by B. Maurey and developed by G. Schechtman (see [M-S]). It also follows from Theorem 3.2 the way (3.5) follows from Theorem 3.1. (This approach gives a constant  $K = 4$  in the exponent.)

For a set  $A \subset \Omega$ , and  $k, q \geq 0$ , consider

$$H(A, q, k) = \left\{ y \in \prod_{k \leq n} \Omega_k; \exists x^1, \dots, x^q \in A; \text{card} \{i; \forall \ell \leq q, y_i \neq x_i^\ell\} \leq k \right\}.$$

For  $q = 1$ , this is exactly the set  $\{d_A \leq k\}$ . The set  $H(A, q, k)$  can be thought of as an “enlargement” of  $A$ , although it does not seem possible to define it as a neighborhood of  $A$  for a distance.

**Theorem 4.1** [T5]. *For some universal constant  $K$ , and all  $k, q \geq 1$ , we have*

$$P(A) \geq 1/2 \Rightarrow P_\bullet(H(A, q, k)) \geq 1 - \left( \frac{K}{k} + \frac{K}{q \log q} \right)^k. \quad (4.2)$$

As stated, this theorem gives information only when  $k, q$  are large. However it is also possible to show that if  $q \geq 2$ ,  $k \geq k_0$ , then

$$P(A) \geq 1/2 \Rightarrow P_*(H(A, q, k)) \geq 1 - \eta^k$$

where  $\eta < 1$  is universal. In contrast with the case  $q = 1$  (4.1), the estimate (4.2) is independent of the number of coordinates (and thus can be extended to the case of an infinity of factors.)

To gain some intuition about (4.2), it is useful to consider the case where  $\Omega_i = \{0, 1\}$ ,  $\mu_i(\{0\}) = 1 - 1/n$ ,  $\mu_i(\{1\}) = 1/n$ , and

$$A = \left\{ (x_i); \sum_{i \leq n} x_i \leq 1 \right\}.$$

In that case,  $P(A) \geq 1/2$  and

$$H(A, q, k) = \left\{ (x_i); \sum_{i \leq n} x_i \leq q + k \right\}.$$

For  $k$  of order  $q \log q$ , simple estimates show that

$$P(H(A, q, k)) \leq 1 - \left( \frac{1}{Kq \log q} \right)^k,$$

which should be compared to (4.2).

Theorem 4.1 has strong implications about the behavior of sums of independent random variables valued in a Banach space. Consider such variables  $X_1, \dots, X_n$  valued into a separable Banach space  $F$ . We now outline a method to obtain bounds on the tails of  $\|\sum_{i \leq n} X_i\|$ . (These bounds can now also be derived from Theorem 3.2, which has a considerably simpler proof than Theorem 4.1. Tail estimates are in particular at the root of classical theorems like strong laws of large numbers and laws of the iterated logarithm.) While this method might look complicated at first glance, it seems to capture the size of these tails in essentially all the situations; see e.g. [T5, L-T1]. Without essential loss of generality, one can assume that the variables are symmetric, i.e.  $X_i$  has the same distribution as  $-X_i$ . Consider a sequence  $(\varepsilon_i)_{i \leq n}$  of Bernoulli random variables, that can be assumed to be independent of  $(X_i)_{i \leq n}$ . Thus  $\sum_{i \leq n} \varepsilon_i X_i$  has the same distribution as  $\sum_{i \leq n} X_i$ . We then write

$$\begin{aligned} \left\| \sum_{i \leq n} \varepsilon_i X_i \right\| &= E_\varepsilon \left\| \sum_{i \leq n} \varepsilon_i X_i \right\| + \left( \left\| \sum_{i \leq n} \varepsilon_i X_i \right\| - E_\varepsilon \left\| \sum_{i \leq n} \varepsilon_i X_i \right\| \right) \\ &:= \text{(I)} + \text{(II)} \end{aligned} \quad (4.3)$$

where  $E_\varepsilon$  is the conditional expectation given  $(X_i)_{i \leq n}$ . Denote by  $\mu_i$  the law of  $X_i$  on  $F$ ; Consider a set  $A \subset F^n$ , and suppose that

$$A \subset \left\{ (x_1, \dots, x_n) \in F^n, E_\varepsilon \left\| \sum_{i \leq n} \varepsilon_i x_i \right\| \leq M \right\}. \quad (4.4)$$

Then it is easy to see that

$$(x_1, \dots, x_n) \in H(A, q, k) \Rightarrow \left\| \sum_{i \leq n} \varepsilon_i x_i \right\| \leq qM + \sum_{i \leq k} \|x_i\|^*$$

where  $\|x_i\|^*$  is the  $i$ -th largest term of the sequence  $(\|x_i\|)_{i \leq n}$ . Suppose now that  $P((X_1, \dots, X_n) \in A) \geq 1/2$  (e.g. if  $M = 2E\|\sum_{i \leq n} \varepsilon_i X_i\|$  and there is equality in (4.4)). It then follows from (4.2) that, if  $k \geq q$

$$P((I) \geq qM + \sum_{i \leq k} \|X_i\|^*) \leq \left(\frac{K}{q}\right)^k. \quad (4.5)$$

On the other hand, if we set

$$\sigma_X^2 = \sup \left\{ \sum_{i \leq n} x^*(X_i)^2; x^* \in E^*, \|x^*\| \leq 1 \right\},$$

it follows from (3.4) that, conditionally on  $X_1, \dots, X_n$

$$P((II) \geq K(1+u)\sigma_X) \leq 4e^{-u^2}. \quad (4.6)$$

To make (4.5), (4.6) usable, it remains to control  $\sum_{i \leq n} \|X_i\|^*$  (which is a problem about real-valued random variables) and  $\sigma_X$ . Several methods have been developed for that purpose; adjusting the various parameters involved has allowed to get bounds of the right order in all the problems studied to date; cf. [L-T2], Chapters 6 to 8.

## References

- [A] R. J. Adler: An introduction to continuity, extrema, and related topics for general Gaussian processes. Forthcoming book
- [B-K] S. Berman, N. Kôno: The maximum of a Gaussian process with non constant variance; a sharp bound for the distribution tail. *Ann. Probab.* **17**, 632–650 (1989)
- [B1] C. Borell: The Brunn-Minkowski inequality in Gauss space. *Invent. math.* **30**, 207–216 (1975)
- [B2] C. Borell: Tail probabilities in Gauss space. *Vector Spaces Measures and Application*, Dublin 1977. (Lecture Notes in Mathematics, vol. 644.) Springer, Berlin Heidelberg New York 1978, pp. 71–82
- [B3] C. Borell: On polynomial chaos and integrability. *Prob. Math. Statist.* **3**, 191–203 (1984)
- [D] R. M. Dudley: The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1**, 290–330 (1987)
- [D-M-W] V. Dobric, M. B. Marcus, M. Weber: The distribution of the large values of the supremum of a Gaussian process
- [E1] A. Ehrhard: Symétrisation dans l'espace de Gauss. *Math. Scand.* **53**, 281–301 (1983)
- [E2] A. Ehrhard: Inégalités isoperimétriques et intégrales de Dirichlet Gaussiennes. *Ann. Sci. Ec. Norm. Sup.* **17**, 317–332 (1984)

- [E3] A. Ehrhard: *Eléments extrémaux pour les inégalités de Brunn-Minkowski gaussiennes*. Ann. Inst. H. Poincaré **22**, 149–168 (1986)
- [F] X. Fernique: *Intégrabilité des vecteurs gaussiens*. C. R. Acad. Sci. Paris **270**, 1698–1699 (1970)
- [F-L-M] T. Figiel, J. Lindenstrauss, V. Milman: *The dimensions of almost spherical sections of convex bodies*. Acta Math. **139**, 52–94 (1977)
- [G-K] V. Goodman, J. Kuelbs: *Rates of clustering of some self similar Gaussian processes*. Manuscript, 1990
- [G-M] M. Gromov, V. D. Milman: *A topological application of the isoperimetric inequality*. Amer. J. Math. **105**, 843–854 (1983)
- [Ha] L. H. Harper: *Optimal numbering and isoperimetric problems on graphs*. J. Comb. Theory **1**, 385–393 (1966)
- [Hoe] W. Hoeffding: *Probability inequalities for sums of bounded random variables*. J. Amer. Statist. Assoc. **58**, 13–30 (1963)
- [Hof] J. Hoffmann-Jørgensen: *Sums of independent Banach space valued random variables*. Studia Math. **52**, 159–186 (1974)
- [J-S] W. B. Johnson, G. Schechtman: *Remarks on Talagrand's deviation inequality for Rademacher functions*. Texas Functional Analysis Seminar 1988–89. The University of Texas, 1989
- [K] J. P. Kahane: *Some random series of functions*. Cambridge University Press, 1985
- [K-S] S. Kwapien, J. Szulga: *Hypercontraction methods for comparison of moments of random series in normed spaces*. Ann. Probab. (to appear)
- [L-S] H. Landau, L. A. Shepp: *On the supremum of a Gaussian process*. Sankhyā **32**, 369–378 (1970)
- [L] M. Ledoux: *Isoperimétrie et inégalités de Sobolev logarithmiques gaussiennes*. C. R. Acad. Sci. Paris **306**, 79–82 (1988)
- [Lé] P. Lévy: *Problèmes concrets d'analyse fonctionnelle*. Gauthier Villars, 1951
- [L-T1] M. Ledoux, M. Talagrand: *Some application of isoperimetric methods to strong limit theorems for sums of independent random variables*, Ann. Probab. **18**, 754–789 (1990)
- [L-T2] M. Ledoux, M. Talagrand: *Probability in Banach spaces*. (Ergebnisse der Mathematik, Bd. 23). Springer, Berlin Heidelberg New York 1991
- [McK] H. P. McKean: *Geometry of differential space*. Ann. Probab. **1**, 197–206 (1973)
- [M1] V. D. Milman: *New proof of the theorem of Dvoretzky on sections of convex bodies*. Funct. Anal. Appl. **5**, 28–37 (1971)
- [M2] V. D. Milman: *The Concentration Phenomenon and Linear Structure of Finite Dimensional Normed Spaces*. Proceedings of the International Congress of Mathematicians, Berkeley 1986, pp. 961–975
- [M3] V. D. Milman: *The heritage of P. Lévy in functional analysis*. Asterisque **157–158**, 273–302 (1988)
- [M-S] V. D. Milman, G. Schechtman: *Asymptotic theory of finite dimensional normed spaces*. (Lecture Notes in Mathematics, vol. 1200.) Springer, Berlin Heidelberg New York 1986
- [O] R. Osserman: *The isoperimetric inequality*. Bull. Amer. Math. Soc. **84**, 1182–1238 (1978)
- [P] G. Pisier: *Probabilistic methods in the geometry of Banach spaces*. Probability and Analysis, Varenna (Italy) 1985. (Lecture Notes in Mathematics, vol. 1206.) Springer, Berlin Heidelberg New York 1986, pp. 167–241
- [S] D. Slepian: *The one sided barrier problem for Gaussian noise*. Bell System Tech. J. **41**, 463–501 (1962)

- [Su] V. N. Sudakov: Gaussian measures, Cauchy measures and  $\varepsilon$ -entropy. *Sov. Math. Dokl.* **10**, 310–315 (1969)
- [S-T] V. N. Sudakov, B. S. Tsirelson: Extremal properties of half spaces for spherically invariant measures. *J. Sov. Math.* **9**, 9–18 (1978) [translated from *Zap. Nauch. Sem. L.O.M.I.* **41**, 14–24 (1974)]
- [T1] M. Talagrand: Sur l'intégrabilité des vecteurs gaussiens. *Z. Wahrsch. Verw. Gebiete* **68**, 1–8 (1984)
- [T2] M. Talagrand: Regularity of Gaussian processes. *Acta Math.* **159**, 99–149 (1987)
- [T3] M. Talagrand: Small tails for the supremum of a Gaussian process. *Ann. Inst. H. Poincaré* **24**, 307–315 (1988)
- [T4] M. Talagrand: An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities. *Proc. Amer. Math. Soc.* **104**, 905–909 (1988)
- [T5] M. Talagrand: Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Ann. Probab.* **17**, 1546–1570 (1989)
- [T6] M. Talagrand: A new isoperimetric inequality and the concentration of measure phenomenon, to appear in *Geometric aspects of functional analysis* (Israel seminar). (Lecture Notes in Mathematics, vol. 1469.) Springer, Berlin Heidelberg New York 1991
- [T7] M. Talagrand: Supremum of canonical processes. Manuscript, 1990
- [T8] M. Talagrand: Regularity of infinitely divisible processes. Manuscript, 1990
- [T9] M. Talagrand: A new isoperimetric inequality for product measure and the tails of sums of independent random variables. Manuscript, 1990