ON TALAGRAND'S DEVIATION INEQUALITIES FOR PRODUCT MEASURES

MICHEL LEDOUX

Abstract. We present a new and simple approach to some of the deviation inequalities for product measures deeply investigated by M. Talagrand in the recent years. Our method is based on functional inequalities of Poincaré and logarithmic Sobolev type and iteration of these inequalities. In particular, we establish with these tools sharp deviation inequalities from the mean on norms of sums of independent random vectors and empirical processes. Concentration for the Hamming distance may also be deduced from this approach.

1. Introduction. Deviation inequalities for convex functions

It is by now classical that if f is a Lipschitz function on $\mathbb{R}^n$ with Lipschitz constant $\|f\|_{\mathrm{Lip}} \le 1$, and if γ_n denotes the canonical Gaussian measure on $\mathbb{R}^n$, for every $t \ge 0$,

$$\gamma_n(f \ge M + t) \le \mathrm{e}^{-t^2/2} \tag{1.1}$$

where M is either the mean or the median of f with respect to γ_n (see Ledoux and Talagrand (1991), Ledoux (1994)). This inequality is part of the so-called concentration of measure phenomenon of isoperimetric type.

In the past years, M. Talagrand developed striking new methods in the investigation of this phenomenon in the case of product measures. These ideas led to definitive progress in an number of various areas such as Probability in Banach spaces, Empirical Processes, Geometric Probability, Statistical Mechanics... The interested reader will find in the important contribution Talagrand (1995a) a complete account on these methods and results (see also Talagrand (1994b)). One of the first results at the starting point of these developments is the following simple inequality for arbitrary product measures Talagrand (1988), Johnson and Schechtman (1987) (see also Maurey (1991)). Let f be a convex Lipschitz function on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \le 1$. Let μ_i, $i = 1, \ldots, n$, be probability measures on $[0, 1]$ and denote by P the product probability measure $\mu_1 \otimes \cdots \otimes \mu_n$. Then, for every $t \ge 0$,

$$P(f \ge M + t) \le 2\,\mathrm{e}^{-t^2/4} \tag{1.2}$$

ESAIM: Probability and Statistics is an electronic journal with URL address **http://www.emath.fr/ps/**

Received by the journal November 8, 1995. Revised April 18, 1996. Accepted for publication May 31, 1996.

where M is a median of f for P. Contrary to the Gaussian case, it is known that the convexity assumption on f is essential (cf. Ledoux and Talagrand (1991), p. 25). The proof of (1.2) is based on the inequality

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

that is established by geometric arguments and a basic induction on the number of coordinates. It is now embedded in some further abstract framework called by M. Talagrand convex hull approximation (cf. Talagrand (1995a), (1994b)). M. Talagrand also introduced a concept of approximation by a finite number of points Talagrand (1989), (1995a), (1994b). These powerful abstract tools have been used in particular to study sharp deviations inequalities for large classes of functions (Talagrand (1994a), (1995a), (1995b)).

The aim of this work is to provide a simple proof of inequality (1.2), as well as of deviations inequalities for classes of functions, based on functional inequalities. Following the basic induction principle, we will work with the only functional inequalities which we know to easily tensorise, namely Poincaré and logarithmic Sobolev inequalities. The proof then reduces to estimates on convex functionals in dimension one which turns out to be trivial. Once the appropriate logarithmic Sobolev inequality holds, it may be turned into a simple differential inequality on Laplace transforms. We apply these ideas to obtain, in Section 2, precise bounds for deviation inequalities on sums of independent vector valued random variables or empirical processes of statistical interest, and motivated by questions by L. Birgé and P. Massart (cf. Talagrand (1995b)). More precisely, if X_i, $i = 1, \dots, n$, are independent random variables with values in some space S, and if $\mathcal{F}$ is a countable class of measurable functions on S, set

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

Then, if $|f| \le C$ for every f in $\mathcal{F}$, and if $\mathbb{E}f(X_i) = 0$ for every $f \in \mathcal{F}$ and $i = 1, \dots, n$, for all $t \ge 0$,

$$\mathbb{P}\big(Z \ge \mathbb{E}(Z) + t\big) \le 3\exp\bigg(-\frac{1}{K}\cdot\frac{t}{C}\log\bigg(1+\frac{Ct}{\sigma^2 + C\mathbb{E}(Z)}\bigg)\bigg)$$

where $\sigma^2 = \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} \mathbb{E}f^2(X_i)$ and $K > 0$ is a numerical constant. The new feature is an exact deviation from the mean rather than only from some multiple of it as in Talagrand (1989), (1994a), Ledoux and Talagrand (1991). This result has been obtained recently by M. Talagrand Talagrand (1995b) as a consequence of a further deepening of his abstract principles. While it is uncertain whether our approach could recover these abstract principles, the deviation inequalities themselves follow rather easily from it. On the abstract inequalities themselves, let us mention here the recent alternate approach by K. Marton (1995a), (1995b) and Dembo (1995) (see also Dembo and Zeitouni (1995)) based on information inequalities and coupling in which the concept of entropy also plays a crucial role. Let us also observe that hypercontraction methods were used in Kwapień and Szulga (1991) to study integrability of norms of sums of independent vector valued random variables. The work by K. Marton also concerns some Markov chain setting. It might be that the functional approach developed in this paper also applies in certain dependent situations. In Section 3, we briefly investigate in the same way deviation inequalities for chaos. In the last section, we emphasize, following S. Bobkov, the basic induction procedure. As is known for example, if g is a function on a product space $\Omega = \Omega_1 \times \cdots \times \Omega_n$ with product probability measure $P = \mu_1 \otimes \cdots \otimes \mu_n$, then

$$\int g^2 \log g^2 dP - \int g^2 dP \log \int g^2 dP \\ \leq \sum_{i=1}^n \int \left[\int g^2 \log g^2 d\mu_i - \int g^2 d\mu_i \log \int g^2 d\mu_i \right] dP.$$

In particular, we easily recover the basic and historical concentration for the Hamming distance (Milman and Schechtman (1986), Talagrand (1995a)) with which we conclude this work.

To introduce to our main argument, we first treat the case of Poincaré or spectral gap inequalities. As before, $\mu_1, \ldots, \mu_n$ are arbitrary probability measures on $[0,1]$ and P is the product probability $P = \mu_1 \otimes \cdots \otimes \mu_n$. We say that a function f on $\mathbb{R}^n$ is separately convex if it is convex in each coordinate.

THEOREM 1.1. *Let f be a separately convex smooth function on $\mathbb{R}^n$. Then, for any product probability P on $[0,1]^n$,*

$$\int f^2 dP - \left(\int f dP \right)^2 \leq \int |\nabla f|^2 dP$$

(where ∇f is the usual gradient of f on $\mathbb{R}^n$ and $|\nabla f|$ denotes its Euclidean length).

After this work was completed, we discovered that this statement has been obtained previously by S. Bobkov (1994) (with the same proof, and a better constant when the μ_i's are centered probability measures on a symmetric interval).

Proof. We will actually prove something more, namely that, for any product probability P on $\mathbb{R}^n$, and any separately convex smooth function f,

$$\int f^2 dP - \left(\int f dP \right)^2 \leq \sum_{i=1}^n \int\int (x_i - y_i)^2 (\partial_i f)^2(x) dP(x) dP(y). \quad (1.3)$$

When P is concentrated on $[0,1]^n$, Theorem 1.1 follows. This important inequality (1.3), and the corresponding one for entropy (1.4), is in fact the form that will be used in a vector valued setting in Section 2 and will be emphasised there as Proposition 2.1. Assume first that $n = 1$. Since $f : \mathbb{R} \to \mathbb{R}$ is convex, for any $x, y \in \mathbb{R}$,

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

Hence
$$\big|f(x)-f(y)\big| \le |x-y| \max\big(|f'(x)|, |f'(y)|\big).$$
Therefore
$$\begin{aligned}\int f^2 dP - \Big(\int f dP\Big)^2 &= \frac{1}{2}\iint \big|f(x)-f(y)\big|^2 dP(x)dP(y)\\ &\le \iint (x-y)^2 f'(x)^2 dP(x)dP(y)\end{aligned}$$
which is the result in this case.

Now, we simply need to classicaly tensorise this one-dimensional Poincaré type inequality. Suppose (1.3) holds for $P_{n-1} = \mu_1 \otimes \cdots \otimes \mu_{n-1}$ and let us prove it for $P_n = P = \mu_1 \otimes \cdots \otimes \mu_n$. Let thus $f : \mathbb{R}^n \to \mathbb{R}$ be separately convex. By Fubini's theorem and the induction hypothesis,
$$\begin{aligned}\int f^2 dP_n &= \int d\mu_n(x_n)\Big[\int f^2(z,x_n)dP_{n-1}(z)\Big]\\ &\le \int d\mu_n(x_n)\Big[\Big(\int f(z,x_n)dP_{n-1}(z)\Big)^2\\ &\quad + \sum_{i=1}^{n-1}\iint (x_i-y_i)^2(\partial_i f)^2(z,x_n)dP_{n-1}(z)dP_{n-1}(y)\Big]\end{aligned}$$
where $z = (x_1,\dots,x_{n-1}) \in \mathbb{R}^{n-1}$. Let $h(x_n) = \int f(z,x_n)dP_{n-1}(z)$. Then h is convex on $\mathbb{R}$ and, by the first step,
$$\int h^2 d\mu_n \le \Big(\int h d\mu_n\Big)^2 + \iint (x_n-y_n)^2 h'(x_n)^2 d\mu_n(x_n)d\mu_n(y_n).$$
Now $\int h d\mu_n = \int f dP_n$ and, by Jensen's inequality,
$$h'(x_n)^2 \le \int (\partial_n f)^2(z,x_n)dP_{n-1}(z).$$
Therefore,
$$\begin{aligned}&\int d\mu_n(x_n)\Big[\Big(\int f(z,x_n)dP_{n-1}(z)\Big)^2\Big]\\ &\qquad \le \Big(\int f dP_n\Big)^2 + \iint (x_n-y_n)^2(\partial_n f)^2(x)dP_n(x)dP_n(y)\end{aligned}$$
from which the conclusion follows. Theorem 1.1 is proved. □

Now, in a setting where a Poincaré type inequality is satisfied, it is known that Lipschitz functions are exponentially integrable Gromov and Milman (1983), Aida et al. (1994). This is however not quite enough to reach Gaussian estimates such as (1.2). This is why we rather have to turn to logarithmic Sobolev inequalities that however are not more difficult.

THEOREM 1.2. *Let g be a smooth function on $\mathbb{R}^n$ such that $\log g^2$ is separately convex ($g^2 > 0$). Then, for any product probability P on $[0,1]^n$,*

$$\int g^2 \log g^2 dP - \int g^2 dP \log \int g^2 dP \leq 4 \int |\nabla g|^2 dP.$$

Proof. As in the proof of Theorem 1.1, we establish that, for any product probability P on $\mathbb{R}^n$, and any smooth function g such that $\log g^2$ is separately convex,

$$\begin{aligned}\int g^2 \log g^2 dP - \int g^2 dP \log \int g^2 dP \\ \leq 4 \sum_{i=1}^{n} \iint (x_i - y_i)^2 (\partial_i g)^2(x) dP(x) dP(y).\end{aligned} \tag{1.4}$$

Start again with $n = 1$. Set $g^2 = \mathrm{e}^f$. Since f and e^f are convex, for all $x, y \in \mathbb{R}$,

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

and

$$\mathrm{e}^{f(x)} - \mathrm{e}^{f(y)} \leq (x-y) f'(x)\, \mathrm{e}^{f(x)}.$$

It follows that, for all $x, y \in \mathbb{R}$,

$$\left[\mathrm{e}^{f(x)} - \mathrm{e}^{f(y)}\right]\left[f(x) - f(y)\right] \leq (x-y)^2 \max\left(f'(x)^2 \mathrm{e}^{f(x)}, f'(y)^2 \mathrm{e}^{f(y)}\right).$$

In another words

$$\left[g^2(x) - g^2(y)\right]\left[\log g^2(x) - \log g^2(y)\right] \leq 4(x-y)^2 \max\left(g'(x)^2, g'(y)^2\right).$$

Hence

$$\begin{aligned}\iint \left[g^2(x) - g^2(y)\right]\left[\log g^2(x) - \log g^2(y)\right] dP(x) dP(y) \\ \leq 8 \iint (x-y)^2 g'(x)^2 dP(x) dP(y).\end{aligned}$$

Now, the left-hand-side of this inequality is equal to

$$2\left[\int g^2 \log g^2 dP - \int g^2 dP \int \log g^2 dP\right]$$

which, by Jensen's inequality, is larger than or equal to

$$2\left[\int g^2 \log g^2 dP - \int g^2 dP \log \int g^2 dP\right].$$

Therefore

$$\int g^2 \log g^2 dP - \int g^2 dP \log \int g^2 dP \leq 4 \iint (x-y)^2 g'(x)^2 dP(x) dP(y)$$

which is the result for $n = 1$ thus.

We tensorise this one-dimensional inequality as in Theorem 1.1. Let us briefly recall this classical argument Gross (1975) for the sake of completeness. (In Section 4 –Proposition 4.1–, we will come back to this iteration procedure in an abstract framework. For pedagogical reasons, we found it easier to present first the argument in this more concrete setting.) Let g be as in the theorem. With the notation of the proof of Theorem 1.1, and the induction hypothesis,

$$\begin{aligned}\int g^2 \log g^2 dP_n &= \int d\mu_n(x_n) \Big[\int g^2(z, x_n) \log g^2(z, x_n) dP_{n-1}(z) \Big] \\ &\le \int d\mu_n(x_n) \Big[\int g^2(z, x_n) dP_{n-1}(z) \log \int g^2(z, x_n) dP_{n-1}(z) \\ &\quad + 4 \sum_{i=1}^{n-1} \int\!\!\int (x_i - y_i)^2 (\partial_i g)^2(z, x_n) dP_{n-1}(z) dP_{n-1}(y) \Big].\end{aligned}$$

Set $h(x_n) = \big(\int g^2(z, x_n) dP_{n-1}(z)\big)^{1/2}$. It is easily seen, by Hölder's inequality, that $\log h^2$ is convex on $\mathbb{R}$. Hence, by the one-dimensional case,

$$\begin{aligned}\int h^2 \log h^2 d\mu_n &\le \int h^2 d\mu_n \log \int h^2 d\mu_n \\ &\quad + 4 \int\!\!\int (x_n - y_n)^2 h'(x_n)^2 d\mu_n(x_n) d\mu_n(y_n).\end{aligned}$$

Now, $\int h^2 d\mu_n = \int g^2 dP_n$ and, by the Cauchy-Schwarz inequality,

$$\begin{aligned}h'(x_n)^2 &= \frac{1}{h(x_n)^2} \Big(\int \partial_n g(z, x_n) g(z, x_n) dP_{n-1}(z) \Big)^2 \\ &\le \int (\partial_n g)^2(z, x_n) dP_{n-1}(z).\end{aligned}$$

The proof of Theorem 1.2 is easily completed. □

With a little more effort, the constant 4 of the logarithmic Sobolev inequality of Theorem 1.2 may be improved to (the probably optimal constant) 2. We need simply improve the estimate of the entropy in dimension one. To this end, recall the variational caracterisation of entropy (Holley and Stroock (1987)) as

$$\begin{aligned}\int g^2 \log g^2 dP &- \int g^2 dP \log \int g^2 dP \\ &= \inf_{c>0} \int \big[g^2 \log g^2 - (\log c + 1) g^2 + c \big] dP.\end{aligned} \tag{1.5}$$

Let thus P be a probability measure concentrated on $[0, 1]$. Set again $g^2 = \mathrm{e}^f$ where f is (smooth and) convex on $\mathbb{R}$. Let then $y \in [0, 1]$ be a point at which f is minimum and take $c = \mathrm{e}^{f(y)}$ (in (1.5)). For every $x \in [0, 1]$,

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

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

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

from which we deduce, together with (1.5), that

$$\int g^2 \log g^2 dP - \int g^2 dP \log \int g^2 dP \le 2 \int {g'}^2 dP.$$

Tensorising this inequality thus similarly yields that if g is smooth on $\mathbb{R}^n$ with $\log g^2$ separately convex, for every product probability P on $[0,1]^n$,

$$\int g^2 \log g^2 dP - \int g^2 dP \log \int g^2 dP \le 2 \int |\nabla g|^2 dP. \tag{1.6}$$

Now, in presence of a logarithmic Sobolev inequality, there is a general procedure that yields concentration inequalities of the type (1.2) via a simple differential inequality on Laplace transforms. This has been shown in Davies and Simon (1984), Aida et al.(1994) and Ledoux (1995) and we recall the simple steps here. In the recent note Bobkov (1995), Talagrand's inequality (1.2) on the cube is deduced in this way from Gross's logarithmic Sobolev inequality on the two point space. Note that (1.1) also follows like that from the Gaussian logarithmic Sobolev inequality Gross (1975). We use below (1.6) rather than Theorem 1.2 in order to improve some numerical constants. (With Theorem 1.2, the constants are simply weaker by a factor 2.)

Let f be a separately convex smooth Lipschitz function on $\mathbb{R}^n$ with Lipschitz norm $\|f\|_{\mathrm{Lip}} \le 1$. Let P be a product probability measure on $[0,1]^n$ which we assume first to be absolutely continuous with respect to Lebesgue's measure. (A simple smoothing procedure will then reduce to this case.) For any $\lambda \ge 0$, apply (1.6) to $g^2 = e^{\lambda f}$. Setting $F(\lambda) = \int e^{\lambda f} dP$, it yields, for any $\lambda \ge 0$,

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \le \frac{\lambda^2}{2} \int |\nabla f|^2 e^{\lambda f} dP. \tag{1.7}$$

Since $\|f\|_{\mathrm{Lip}} \le 1$, $|\nabla f| \le 1$ almost everywhere, and thus also P-almost surely since P is assumed to be absolutely continuous with respect to Lebesgue's measure. We therefore get the differential inequality

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \le \frac{\lambda^2}{2} F(\lambda), \quad \lambda \ge 0,$$

which can easily be integrated. If we let $H(\lambda) = \frac{1}{\lambda} \log F(\lambda)$, $\lambda > 0$, it reduces to $H'(\lambda) \le 1/2$. Since

$$H(0) = \lim_{\lambda \to 0} \frac{1}{\lambda} \log F(\lambda) = \frac{F'(0)}{F(0)} = \int f dP,$$

it follows that, for every $\lambda \geq 0$, $H(\lambda) \leq \int f dP + \lambda/2$. Therefore, for every $\lambda \geq 0$,

$$F(\lambda) = \int e^{\lambda f} dP \leq e^{\lambda \int f dP + \lambda^2/2}.$$

Let now P be arbitrary. Any smooth (product) convolution of P will satisfy the preceding inequality and thus, since f is continuous, P also. The smoothness assumption on f may be dropped similarly. Summarizing, we obtained that for any separately convex Lipschitz function f on $\mathbb{R}^n$ with $\|f\|_{\mathrm{Lip}} \leq 1$, and any product probability P on $[0,1]^n$, for every $\lambda \geq 0$,

$$\int e^{\lambda f} dP \leq e^{\lambda \int f dP + \lambda^2/2}.$$

By Chebyshev's inequality, for every $\lambda, t \geq 0$,

$$P\big(f \geq \int f dP + t\big) \leq e^{-\lambda t + \lambda^2/2}.$$

Optimising in λ yields the following corollary.

Corollary 1.3. *Let f be a separately convex Lipschitz function on $\mathbb{R}^n$ with Lipschitz constant $\|f\|_{\mathrm{Lip}} \leq 1$. Then, for every $t \geq 0$,*

$$P\big(f \geq \int f dP + t\big) \leq e^{-t^2/2}.$$

This inequality is the analogue of (1.2) with the mean instead of the (a) median M and the improved bound $e^{-t^2/2}$.

M. Talagrand (1988) (see also Johnson and Schechtman (1987), Maurey (1991), Talagrand (1995a), (1994b)) actually also showed deviation inequalities under the level M, that is an inequality for $-f$ (f convex). It yields a concentration result of the type

$$P\big(|f - M| \geq t\big) \leq 4\, e^{-t^2/4}, \quad t \geq 0.$$

We have not been able so far to deduce such a result with our methods. At a weak level, Theorem 1.1 indicates that for f separately convex with $\|f\|_{\mathrm{Lip}} \leq 1$,

$$P\big(\big|f - \int f dP\big| \geq t\big) \leq \frac{1}{t^2}, \quad t > 0.$$

This inequality ensures in any case that a median M (of f for P) is close to the mean $\int f dP$. Indeed, if $t > \sqrt{2}$,

$$P\big(\big|f - \int f dP\big| \leq t\big) > \frac{1}{2}$$

so that the definition of a median implies that

$$\big|M - \int f dP\big| \leq \sqrt{2}\,.$$

That concentration inequalities around the mean or the median are equivalent up to numerical constants is a well-known issue (cf. e.g. Milman and Schechtman (1986), p. 142). Although deviation inequalities above the mean or the median are the useful inequalities in Probability and its applications, concentration inequalities are sometimes important issues (e.g. in Geometry of Banach spaces (Milman, and Schechtman (1986)), percolation, spin glasses... Talagrand (1995a)).

Corollary 1.3 of course extends to probability measures μ_i supported on $[a_i, b_i]$, $i = 1, \ldots, n$, (following for example (1.4) of the proof of Theorem 1.2, or by scaling). In particular, if P is a product measure on $[a,b]^n$ and if f is separately convex on $\mathbb{R}^n$ with Lipschitz constant less than or equal to 1, for every $t \geq 0$,

$$P\big(f \geq \textstyle\int f dP + t\big) \leq \mathrm{e}^{-t^2/2(b-a)^2}. \tag{1.8}$$

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

$$\mathbb{P}\bigg(\bigg\|\sum_{i=1}^n \eta_i a_i\bigg\| \geq \mathbb{E}\bigg\|\sum_{i=1}^n \eta_i a_i\bigg\| + t\bigg) \leq \mathrm{e}^{-t^2/8\sigma^2} \tag{1.9}$$

where

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

2. Sharp bounds on norms of random vectors and empirical processes

In the second part of this paper, we turn to the case where the μ_i's are probability measures on some Banach spaces E_i. This will allow us to investigate deviation inequalities for norms of sums of independent vector valued random variables as well as empirical processes. The main idea will be to use the vector valued version of the basic inequalities (1.3) and (1.4) of Theorems 1.1 and 1.2 that we emphasize first.

For simplicity, we assume that $E_i = E$, $i = 1, \ldots, n$, where E is a real separable Banach space with norm $\|\cdot\|$. If $f : E \to \mathbb{R}$ is smooth enough, let $Df(x)$, $x \in E$, be the element of the dual space E^* of E defined as

$$\lim_{t \to 0} \frac{f(x+ty) - f(x)}{t} = \big\langle Df(x), y \big\rangle, \quad y \in E.$$

If f is convex, as in the scalar case, for every $x, y \in E$,

$$f(x) - f(y) \leq \big\langle Df(x), x - y \big\rangle.$$

Therefore, if f is a real valued separately convex smooth function on E^n, the proofs of inequalities (1.3) and (1.4) in Theorems 1.1 and 1.2 immediately extend to this vector valued to yield the following statement, of possible independent interest.

PROPOSITION 2.1. *Let f be a real valued separately convex smooth function on E^n. Then for any product probability $P = \mu_1 \otimes \cdots \otimes \mu_n$ on E^n*

$$\int f^2 dP - \left(\int f dP\right)^2 \leq \sum_{i=1}^n \iint \langle D_i f(x), x_i - y_i \rangle^2 dP(x) dP(y)$$

and

$$\int f e^f dP - \int e^f dP \log \int e^f dP \leq \sum_{i=1}^n \iint \langle D_i f(x), x_i - y_i \rangle^2 e^{f(x)} dP(x) dP(y).$$

(Here $D_i f(x)$ denotes the i-th partial derivative of f at the point x.)

These inequalities are of particular interest when f is given by $f(x) = \| \sum_{i=1}^n x_i \|$, $x = (x_1, \ldots, x_n) \in E^n$. Since we however run into various regularity questions, let us first illustrate how the preceding statement and inequalities may be expressed in a finite dimensional setting. Let us thus first assume that $E = \mathbb{R}^N$ and consider

$$f(x) = \max_{1 \leq k \leq N} \sum_{i=1}^n x_i^k \quad \text{or} \quad \max_{1 \leq k \leq N} \left| \sum_{i=1}^n x_i^k \right|.$$

As is easily seen, the convexity properties of this functional f still ensure that, for every $i = 1, \ldots, n$ and $x, y \in E^n$ with $x_j = y_j$, $j \neq i$,

$$f(x) - f(y) \leq \sum_{k=1}^N \alpha_k |x_i^k - y_i^k| \tag{2.1}$$

where $\alpha_k = \alpha_k(x) = I_{A_k}(x)$ and $(A_k)_{1 \leq k \leq N}$ is a partition of E^n with

$$A_k \subset \left\{ z \in E^n ; f(z) = \sum_{i=1}^n z_i^k \quad \text{or} \quad \left| \sum_{i=1}^n z_i^k \right| \right\}.$$

In particular, $\alpha_k \geq 0$ and $\sum_{k=1}^N \alpha_k = 1$. Using (2.1), the inequalities of Proposition 2.1 hold similarly for this functional f and we thus get respectively (under proper integrability conditions)

$$\int f^2 dP - \left(\int f dP\right)^2 \leq \iint \max_{1 \leq k \leq N} \sum_{i=1}^n (x_i^k - y_i^k)^2 dP(x) dP(y) \tag{2.2}$$

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

$$\begin{aligned} \int \lambda f e^{\lambda f} dP - \int e^{\lambda f} dP \log \int e^{\lambda f} dP \\ \leq \iint \max_{1 \leq k \leq N} \sum_{i=1}^n (x_i^k - y_i^k)^2 e^{\lambda f(x)} dP(x) dP(y). \end{aligned} \tag{2.3}$$

As announced, by finite dimensional approximation and monotone convergence, the preceding inequalities extend to

$$f(x) = \Big\| \sum_{i=1}^n x_i \Big\| = \sup_{\xi \in E^*, \|\xi\| \le 1} \sum_{i=1}^n \langle \xi, x_i \rangle, \quad x \in E^n,$$

on an arbitrary separable Banach space $(E, \|\cdot\|)$ with dual space E^*. To state the corresponding inequalities, let us use probabilistic notation and consider independent random variables X_i, $i = 1, \ldots, n$, on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in E (with law μ_i respectively). Write $S_n = \sum_{i=1}^n X_i$ and let also Y_i, $i = 1, \ldots, n$, be an independent copy of the sequence X_i, $i = 1, \ldots, n$. Assume that $\mathbb{E}\|X_i\|^2 < \infty$ for every $i = 1, \ldots, n$. With this notation, (2.2) implies that

$$\begin{aligned} \mathbb{E}\|S_n\|^2 - \big(\mathbb{E}\|S_n\|\big)^2 &\le \mathbb{E}\bigg(\sup_{\|\xi\| \le 1} \sum_{i=1}^n \langle \xi, X_i - Y_i \rangle^2 \bigg) \\ &\le 4\, \mathbb{E}\bigg(\sup_{\|\xi\| \le 1} \sum_{i=1}^n \langle \xi, X_i \rangle^2 \bigg). \end{aligned} \tag{2.4}$$

(Recall the classical Yurinskii bound based on the martingale method – cf. Ledoux and Talagrand (1991) – only yields

$$\mathbb{E}\|S_n\|^2 - \big(\mathbb{E}\|S_n\|\big)^2 \le \sum_{i=1}^n \mathbb{E}\|X_i\|^2.)$$

Note that if the X_i's are centered, (2.4) may be slightly improved to

$$\mathbb{E}\|S_n\|^2 - \big(\mathbb{E}\|S_n\|\big)^2 \le \sup_{\|\xi\| \le 1} \sum_{i=1}^n \mathbb{E}\langle \xi, X_i \rangle^2 + \mathbb{E}\bigg(\sup_{\|\xi\| \le 1} \sum_{i=1}^n \langle \xi, X_i \rangle^2 \bigg).$$

Similarly with (2.3), denote by $F(\lambda) = \mathbb{E}(\mathrm{e}^{\lambda \|S_n\|})$, $\lambda \ge 0$, the Laplace transform of $\|S_n\|$, assumed to be finite. Then, for every $\lambda \ge 0$,

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \le \lambda^2 \mathbb{E}\bigg(\sup_{\|\xi\| \le 1} \sum_{i=1}^n \langle \xi, X_i - Y_i \rangle^2 \mathrm{e}^{\lambda \|S_n\|} \bigg).$$

Denote by Σ^2 the random variable $\sup_{\|\xi\| \le 1} \sum_{i=1}^n \langle \xi, X_i \rangle^2$. Then, for $\lambda \ge 0$,

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \le 2\mathbb{E}(\Sigma^2) \lambda^2 F(\lambda) + 2\lambda^2 \mathbb{E}\big(\Sigma^2 \mathrm{e}^{\lambda \|S_n\|}\big). \tag{2.5}$$

With the same approximation procedure, inequalities (2.4) and (2.5) also hold for more general sums (empirical processes)

$$Z = \sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i) \quad \text{or} \quad \sup_{f \in \mathcal{F}} \Big| \sum_{i=1}^n f(X_i) \Big| \tag{2.6}$$

where the X_i's are independent random variables with values in some space S and $\mathcal{F}$ is a countable class of (bounded) measurable functions on S (start again with a finite class $\mathcal{F}$). In this case, $\Sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^n f^2(X_i)$. This point of view slightly generalises the setting of Banach space valued random variables and we adopt this language below. We summarise in this notation the results obtained so far.

PROPOSITION 2.2. *Let Z be as above and set $\Sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^n f^2(X_i)$. Then*

$$\mathbb{E}(Z^2) - \mathbb{E}(Z)^2 \leq 4\,\mathbb{E}(\Sigma^2)$$

and, denoting by $F(\lambda) = \mathbb{E}(\mathrm{e}^{\lambda Z})$, $\lambda \geq 0$, the Laplace transform of Z,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq 2\mathbb{E}(\Sigma^2)\lambda^2 F(\lambda) + 2\lambda^2 \mathbb{E}\big(\Sigma^2 \mathrm{e}^{\lambda Z}\big) \tag{2.7}$$

for every $\lambda \geq 0$.

The preceding differential inequality (2.7) on the Laplace transform of Z will be the key to the Gaussian bounds on Z. In order to describe the Poissonian behavior, it should be completed with a somewhat different inequality, that is however also obtained via logarithmic Sobolev inequalities. (The reader only interested in Gaussian estimates might want to skip this sligthly more technical part at first reading.) To this end, we simply estimate in a different way entropy in dimension one. We start again in a finite dimensional setting and recall

$$f(x) = \max_{1 \leq k \leq N} \sum_{i=1}^n x_i^k$$

on $E = \mathbb{R}^N$, $x = (x_1, \ldots, x_n) \in E^n$. We will assume here that the μ_i's are concentrated on $[0,1]^N \subset E$. By the variational caracterisation of entropy (1.5), for every $i = 1, \ldots, n$ and $\lambda \geq 0$,

$$\int \lambda f \mathrm{e}^{\lambda f} d\mu_i - \int \mathrm{e}^{\lambda f} d\mu_i \log \int \mathrm{e}^{\lambda f} d\mu_i = \inf_{c>0} \int \big[\lambda f \mathrm{e}^{\lambda f} - (\log c + 1)\,\mathrm{e}^{\lambda f} + c\big]\, d\mu_i$$

(in which it is understood that we integrate $f(x) = f(x_1, \ldots, x_n)$ with respect to x_i with the other coordinates fixed). For every $x = (x_1, \ldots, x_n)$ in E^n, and $i = 1, \ldots, n$, set $y = (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$. Now, if $c = \mathrm{e}^{\lambda f(y)}$, $\lambda \geq 0$, for every x,

$$\lambda f(x)\,\mathrm{e}^{\lambda f(x)} - (\log c + 1)\,\mathrm{e}^{\lambda f(x)} + c = \big[\lambda\big(f(x) - f(y)\big) - 1 + \mathrm{e}^{-\lambda(f(x)-f(y))}\big]\mathrm{e}^{\lambda f(x)}.$$

If $x_i \in [0,1]^N$,

$$f(x) = f(x_1, \ldots, x_n) \geq f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = f(y).$$

Furthermore, by (2.1),

$$0 \leq f(x) - f(y) \leq \sum_{k=1}^N \alpha_k x_i^k = \langle \alpha, x_i \rangle.$$

The function $u - 1 + \mathrm{e}^{-u}$ is increasing in $u \geq 0$. Therefore,

$$\left[\lambda\big(f(x) - f(y)\big) - 1 + \mathrm{e}^{-\lambda(f(x)-f(y))}\right]\mathrm{e}^{\lambda f(x)} \leq \left[\lambda\langle\alpha, x_i\rangle - 1 + \mathrm{e}^{-\lambda\langle\alpha, x_i\rangle}\right]\mathrm{e}^{\lambda f(x)}.$$

Now, if $x_i \in [0,1]^N$, and thus $0 \leq \langle\alpha, x_i\rangle \leq 1$, and if $\lambda \geq \frac{1}{4}$,

$$1 - \mathrm{e}^{-\lambda\langle\alpha, x_i\rangle} \geq \frac{1}{8}\,\langle\alpha, x_i\rangle.$$

Since the μ_i's are concentrated on $[0,1]^N$, it follows from the preceding that, for every $x \in E^n$, $i = 1, \ldots, n$ and $\lambda \geq \frac{1}{4}$,

$$\int \lambda f \mathrm{e}^{\lambda f} d\mu_i - \int \mathrm{e}^{\lambda f} d\mu_i \log \int \mathrm{e}^{\lambda f} d\mu_i \leq \Big(\lambda - \frac{1}{8}\Big) \int \langle\alpha, x_i\rangle \mathrm{e}^{\lambda f(x)} d\mu_i(x_i).$$

Tensorising this inequality as in the proof of Theorem 1.2 (cf. Proposition 4.1 below) immediately yields, for $\lambda \geq \frac{1}{4}$ thus,

$$\begin{aligned}\int \lambda f \mathrm{e}^{\lambda f} dP - \int \mathrm{e}^{\lambda f} dP \log \int \mathrm{e}^{\lambda f} dP &\leq \Big(\lambda - \frac{1}{8}\Big) \sum_{i=1}^{n} \int \langle\alpha, x_i\rangle \mathrm{e}^{\lambda f(x)} dP(x) \\ &\leq \Big(\lambda - \frac{1}{8}\Big) \int f \mathrm{e}^{\lambda f} dP\end{aligned}$$

since

$$\sum_{i=1}^{n} \langle\alpha, x_i\rangle = \sum_{i=1}^{n}\sum_{k=1}^{N} \alpha_k x_i^k = \max_{1 \leq k \leq N} \sum_{i=1}^{n} x_i^k = f(x).$$

In another words, if F is the Laplace transform of f, and when the μ_i's are concentrated on $[0,1]^N$ thus,

$$F'(\lambda) \leq 8F(\lambda)\log F(\lambda), \quad \lambda \geq \tfrac{1}{4}\,.$$

Note that since $u - 1 + \mathrm{e}^{-u} \leq \frac{u^2}{2}$ for $u \geq 0$, the preceding argument shows that (2.3) may be improved in this setting to

$$\begin{aligned}\int \lambda f \mathrm{e}^{\lambda f} dP - \int \mathrm{e}^{\lambda f} dP \log \int \mathrm{e}^{\lambda f} dP& \\ \leq \frac{\lambda^2}{2} \int \max_{1 \leq k \leq N} \sum_{i=1}^{n} (x_i^k)^2 \mathrm{e}^{\lambda f(x)} dP(x).&\end{aligned} \tag{2.8}$$

For simplicity however, we will not use this below.

In probabilistic notation, the preceding argument for example applies to the case the supremum Z in (2.6) is defined with a class $\mathcal{F}$ of functions f such that $0 \leq f \leq 1$. We may therefore state.

Proposition 2.3. *Let Z be defined as in (2.6) with a class $\mathcal{F}$ of functions f such that $0 \leq f \leq 1$ and denote by F its Laplace transform. Then*

$$F'(\lambda) \leq 8F(\lambda)\log F(\lambda) \tag{2.9}$$

for every $\lambda \geq \frac{1}{4}$.

The differential inequalities (2.7) and (2.9) on Laplace transforms of Propositions 2.2 and 2.3 may thus be used to yield sharp bounds on the tail of supremum Z over a class $\mathcal{F}$ of functions. (While (2.7) is used for general classes, (2.9) only applies as we have seen to classes of functions f such that $0 \leq f \leq 1$.) More precisely, they will provide precise deviation inequalities from the mean of statistical interest in which (2.7) will be used to describe the Gaussian behavior and (2.9) the Poissonian behavior. The following statement is a first result in this direction. It has been established recently by M. Talagrand (1995b). The proof here is elementary.

THEOREM 2.4. *Assume that* $0 \leq f \leq 1$, $f \in \mathcal{F}$. *Then, for every* $t \geq 0$,

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + t\big) \leq \exp\bigg(-\frac{t}{K}\,\log\bigg(1 + \frac{t}{\mathbb{E}(Z)}\bigg)\bigg).$$

for some numerical constant $K > 0$.

Proof. We first show the main Gaussian bound

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + t\big) \leq \exp\bigg(-\frac{1}{K}\,\min\bigg(t\,,\ \frac{t^2}{\mathbb{E}(Z)}\bigg)\bigg), \quad t \geq 0, \tag{2.10}$$

for some numerical $K > 0$, using the differential inequality (2.7) of Proposition 2.2. In the process of the proof, we found it easier to write down explicitely some numerical constants. (These constants are not sharp and we did not try to improve them. Some sharper constants may however be obtained through (2.8).) Since $0 \leq f \leq 1$ for every $f \in \mathcal{F}$,

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

Hence, (2.7) reads in this case, for every $\lambda \geq 0$,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq 2\mathbb{E}(Z)\lambda^2 F(\lambda) + 2\lambda^2 F'(\lambda).$$

Setting, as in Section 1, $H(\lambda) = \frac{1}{\lambda}\log F(\lambda)$, we see that

$$H'(\lambda) \leq 2\mathbb{E}(Z) + 2\,\frac{F'(\lambda)}{F(\lambda)}\,.$$

Therefore, for every $\lambda \geq 0$,

$$H(\lambda) \leq H(0) + 2\lambda\mathbb{E}(Z) + 2\log F(\lambda).$$

Since $H(0) = \mathbb{E}(Z)$, we get that, for every $\lambda \geq 0$,

$$F(\lambda) = \mathbb{E}\big(\mathrm{e}^{\lambda Z}\big) \leq \mathrm{e}^{\lambda\mathbb{E}(Z) + 2\lambda^2\mathbb{E}(Z) + 2\lambda\log F(\lambda)}. \tag{2.11}$$

When $\lambda \leq \frac{1}{4}$, (2.11) implies that

$$F(\lambda) \leq \mathrm{e}^{\lambda \mathbb{E}(Z)+2\lambda^2\mathbb{E}(Z)} F(\lambda)^{1/2}$$

so, when $0 \leq \lambda \leq \frac{1}{4}$,

$$F(\lambda) \leq \mathrm{e}^{2\lambda \mathbb{E}(Z)+4\lambda^2\mathbb{E}(Z)} \leq \mathrm{e}^{3\lambda\mathbb{E}(Z)}. \tag{2.12}$$

Taking this estimate back in (2.11) yields, always for $0 \leq \lambda \leq \frac{1}{4}$,

$$\mathbb{E}\big(\mathrm{e}^{\lambda[Z-\mathbb{E}(Z)]}\big) \leq \mathrm{e}^{2\lambda^2\mathbb{E}(Z)+6\lambda^2\mathbb{E}(Z)} \leq \mathrm{e}^{8\lambda^2\mathbb{E}(Z)}.$$

Now, by Chebyshev's inequality,

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z)+t\big) \leq \mathrm{e}^{-\lambda t+8\lambda^2\mathbb{E}(Z)}.$$

Choose $\lambda = t/16\mathbb{E}(Z)$ if $t \leq 4\mathbb{E}(Z)$ and $\lambda = \frac{1}{4}$ if $t \geq 4\mathbb{E}(Z)$ so that, for every $t \geq 0$,

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z)+t\big) \leq \exp\bigg(-\frac{1}{8}\,\min\bigg(t\,,\,\frac{t^2}{4\mathbb{E}(Z)}\bigg)\bigg)$$

and (2.10) is established.

We now prove, using Proposition 2.3, that

$$\mathbb{P}(Z \geq t) \leq \exp\bigg(-\frac{t}{K}\log\frac{t}{\mathbb{E}(Z)}\bigg) \tag{2.13}$$

for every $t \geq K\mathbb{E}(Z)$ for some large enough numerical constant K. This inequality together with (2.10) yields the full conclusion of Theorem 2.4. Integrating (2.9) shows that

$$F(\lambda) \leq \mathrm{e}^{\log F(1/4)\mathrm{e}^{8\lambda}}, \quad \lambda \geq \tfrac{1}{4}\,.$$

Furthermore, by (2.12), $\log F(1/4) \leq \mathbb{E}(Z)$. By Chebyshev's inequality, we get that, for every $\lambda \geq \frac{1}{4}$ and $t \geq 0$,

$$\mathbb{P}(Z \geq t) \leq \mathrm{e}^{-\lambda t+\mathbb{E}(Z)\mathrm{e}^{8\lambda}}.$$

Choose then $\lambda = \frac{1}{8}\log(t/\mathbb{E}(Z))$ provided that $t \geq 8\mathbb{E}(Z)$. The claim (2.13) easily follows and the proof of Theorem 2.4 is thus complete. □

Bounds on general sums are a little more involved but noch much.

THEOREM 2.5. *Under the previous notation, assume that $|f| \leq C$ for every $f \in \mathcal{F}$ and recall $\Sigma^2 = \sup_{f\in\mathcal{F}}\sum_{i=1}^n f^2(X_i)$. Then, for every $t \geq 0$,*

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z)+t\big) \leq 3\exp\bigg(-\frac{1}{K}\cdot\frac{t}{C}\log\bigg(1+\frac{Ct}{\mathbb{E}(\Sigma^2)+C\mathbb{E}(Z)}\bigg)\bigg)$$

for some numerical constant $K > 0$.

As is classical in Probability in Banach spaces (cf. Ledoux and Talagrand (1991), Lemmas 6.6 and 6.3), if $\mathbb{E} f(X_i) = 0$ for every $f \in \mathcal{F}$ and $i = 1, \ldots, n$,

$$\mathbb{E}(\Sigma^2) \le \sigma^2 + 8C\mathbb{E}(\overline{Z})$$

where $\sigma^2 = \sup_{f\in\mathcal{F}} \sum_{i=1}^n \mathbb{E} f^2(X_i)$ and $\overline{Z} = \sup_{f\in\mathcal{F}} \big|\sum_{i=1}^n f(X_i)\big|$ (if Z is defined without absolute values). Hence Theorem 2.5 immediately yields the following corollary. This type of estimate corresponds to the classical exponential bounds for sums of independent real valued random variables, with a Gaussian behavior for the small values of t and a Poissonian behavior for the large values. It is as general and sharp as possible (besides numerical constants) to recover all the vector valued extensions of classical limit theorems and bounds on tails for sums of independent random variables (cf. Ledoux and Talagrand (1991), Chapters 6 and 8).

Corollary 2.6. *Assume that* $|f| \le C$, $f \in \mathcal{F}$, *and that* $\mathbb{E} f(X_i) = 0$ *for every* $f \in \mathcal{F}$ *and* $i = 1, \ldots, n$. *Recall* $\sigma^2 = \sup_{f\in\mathcal{F}} \sum_{i=1}^n \mathbb{E} f^2(X_i)$. *Then, for every* $t \ge 0$,

$$\mathbb{P}\big(Z \ge \mathbb{E}(Z) + t\big) \le 3 \exp\bigg(-\frac{1}{K} \cdot \frac{t}{C} \log\bigg(1 + \frac{Ct}{\sigma^2 + C\mathbb{E}(\overline{Z})}\bigg)\bigg)$$

for some numerical constant $K > 0$.

As announced, Theorems 2.4 and 2.5 were obtained recently by M. Talagrand (1995b) as a further development of his abstract investigation of isoperimetric and concentration inequalities in product spaces. (Talagrand's formulation of Theorem 2.5 actually only involves $\mathbb{E}(\Sigma^2)$ in the logarithmic factor rather than $\mathbb{E}(\Sigma^2) + C\mathbb{E}(Z)$. For the applications through Corollary 2.6, this however does not make any difference.) The self-contained proofs presented here are much simpler. The main interest of these statements lies in the exact control of the deviation from the mean, that is the Gaussian estimate for the small values of t. The previous known bounds only concerned $t \ge K\mathbb{E}(Z)$ where $K > 0$ is some numerical constant. They were obtained by M. Talagrand as a consequence of either his abstract control by a finite number of points, or, but with some more efforts, of the convex hull approximation (cf. Talagrand (1989), (1994a), (1995a)). The new feature of Theorems 2.4 and 2.5 is that they allow deviation inequalities exactly from the mean, a result of strong statistical interest. That such bounds may be obtained is considered by M. Talagrand in his recent paper Talagrand (1995b) as "a result at the center of the theory".

Now, we turn to the proof of Theorem 2.5. It is similar to that of Theorem 2.4.

Proof of Theorem 2.5. We may assume by homogeneity that $C = 1$. We start again with the main Gaussian bound

$$\mathbb{P}\big(Z \ge \mathbb{E}(Z) + t\big) \le 2 \exp\bigg(-\frac{1}{K} \min\bigg(t\,, \frac{t^2}{\mathbb{E}(\Sigma^2) + \mathbb{E}(Z)}\bigg)\bigg) \tag{2.14}$$

for every $t \geq 0$ and some numerical constant $K > 0$. We use the differential inequality (2.7)

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq 2\mathbb{E}(\Sigma^2)\lambda^2 F(\lambda) + 2\lambda^2\mathbb{E}\big(\Sigma^2 e^{\lambda Z}\big), \quad \lambda \geq 0.$$

We first study the term $\mathbb{E}(\Sigma^2 e^{\lambda Z})$. We can write, for every $0 \leq \lambda \leq \frac{1}{8}$,

$$\begin{aligned}\mathbb{E}\big(\Sigma^2 e^{\lambda Z}\big) &\leq \int_{\{\Sigma^2 \leq 6\mathbb{E}(\Sigma^2)+Z\}} \Sigma^2 e^{\lambda Z} d\mathbb{P} + \int_{\{\Sigma^2 \geq 6\mathbb{E}(\Sigma^2)+Z\}} \Sigma^2 e^{\lambda Z} d\mathbb{P} \\ &\leq 6\mathbb{E}(\Sigma^2)F(\lambda) + F'(\lambda) + e^{-6\mathbb{E}(\Sigma^2)/8}\mathbb{E}\big(\Sigma^2 e^{\Sigma^2/8}\big) \\ &\leq 6\mathbb{E}(\Sigma^2)F(\lambda) + F'(\lambda) + 8\,e^{-3\mathbb{E}(\Sigma^2)/4}\mathbb{E}\big(e^{\Sigma^2/4}\big).\end{aligned}$$

Since $\Sigma^2 = \sup_{f\in\mathcal{F}}\sum_{i=1}^n f^2(X_i)$ and $0 \leq f^2 \leq 1$, $f \in \mathcal{F}$, by Theorem 2.4, more precisely (2.12) of the proof of Theorem 2.4,

$$\mathbb{E}\big(e^{\Sigma^2/4}\big) \leq e^{3\mathbb{E}(\Sigma^2)/4}.$$

We thus obtained that for every $0 \leq \lambda \leq \frac{1}{8}$,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq \big(14\mathbb{E}(\Sigma^2) + 16\big)\lambda^2 F(\lambda) + 2\lambda^2 F'(\lambda). \tag{2.15}$$

From here, the argument is exactly the same as the one used in the proof of Theorem 2.4. The differential inequality (2.15) yields, for $0 \leq \lambda \leq \frac{1}{8}$,

$$F(\lambda) \leq e^{\lambda\mathbb{E}(Z) + (14\mathbb{E}(\Sigma^2)+16)\lambda^2 + 2\lambda\log F(\lambda)}. \tag{2.16}$$

In particular,

$$F(\lambda) \leq e^{\lambda\mathbb{E}(Z) + (14\mathbb{E}(\Sigma^2)+16)\lambda^2} F(\lambda)^{1/2}$$

and thus ($\lambda \leq \frac{1}{8}$)

$$F(\lambda) \leq e^{2\lambda\mathbb{E}(Z) + 2(14\mathbb{E}(\Sigma^2)+16)\lambda^2} \leq e^{2\lambda\mathbb{E}(Z) + (7/2)\lambda\mathbb{E}(\Sigma^2) + 4\lambda}.$$

Coming back to (2.16),

$$F(\lambda) \leq e^{\lambda\mathbb{E}(Z) + (14\mathbb{E}(\Sigma^2)+16)\lambda^2 + 4\lambda^2\mathbb{E}(Z) + 7\lambda^2\mathbb{E}(\Sigma^2) + 8\lambda^2}.$$

Hence, for every $0 \leq \lambda \leq \frac{1}{8}$,

$$\mathbb{E}\big(e^{\lambda[Z-\mathbb{E}(Z)]}\big) \leq 2\,e^{21\lambda^2\mathbb{E}(\Sigma^2) + 4\lambda^2\mathbb{E}(Z)}.$$

Optimising in $\lambda \leq \frac{1}{8}$ together with Chebyshev's inequality, it follows that, for every $t \geq 0$,

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + t\big) \leq 2\exp\left(-\frac{1}{4}\min\left(\frac{t}{4}, \frac{t^2}{21\mathbb{E}(\Sigma^2) + 4\mathbb{E}(Z)}\right)\right).$$

Inequality (2.14) is thus established.

We turn to the Poissonian bound

$$\mathbb{P}(Z \geq t) \leq 3\exp\left(-\frac{t}{K}\log\left(\frac{t}{\mathbb{E}(\Sigma^2)}\right)\right) \tag{2.17}$$

for $t \geq K(\mathbb{E}(\Sigma^2) + \mathbb{E}(Z))$ with some numerical constant K. Together with (2.14), the proof of Theorem 2.5 will be complete. We follow the truncation argument of Talagrand (1995b). For every $t \geq 0$,

$$\mathbb{P}(Z \geq 3t) \leq \mathbb{P}(Z_\rho \geq 2t) + \mathbb{P}(W \geq t)$$

where

$$Z_\rho = \sup_{f\in\mathcal{F}_\rho}\sum_{i=1}^n f(X_i) \quad \text{or} \quad \sup_{f\in\mathcal{F}_\rho}\left|\sum_{i=1}^n f(X_i)\right|$$

with $\mathcal{F}_\rho = \{fI_{\{|f|\leq\rho\}}; f \in \mathcal{F}\}$, $\rho > 0$ to be determined, and

$$W = \sup_{f\in\mathcal{F}}\sum_{i=1}^n \big|f(X_i)\big| I_{\{|f(X_i)|>\rho\}}.$$

We use (2.14) for Z_ρ to get that, for some constant K_1 and by homogeneity,

$$\mathbb{P}\big(Z_\rho \geq \mathbb{E}(Z_\rho) + t\big) \leq 2\exp\left(-\frac{t}{K_1\rho}\right) \tag{2.18}$$

provided that $t \geq \mathbb{E}(\Sigma^2)/\rho + \mathbb{E}(Z_\rho)$. We may apply Theorem 2.4, more precisely (2.13), to W to get

$$\mathbb{P}(W \geq t) \leq \exp\left(-\frac{t}{K_2}\log\frac{t}{\mathbb{E}(W)}\right) \tag{2.19}$$

if $t \geq K_2\mathbb{E}(W)$ for some constant K_2 which we may assume ≥ 4. Let now t be fixed such that $t \geq K_2^2(\mathbb{E}(\Sigma^2)+\mathbb{E}(Z))$ and choose $\rho = \rho(t) = \sqrt{\mathbb{E}(\Sigma^2)/t}$. Then, since $W \leq \Sigma^2/\rho$,

$$t \geq K_2\sqrt{t\,\mathbb{E}(\Sigma^2)} = K_2\,\frac{\mathbb{E}(\Sigma^2)}{\rho} \geq K_2\mathbb{E}(W)$$

and

$$\big|\mathbb{E}(Z_\rho) - \mathbb{E}(Z)\big| \leq \mathbb{E}(W) \leq \sqrt{t\,\mathbb{E}(\Sigma^2)} \leq \frac{t}{K_2} \leq \frac{t}{4}$$

so that $\mathbb{E}(Z_\rho) \leq t/2$ since $t \geq 4\mathbb{E}(Z)$. Therefore (2.18) and (2.19) hold and, by the choice of ρ,

$$\begin{aligned}\mathbb{P}(Z_\rho \geq 2t) &\leq \mathbb{P}\big(Z_\rho \geq \mathbb{E}(Z_\rho) + t\big)\\ &\leq 2\exp\left(-\frac{t^{3/2}}{K_1\sqrt{\mathbb{E}(\Sigma^2)}}\right) \leq 2\exp\left(-\frac{t}{K_1}\log\frac{t}{\mathbb{E}(\Sigma^2)}\right)\end{aligned}$$

and

$$\mathbb{P}(W \geq t) \leq \exp\left(-\frac{t}{K_2}\log\frac{t}{\mathbb{E}(W)}\right) \leq \exp\left(-\frac{t}{2K_2}\log\frac{t}{\mathbb{E}(\Sigma^2)}\right).$$

(2.17) follows and Theorem 2.5 is therefore established. □

3. Deviation inequalities for chaos

In this section, we come back to the setting of the first part, but we will be interested in sharp deviation inequalities for chaos in the spirit of (1.9). Assume thus we are given independent random variables η_i, $i = 1, \ldots, n$ such that $|\eta_i| \leq 1$ almost surely for every i. Let also a_{ij}, $i, j = 1, \ldots, n$, be elements in some Banach space $(E, \|\cdot\|)$ such that $a_{ij} = a_{ji}$ and $a_{ii} = 0$. We are interested in the deviation of the random variable $Z = \|\sum_{i,j=1}^n a_{ij}\eta_i\eta_j\|$ from its mean. We are thus dealing with the function on $\mathbb{R}^n$ defined by

$$f(x) = \left\|\sum_{i,j=1}^n a_{ij}x_ix_j\right\|, \quad x = (x_1, \ldots, x_n).$$

To study functional inequalities for such a function, we make advantage of the fact that Theorems 1.1 and 1.2 hold for separately convex functions, which is precisely the case with this f $(a_{ii} = 0)$. Let

$$\Sigma = \sup_{\|\xi\|\leq 1}\left(\sum_{i=1}^n\left(\sum_{j=1}^n \langle\xi, a_{ij}\rangle\eta_j\right)^2\right)^{1/2} = \sup_{\|\xi\|\leq 1}\sup_{|\alpha|\leq 1}\sum_{i,j=1}^n \alpha_i\eta_j\langle\xi, a_{ij}\rangle$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ and $|\alpha|$ denotes its Euclidean norm. (Without the symmetry assumption on the a_{ij}'s, we would need to consider also the expression symmetric in i and j.) Set also

$$\sigma = \sup_{\|\xi\|\leq 1}\sup_{|\alpha|\leq 1}\sup_{|\beta|\leq 1}\sum_{i,j=1}^n \alpha_i\beta_j\langle\xi, a_{ij}\rangle.$$

Deviation inequalities for chaos have been obtained e.g. in Ledoux and Talagrand (1991), Chapter 4. Again, they concern deviation from a multiple of the median M and allow only a control of the probabilities $\mathbb{P}(Z \geq 2M+t)$. A somewhat more precise version of the following statement for symmetric Bernoulli random variables is established in Talagrand (1995b).

Theorem 3.1. *Under the preceding notation, for every $t \geq 0$,*

$$\mathbb{P}\big(Z \geq \mathbb{E}(Z) + t\big) \leq 2\exp\left(-\frac{1}{K}\min\left(\frac{t}{\sigma}, \frac{t^2}{\sigma\mathbb{E}(Z) + \mathbb{E}(\Sigma)^2}\right)\right)$$

for some numerical constant $K > 0$. Furthermore, there is a numerical $K > 0$ such that, for every $0 < \varepsilon \leq 1$ and every $t \geq 0$,

$$\mathbb{P}\big(Z \geq (1+\varepsilon)\mathbb{E}(Z) + t\big) \leq 2\exp\left(-\frac{1}{K}\min\left(\frac{\varepsilon t}{\sigma}, \frac{t^2}{\mathbb{E}(\Sigma)^2}\right)\right).$$

Proof. Denote by $F(\lambda) = \mathbb{E}(e^{\lambda Z})$, $\lambda \geq 0$, the Laplace transform of Z. Since as we have seen Z is given by a separately convex function, the basic inequalities of Section 1 apply. In particular (1.7) implies that (recall $|\eta_i| \leq 1$), for every $\lambda \geq 0$,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq 8\lambda^2 \mathbb{E}\big(\Sigma^2 e^{\lambda Z}\big). \tag{3.1}$$

We may and do assume by homogeneity that $\sigma = 1$. Observe that Σ is also given by a separately convex function with furthermore a Lipschitz constant less than or equal to $\sigma = 1$. Therefore, by (1.8),

$$\mathbb{P}\big(\Sigma \geq \mathbb{E}(\Sigma) + t\big) \leq e^{-t^2/8}, \quad t \geq 0. \tag{3.2}$$

We have now just, more or less, to properly combine (3.1) and (3.2). As in the proof of Theorem 2.5, one may write, for every $0 \leq \lambda \leq \lambda_0$,

$$\begin{aligned}\mathbb{E}\big(\Sigma^2 e^{\lambda Z}\big) &\leq \int_{\{\Sigma^2 \leq 8\mathbb{E}(\Sigma)^2 + Z\}} \Sigma^2 e^{\lambda Z} d\mathbb{P} + \int_{\{\Sigma^2 \geq 8\mathbb{E}(\Sigma)^2 + Z\}} \Sigma^2 e^{\lambda Z} d\mathbb{P} \\ &\leq 8\mathbb{E}(Z)^2 F(\lambda) + F'(\lambda) + e^{-8\lambda_0 \mathbb{E}(\Sigma)^2} \mathbb{E}\big(\Sigma^2 e^{\lambda_0 \Sigma^2}\big).\end{aligned}$$

Now, by the integration by parts formula,

$$\begin{aligned}\mathbb{E}\big(\Sigma^2 e^{\lambda_0 \Sigma^2}\big) &\leq \frac{1}{\lambda_0}\, \mathbb{E}\big(e^{2\lambda_0 \Sigma^2}\big) \\ &\leq \frac{1}{\lambda_0}\, e^{8\lambda_0 \mathbb{E}(\Sigma)^2} + \frac{4}{\lambda_0} \int_{2\mathbb{E}(\Sigma)}^{\infty} \mathbb{P}(\Sigma \geq t)\, t\, e^{2\lambda_0 t^2} dt \\ &\leq \frac{1}{\lambda_0}\, e^{8\lambda_0 \mathbb{E}(\Sigma)^2} + \frac{4}{\lambda_0} \int_0^{\infty} t\, e^{2\lambda_0 t^2 - t^2/32} dt\end{aligned}$$

where we used (3.2). Set $\lambda_0 = 1/128$. Calculus shows that

$$e^{-8\lambda_0 \mathbb{E}(\Sigma)^2} \mathbb{E}\big(\Sigma^2 e^{\lambda_0 \Sigma^2}\big) \leq K = 128 + 128^2.$$

The differential inequality (3.1) then reads,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq 64\mathbb{E}(\Sigma)^2 \lambda^2 F(\lambda) + 8\lambda^2 F'(\lambda) + 8K\lambda^2, \quad 0 \leq \lambda \leq \lambda_0.$$

Its integration gives

$$F(\lambda) \leq e^{\lambda \mathbb{E}(Z) + (64\mathbb{E}(\Sigma)^2 + 8K)\lambda^2 + 8\lambda \log F(\lambda)}. \tag{3.3}$$

Arguing as in the proofs of Theorems 2.4 and 2.5, we get that, for every $0 \leq \lambda \leq \lambda_0$,

$$\mathbb{E}\big(e^{\lambda[Z - \mathbb{E}(Z)]}\big) \leq 2\, e^{K'(\mathbb{E}(Z) + \mathbb{E}(\Sigma)^2)\lambda^2}$$

where K' is some further numerical constant. A simple use of Chebyshev's inequality then yields the first part of Theorem 3.1.

If $16\lambda \le \varepsilon/1+\varepsilon$ for some $\varepsilon > 0$ small enough, (3.3) implies that

$$\mathbb{E}\big(e^{\lambda Z}\big) \le e^{(1+\varepsilon)\mathbb{E}(Z)\lambda + (1+\varepsilon)(64\mathbb{E}(\Sigma)^2+8K)\lambda^2}$$

from which the second part of the statement follows. The proof of Theorem 3.1 is complete. □

4. Iteration and concentration for the Hamming distance

In the last part of this work, we first isolate the basic iteration procedure for functions on a product space. We learned the argument in its full generality from S. Bobkov. We adopt a somewhat general formulation in order to include in the same pattern Poincaré and logarithmic Sobolev inequalities. The statement we present is the general iteration result which reduces to estimates in dimension one. At least in case of variance and entropy, it is a well-known statement.

Let Φ be a convex function on some closed interval of $\mathbb{R}$. If μ is a probability measure, consider the non-negative functional

$$E_{\Phi,\mu}(g) = \int \Phi(g)\,d\mu - \Phi\Big(\int g\,d\mu\Big)$$

(under appropriate range and integrability conditions on g). We consider convex functions Φ such that $E_{\Phi,\mu}$ defines a convex functional for every μ in the sense that

$$E_{\Phi,\mu}\Big(\sum_k \alpha_k g_k\Big) \le \sum_k \alpha_k E_{\Phi,\mu}(g_k) \tag{4.1}$$

for every $\alpha_k \ge 0$ with $\sum_k \alpha_k = 1$ and functions g_k. A first example of such convex functionals is the variance ($\Phi(x) = x^2$) since

$$\int g^2 d\mu - \Big(\int g\,d\mu\Big)^2 = \frac{1}{2}\int\!\!\int \big|g(x)-g(y)\big|^2 d\mu(x)d\mu(y).$$

Another one is entropy ($\Phi(x) = x\log x$ on $[0,\infty)$) since,

$$\int g\log g\,d\mu - \int g\,d\mu \log\int g\,d\mu = \sup\int fg\,d\mu$$

where the supremum is running over all f's with $\int e^f d\mu \le 1$. Further examples of interest have been described by S. Bobkov (private communication) via analytical conditions on Φ.

Now, consider $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1,\dots,n$, arbitrary probability spaces. Denote by $P = P_n$ the product probability $\mu_1 \otimes \cdots \otimes \mu_n$ on the product space $\Omega = \Omega_1 \times \cdots \times \Omega_n$. A generic point in Ω is denoted $x = (x_1,\dots,x_n)$. If g is a function on the product space Ω, for every $1 \le i \le n$, let g_i be the function on Ω_i defined by

$$g_i(x_i) = g(x_1,\dots,x_i,\dots,x_n)$$

with x_j, $j \neq i$, fixed.

PROPOSITION 4.1. *Let Φ be convex satisfying (4.1). Then, for any g and any product probability $P = \mu_1 \otimes \cdots \otimes \mu_n$ on Ω,*

$$E_{\Phi,P}(g) \leq \sum_{i=1}^{n} \int E_{\Phi,\mu_i}(g_i) dP.$$

Proof. By induction on n. The case $n = 1$ is of course trivial. Assume the proposition holds for P_{n-1} and let us prove it for $P_n = P$. Write, by Fubini's theorem and the induction hypothesis,

$$\begin{aligned}\int \Phi(g) d\mu &= \int d\mu_n(x_n) \bigg[\int \Phi\big(g(z, x_n)\big) dP_{n-1}(z) \bigg] \\ &\leq \int d\mu_n(x_n) \bigg[\Phi\bigg(\int g(z, x_n) dP_{n-1}(z) \bigg) \\ &\qquad + \sum_{i=1}^{n-1} \int E_{\Phi,\mu_i}\big(g_i(\cdot, x_n)\big) dP_{n-1} \bigg]\end{aligned}$$

where, as usual, $z = (x_1, \ldots, x_{n-1})$. Now, by the convexity property (4.1),

$$\int d\mu_n(x_n) \bigg[\Phi\bigg(\int g(z, x_n) dP_{n-1}(z) \bigg) \bigg] \leq \Phi\bigg(\int g dP_n \bigg) + \int E_{\Phi,\mu_n}(g_n) dP_n$$

from which the result follows. □

As an application we get for example, for any non-negative function g on Ω (or equivalently for g^2),

$$\begin{aligned}\int g \log g dP - \int g dP \log \int g dP \\ \leq \sum_{i=1}^{n} \int \bigg[\int g \log g d\mu_i - \int g d\mu_i \log \int g d\mu_i \bigg] dP.\end{aligned} \tag{4.2}$$

This reduces to estimates of the entropy in dimension one. In Sections 1 and 2, we present some gradient one-dimensional bounds for convex functions and perform there the tensorisation directly on the gradient estimates (a procedure that is somewhat simpler than to go through Proposition 4.1). However, once (4.2) has been isolated, the proofs reduce to dimension one in a really straightforward manner. As an application, we next observe that a trivial one-dimensional estimate similarly yields the classical concentration for the Hamming metric. This example was an important step in the development of the abstract general inequalities (cf. Talagrand (1995a)).

On $\Omega = \Omega_1 \times \cdots \times \Omega_n$, consider the Hamming distance given by

$$d(x, y) = \mathrm{Card}\{1 \leq i \leq n; x_i \neq y_i\}, \quad x, y \in \Omega.$$

It was shown in Milman and Schechtman (1986) (in a particular case that however trivially extends to the general case) and Talagrand (1995a) that for every Lipschitz function f on (Ω, d) with Lipschitz constant $\|f\|_{\rm Lip} \leq 1$ and every product probability P on Ω,

$$P\big(\big|f - \textstyle\int f dP\big| \geq t\big) \leq 2\,{\rm e}^{-t^2/4n} \tag{4.3}$$

for every $t \geq 0$.

To deduce this inequality from our approach, we apply (4.2) to ${\rm e}^{\lambda f}$ with $\lambda \in \mathbb{R}$ and $\|f\|_{\rm Lip} \leq 1$ with respect to the Hamming metric. We are then reduced to the estimate of the entropy of ${\rm e}^{\lambda f}$ in dimension one. Let thus first $n = 1$, and f on Ω with $|f(x) - f(y)| \leq c$ for every x, y and some $c > 0$. By Jensen's inequality,

$$\begin{aligned}\int f{\rm e}^f dP - \int {\rm e}^f dP \log \int {\rm e}^f dP \\ \leq \frac{1}{2} \int\!\!\int \big[f(x) - f(y)\big]\big[{\rm e}^{f(x)} - {\rm e}^{f(y)}\big] dP(x) dP(y).\end{aligned}$$

For every x, y,

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

where we used that $(u - v)({\rm e}^u - {\rm e}^v) \leq \frac{1}{2}(u - v)^2({\rm e}^u + {\rm e}^v)$ for every $u, v \in \mathbb{R}$. Therefore,

$$\int f{\rm e}^f dP - \int {\rm e}^f dP \log \int {\rm e}^f dP \leq \frac{c^2}{2} \int {\rm e}^f dP. \tag{4.4}$$

Let thus f on the product space $\Omega = \Omega_1 \times \ldots \times \Omega_n$ such that $\|f\|_{\rm Lip} \leq 1$ with respect to d. Applying (4.2) to ${\rm e}^{\lambda f}$, $\lambda \in \mathbb{R}$, and then (4.4),

$$\lambda \int f{\rm e}^{\lambda f} dP - \int {\rm e}^{\lambda f} dP \log \int {\rm e}^{\lambda f} dP \leq \frac{n\lambda^2}{2} \int {\rm e}^{\lambda f} dP.$$

That is, setting $F(\lambda) = \int {\rm e}^{\lambda f} dP$,

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \leq \frac{n\lambda^2}{2} F(\lambda), \quad \lambda \in \mathbb{R}.$$

This is the basic differential inequality of this work. As in Section 1, its integration shows that, for every $\lambda \in \mathbb{R}$,

$$\int {\rm e}^{\lambda f} dP \leq {\rm e}^{\lambda \int f dP + n\lambda^2/2}.$$

By Chebyshev's inequality, for every $t \geq 0$,

$$P\big(f \geq \textstyle\int f dP + t\big) \leq {\rm e}^{-t^2/2n}. \tag{4.5}$$

Together with the same inequality for $-f$, we find again (4.3), with a somewhat better constant. (This is however not quite the optimal bound obtained with the martingale method in McDiarmid (1989), see also Talagrand (1995a).) The same argument works for the Hamming metrics $\sum_{i=1}^n a_i I_{\{x_i \neq y_i\}}$, $a_i \geq 0$.

The preceding development has also some interesting consequences to the concept of penalties introduced by M. Talagrand in Talagrand (1995a). Assume for simplicity that all the probability spaces $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, \ldots, n$, are identical. Let h be non-negative symmetric on $\Omega_1 \times \Omega_1$ and equal to 0 on the diagonal, and consider a function f on the product space Ω such that

$$\big|f(x) - f(y)\big| \leq \sum_{i=1}^n h(x_i, y_i), \quad x, y \in \Omega.$$

Then, if $F(\lambda) = \int \mathrm{e}^{\lambda f} dP$, we get as before

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \leq \lambda^2 \sum_{i=1}^n \iint h(x_i, y_i)^2 \mathrm{e}^{\lambda f(x)} dP(x) dP(y), \quad \lambda \in \mathbb{R}.$$

Set $\|h\|_\infty = \sup |h(x,y)|$ and $\|h\|_2^2 = \iint h(x,y)^2 d\mu_1(x) d\mu_1(y)$. Then, arguing as in the proofs of Theorem 2.5 or Theorem 3.1, we easily obtain that, for some numerical constant $K > 0$, and every $t \geq 0$,

$$P\big(f \geq \textstyle\int f dP + t\big) \leq 2 \exp\bigg(-\frac{1}{K} \, \min\bigg(\frac{t}{\|h\|_\infty}, \frac{t^2}{n\|h\|_2^2}\bigg)\bigg).$$

This inequality resembles the classical Bernstein inequality and was obtained by M. Talagrand (1995a) as a consequence of the study of penalties in this context.

Further work in the directions explored in this paper is still in progress.

Acknowledgements

I thank M. Talagrand for presenting and explaining to me over the years his work on product measures that finally led to the ideas developed here. I am grateful to S. Bobkov for his numerous observations and kind permission to include here Proposition 4.1. Thanks are also due to S. Kwapień, P. Massart, L. Saloff-Coste for several helpful comments and remarks and to P.-M. Samson for a careful reading of the manuscript.

References

Aida, S., Masuda, T. and Shigekawa. I, (1994), Logarithmic Sobolev inequalities and exponential integrability, *J. Funct. Anal.* **126** 83–101.

BOBKOV, S. (1994), Some extremal properties of Bernoulli distribution, *Theory Prob. Appl.* to appear.

BOBKOV, S. (1995), On Gross' and Talagrand's inequalities on the discrete cube, *Vestnik of Syktyvkar University* **Ser. 1, 1** 12–19 (in Russian).

BOBKOV, S. Private communication.

DAVIES, E. B. and SIMON, B. (1984), Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians, *J. Funct. Anal.* **59** 335–395.

DEMBO, A. (1995), Information inequalities and concentration of measure, *Ann. Probability,* to appear.

DEMBO, A. and ZEITOUNI, O. (1995), Transportation approach to some concentration inequalities in product spaces, *Preprint.*

GROMOV, M. and MILMAN, V. D. (1983), A topological application of the isoperimetric inequality, *Amer. J. Math.* **105** 843–854.

GROSS, L. (1975), Logarithmic Sobolev inequalities, *Amer. J. Math.* **97** 1061–1083.

HOLLEY, R. and STROOCK, D. (1987), Logarithmic Sobolev inequalities and stochastic Ising models, *J. Statist. Phys.* **46** 1159–1194.

JOHNSON, W. B. and SCHECHTMAN, G. (1987), Remarks on Talagrand's deviation inequality for Rademacher functions, *Longhorn Notes* Texas.

KWAPIEŃ, S. and SZULGA, J. (1991), Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces, *Ann. Probability* **19** 369–379.

LEDOUX, M. (1995), Remarks on logarithmic Sobolev constants, exponential integrability and bounds on the diameter, *J. Math. Kyoto Univ.* **35** 211–220.

LEDOUX, M. (1994), Isoperimetry and Gaussian Analysis, *Ecole d'Eté de Probabilités de St-Flour Lecture Notes in Math.*, Springer-Verlag, to appear.

LEDOUX, M. and TALAGRAND, M. (1991), Probability in Banach spaces (Isoperimetry and processes), *Ergebnisse der Mathematik und ihrer Grenzgebiete* Springer-Verlag.

McDIARMID, C. (1989), On the method of bounded differences. Surveys in Combinatorics. *London Math. Soc. Lecture Notes* **141** 148–188 Cambridge Univ. Press.

MARTON, K. (1995a), Bounding $\bar{d}$-distance by information divergence: a method to prove measure concentration, *Ann. Probability,* to appear.

MARTON, K. (1995b), A concentration of measure inequality for contracting Markov chains, *Geom. and Funct. Anal.,* to appear.

MAUREY, B. (1991), Some deviations inequalities, *Geometric and Funct. Anal.* **1** 188–197.

MILMAN, V. D. and SCHECHTMAN, G. (1986), Asymptotic theory of finite dimensional normed spaces, *Lecture Notes in Math.* **1200** Springer-Verlag.

TALAGRAND, M. (1988), An isoperimetric theorem on the cube and the Khintchine-Kahane inequalities, *Proc. Amer. Math. Soc.* **104** 905–909.

TALAGRAND, M. (1989), Isoperimetry and integrability of the sum of independent Banach space valued random variables, *Ann. Probability* **17** 1546–1570.

TALAGRAND, M. (1994a), Sharper bounds for Gaussian and empirical processes. *Ann. Probability* **22** 28–76.

TALAGRAND, M. (1995a), Concentration of measure and isoperimetric inequalities in product spaces, *Publications Mathématiques de l'I.H.E.S.* **81** 73–205.

TALAGRAND, M. (1994b), A new look at independence, *Ann. Probability* to appear.

TALAGRAND, M. (1995b), New concentration inequalities in product spaces, *Invent. math.* to appear.