Some deviation inequalities

by Bernard Maurey

September 1990

Abstract. We introduce a concentration property for probability measures on \( \mathbb{R}^n \), which we call Property (\( \tau \)); we show that this property has an interesting stability under products and contractions (Lemmas 1, 2, 3). Using property (\( \tau \)), we give a short proof for a recent deviation inequality due to Talagrand. In a third section, we also recover known concentration results for Gaussian measures using our approach.

Introduction. Very roughly speaking, a concentration of measure phenomenon for a probability measure \( \mu \) means that, given any measurable subset \( A \) such that \( \mu(A) \geq \frac{1}{2} \), the enlargements of \( A \), in a sense to be made precise, almost have measure 1. One important example of this situation is a consequence of Paul Lévy’s isoperimetric inequality for the unit sphere \( S_n \) of \( \mathbb{R}^{n+1} \). This consequence is the following: Let \( \mu_n \) denote the normalized rotation invariant measure on \( S_n \); for every measurable subset \( A \) of \( S_n \) such that \( \mu_n(A) \geq \frac{1}{2} \) we have

\[
\mu_n \{ x \in S_n; x \notin A_{\varepsilon} \} \leq \sqrt{\frac{n}{8}} e^{-n\varepsilon^2/2}
\]

where \( A_{\varepsilon} \) denotes the subset of \( S_n \) of all points whose geodesic distance to \( A \) is less than \( \varepsilon \) (see [8]). This result is crucial for the proof of Dvoretzky’s theorem (about almost spherical sections of convex bodies) as given by V. Milman [7] (see also Figiel, Lindenstrauss and Milman [4]). Using Lévy’s result, Borell [1] was able to prove an analogous isoperimetric result for the Gaussian measure on \( \mathbb{R}^n \) (see also Ehrhard [3]). As it is the case for the sphere, Borell’s isoperimetric result implies a Gaussian concentration of measure principle. Later, Pisier and the author gave a very simple proof for the Gaussian concentration principle which is needed for the proof of Dvoretzky’s theorem (see [9], [8] and [10]). Recently, Talagrand proved [12] a concentration principle for measures on \( \mathbb{R}^n \) with exponential densities which is stronger than the Gaussian one (Corollary 1 below; this Corollary appears in [12] as a consequence of a more precise isoperimetric inequality which does not follow from our proof). We give here a proof for Talagrand’s concentration result, using a property which we call Property (\( \tau \)). This property is defined in section I, and general stability results about it are stated (Lemmas 1, 2, 3); we also explain in section I why property (\( \tau \)) is related to concentration (Lemma 4). In section II, we show how to get Talagrand’s inequality using property (\( \tau \)). In section III we recover the Gaussian case along the same lines (Corollary 2). In a last section, we study a variant of property (\( \tau \)), the convex property (\( \tau \)); this variant is related to an other deviation inequality due to Talagrand [13] (see Corollary 5).

I. Property (\( \tau \))

Let \( f \) and \( g \) be two measurable functions on \( \mathbb{R}^n \); we denote by \( f \otimes g \) the inf-convolution of \( f \) and \( g \),

\[
(f \otimes g)(x) = \inf \left\{ f(x-y) + g(y); y \in \mathbb{R}^n \right\}.
\]

If \( \mu \) is a probability measure on \( \mathbb{R}^n \) and \( w \) a positive measurable function on \( \mathbb{R}^n \), we say that the couple \( (\mu, w) \) satisfies the Property (\( \tau \)) if for every bounded measurable function \( \varphi \) on \( \mathbb{R}^n \) we have

\[
\left( \int e^{\varphi \otimes w} \, d\mu \right) \left( \int e^{-\varphi} \, d\mu \right) \leq 1.
\]

If we adopt the convention \(+\infty, 0, 1\), we see easily that the above inequality extends to all \( \mathbb{R} \)-valued measurable functions \( \varphi \).

The definition of property (\( \tau \)) was motivated by Talagrand’s isoperimetric inequality for the cube, as the careful reader of [13] will notice.
Lemma 1. If $(\mu_i, w_i)$ satisfies $\tau$ on $\mathbb{R}^{n_i}$ for $i = 1, 2$, then $(\mu_1 \otimes \mu_2, w)$ satisfies $\tau$ on $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, where
$$w(x_1, x_2) = w_1(x_1) + w_2(x_2).$$
Proof: Consider $\varphi(x) = \varphi(x, y)$ and apply $\tau$ to $\psi(y) = \log(\int e^{\varphi} \cdot w \, d\mu_1)$.

Lemma 2. If $(\mu_i, w_i)$ satisfies $\tau$ on $\mathbb{R}^n$ for $i = 1, 2$, then $(\mu_1 \ast \mu_2, w_1 \ast w_2)$ satisfies $\tau$ on $\mathbb{R}^n$.

Lemma 3. Let $(\mu_1, w_1)$ satisfy $\tau$ on $\mathbb{R}^{n_1}$. Let $w_2$ be a positive measurable function on $\mathbb{R}^{n_2}$ and $F$ a mapping from $\mathbb{R}^{n_1}$ to $\mathbb{R}^{n_2}$ such that $w_2(Fx - Fy) \leq w_1(x - y)$ for every pair $x, y$. Let $\mu_2$ be the image probability measure on $\mathbb{R}^{n_2}$ defined by $\mu_2 = F(\mu_1)$. Then $(\mu_2, w_2)$ satisfies $\tau$.

Proof: Consider $\varphi \circ F \circ w_1 \geq (\varphi \circ w_2) \circ F$.

Lipschitz maps like the above $F$ were used by Pisier [9] in a slightly different context. Following him, we will be able in section III to pass from the Gaussian case to the uniform probability measure on $[0, 1]^n$ (Corollary 4).

We have perhaps to explain why property $\tau$ for a couple $(\mu, w)$ is a concentration of measure property.

Lemma 4. Assume that $(\mu, w)$ satisfies $\tau$ on $\mathbb{R}^n$. For every measurable subset $A$ of $\mathbb{R}^n$ and every positive real number $t$, we have
$$\mu\{x \notin A + \{w < t\} \} \leq (\mu(A))^{-1} e^{-t}.$$  

Proof: Let $\xi$ be a measurable subset of $\mathbb{R}^n$ and denote by $\varphi_{\xi}$ the function equal to 0 on $\xi$ and $+\infty$ outside.

We observe that $(\varphi_{\xi} \ast w_i)(x) \geq t$ when $x \notin \xi + \{w < t\} = \{a + y; a \in \xi, w(y) < t\}$. Property $\tau$ implies that
$$\int e^{\varphi_{\xi} \ast w_i} \, d\mu_i \leq (\mu(A))^{-1}$$
and we conclude with Tschebycheff’s inequality.

II. Talagrand’s deviation inequality

Let us define a function $W$ on $\mathbb{R}$ by
$$W(t) = \frac{1}{18} t^2 \quad \text{for} \quad |t| \leq 2, \quad \frac{2}{9} (|t| - 1) \quad \text{otherwise}.$$  

and let $\mu_c$ be the probability measure on $\mathbb{R}$ with density $1_{(0, \infty)}(x) e^{-x}$.

Proposition. The couple $(\mu_c, W)$ satisfies $\tau$.

It follows from Lemma 2 that $(\xi, U)$ also satisfies $\tau$, where $\xi$ is the convolution of $\mu_c$ and its symmetric image on $(-\infty, 0)$, and $U = W \ast W$. It is easy to see that $\xi$ has density $\frac{1}{2} e^{-|x|}$ on $\mathbb{R}$, and that
$$U(t) = 2W(t/2) = \frac{1}{36} t^2 \quad \text{for} \quad |t| \leq 4, \quad \frac{2}{9} (|t| - 2) \quad \text{otherwise}.$$  

We deduce now from Lemma 1 that the couple $(\xi_n, U_n)$ satisfies $\tau$ on $\mathbb{R}^n$ for every $n$, where $\xi_n$ is the product of $n$ copies of $\xi$ and $U_n(x) = \sum_{i=1}^n U(x_i)$. The idea of working with functions like $W$ or $U_n$ comes from Talagrand [12].

Theorem 1. The couple $(\xi_n, U_n)$ satisfies $\tau$ for every integer $n$. In particular, for every measurable subset $A$ of $\mathbb{R}^n$ we have if we set $\rho_A = \varphi_A \ast U_n$
$$\int e^{\rho_A} \, d\xi_n \leq (\xi_n(A))^{-1}.$$  

Corollary 1 (Talagrand). For every $t > 0$,
$$\xi_n \{x; x \notin A + 6\sqrt{t} B_2 + 9t B_1 \} \leq (\xi_n(A))^{-1} e^{-t}.$$  

where $B_2$ and $B_1$ are respectively the usual $\ell_2^2$ and $\ell_1^n$ balls.

Proof: According to Lemma 4, we need only show that
$$\{U_n < t\} \subset 6\sqrt{t} B_2 + 9t B_1.$$  

Assume $U_n(x) < t$, and define $y$ and $z$ in the following way: $y_i = x_i$ if $|x_i| \leq 4$, $y_i = 0$ otherwise; $z_i = x_i$ if $|x_i| > 4$, $z_i = 0$ otherwise. Then $x = y + z$ and it is easy to check that $\|y\|_2 \leq 6\sqrt{t}$, $\|z\|_1 \leq 9t$. 

2
We present now the proof of the above Proposition. Let \( \varphi \) be a bounded measurable function on \((0, +\infty)\), and let \( \psi \) denote the function \( \varphi \circ W \).

Let \( I_0 = \int_0^\infty e^{-\varphi(x)-x} \, dx \) and \( I_1 = \int_0^\infty e^{\psi(y)-y} \, dy \). For \( t \in (0, 1) \), we define \( x(t) \) and \( y(t) \) by the relations

\[
\int_0^{x(t)} e^{-\varphi(x)-x} \, dx = tI_0, \quad \int_0^{y(t)} e^{\psi(y)-y} \, dy = tI_1.
\]

We obtain by differentiation

\[
x'(t) = I_0 e^{\varphi(x(t))+x(t)}, \quad y'(t) = I_1 e^{-\psi(y(t))+y(t)}.
\]

Taking into account the fact that \( \psi(y(t)) \leq \varphi(x(t)) + W(x(t) - y(t)) \), we obtain

\[
y'(t) \geq I_1 e^{-\varphi(x(t))-W(x(t)-y(t))+y(t)}.
\]

Let now \( z(t) = \frac{1}{2}(x(t) + y(t)) - W(x(t) - y(t)) \). We have

\[
z'(t) = \left(1 - 2W'(x-y)\right)I_0 e^{\frac{\varphi(x)}{2}} + \left(1 + 2W'(x-y)\right)I_1 e^{-W(x-y)+y} e^{-\varphi(x)}
\]

\[
\geq \sqrt{1 - 4W'(x-y)^2} \sqrt{I_0 I_1} e^{\frac{\varphi(x+y)}{2}} e^{\frac{1}{2}W(x-y)}
\]

\[
= \sqrt{I_0 I_1} e^{z(t)} \sqrt{1 - 4W'(x-y)^2} e^{\frac{1}{2} W(x-y)}.
\]

We claim that for every \( s \)

\[
(1 - 4W'(s)^2) e^{W(s)} \geq 1.
\]

It will then follow that \( e^{-z(t)} z'(t) \geq \sqrt{I_0 I_1} \), which yields after integrating between 0 and 1

\[
1 \geq \sqrt{I_0 I_1}
\]

and this is our Proposition.

Proof of the claim: We only consider \( s \geq 0 \) since \( W \) is even. For \( s \geq 2 \), \( W' \) is constant and \( W \) increasing, so it is enough to check the case \( 0 \leq s \leq 2 \); this reduces to

\[
e^{-u/18} \leq 1 - 4u/81 \quad \text{for} \quad u \in (0, 4)
\]

which is proved using elementary calculus.

**III. The Gaussian case**

Let \( \gamma \) be the standard Gaussian probability measure on \( R \), with density \( \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \), and \( \gamma_n \) the product of \( n \) copies of \( \gamma \). Throughout this section, the norm will be the Euclidean norm on \( R^n \).

**Theorem 2.** The couple \((\gamma_n, \frac{1}{2} \|x\|^2)\) satisfies \((\tau)\) for every integer \( n \).

Proof: We check first that \((\gamma, x^2/4)\) has property \((\tau)\) on \( R \); the proof is similar to the proof of the Proposition, but simpler: \( x(t) \) and \( y(t) \) are defined in a similar fashion, and \( z(t) \) is simply equal to \( \frac{1}{2}(x(t) + y(t)) \). It follows from Lemma 1 that \((\gamma_n, \frac{1}{2} \|x\|^2)\) has property \((\tau)\) for every integer \( n \).

We can also give a direct proof using the functional Brunn-Minkowski inequality due to Prekopa and Leindler [11], [6] (see also [10]): If \( f, g, h \) are bounded below measurable functions on \( R^n \) such that for all \( x \) and \( u \) we have \( \frac{1}{2} (f(x+u) + g(x-u)) \geq h(x) \), then

\[
\left( \int e^{-f(x)} \, dx \right) \left( \int e^{-g(x)} \, dx \right) \leq \left( \int e^{-h(x)} \, dx \right)^2.
\]

We apply this inequality to \( f(x) = \varphi(x) + \frac{1}{2} \|x\|^2 \), \( g(y) = -\psi(y) + \frac{1}{2} \|y\|^2 \) and \( h(z) = \frac{1}{2} \|z\|^2 \), where we have set \( \psi = \varphi \circ w \), \( w(y) = \frac{1}{2} \|y\|^2 \).
**Remark 1.** The second proof of Theorem 2 only uses the uniform convexity properties of $-\log f$, where $f$ is the density of $\mu$.

**Remark 2.** As pointed out by Talagrand, the Gaussian concentration result is a consequence of Corollary 1, using a suitable map that transforms $\xi_n$ into $\gamma_n$. More precisely, Lemma 3 and Theorem 1 imply that for some $a > 0$, the couple $(\gamma_n, a\|x\|^2)$ satisfies $(\tau)$ for every $n$. However, the proof of Theorem 2 gives a better constant $a$ and is simpler.

We will show now that Theorem 2 allows to recover the main conclusion of the Gaussian concentration result of [9].

**Corollary 2.** Let $\varphi$ be a 1-Lipschitz function on $R^n$, and $X, Y$ two independent $n$-dimensional Gaussian vectors with distribution equal to $\gamma_n$. For every real number $\lambda$ we have

$$Ee^{\frac{\lambda}{\sqrt{2}}(\varphi(X) - \varphi(Y))} \leq e^{\lambda^2/2}. $$

**Remark 3.** This inequality is optimal since there is equality when $\varphi$ is a norm-one linear functional.

Proof: Let $\psi = \frac{\lambda}{\sqrt{2}}w$, where $\varphi$ is 1-Lipschitz on $R^n$, $w(y) = \frac{4}{n}\|y\|^2$ and $\lambda > 0$. It is enough to apply $(\tau)$ and notice that $\psi(x) \geq \frac{\lambda}{\sqrt{2}}(x) - \lambda^2/2$. Let $y$ be such that

$$\psi(x) = \frac{\lambda}{\sqrt{2}}(y) + \frac{1}{4}\|x - y\|^2. $$

Then

$$\psi(x) \geq \frac{\lambda}{\sqrt{2}}(x) - \frac{\lambda}{\sqrt{2}}\|x - y\| + \frac{1}{4}\|x - y\|^2 \geq \frac{\lambda}{\sqrt{2}}(x) + \min\{\frac{1}{4}u^2 - \frac{\lambda}{\sqrt{2}}u; u \in R\} = \frac{\lambda}{\sqrt{2}}(x) - \lambda^2/2. $$

The first part of the next Corollary is known (it is a Poincaré-type inequality due to Chen [2]).

**Corollary 3.** If $\varphi$ is a Lipschitz function on $R^n$, we have

$$\frac{1}{2}\int (\varphi(x) - \varphi(y))^2 d\gamma_n(x)d\gamma_n(y) \leq \int \|\nabla \varphi\|^2 d\gamma_n. $$

More generally, this result holds for every probability measure $\mu$ on $R^n$ such that $(\mu, w)$ satisfies $(\tau)$ for a function $w$ convex and greater than $\frac{4}{n}\|x\|^2$ in a neighborhood of 0.

Proof: Let $u$ be a convex function such that $u \leq w$ and $u(x) = \frac{1}{4}\|x\|^2$ in a neighborhood of 0; assume that $\varphi$ is a compactly supported $C^1$-function. For $t > 0$ consider $\varphi_t = t\varphi$ and $\psi_t = \varphi_t w$. One can check that

$$\lim_{t \to 0} \frac{\psi_t(x) - \varphi_t(x)}{t^2} = -\|\nabla \varphi(x)\|^2 $$

and the result follows easily from the property $(\tau)$ of $(\mu, u)$ applied to $\varphi_t$, when $t \to 0$.

**Remark 4.** If $(\mu, w)$ satisfies $(\tau)$ on $R$, where $\mu$ is such that $\int (x - y)^2 d\mu(x)d\mu(y) > 1$, we can apply Corollary 3 to $\varphi(x) = x$ to conclude that $\{x; w(x) \geq \frac{1}{2}x^2\}$ is not a neighborhood of 0. One can also show that if $(\mu, w)$ satisfies $(\tau)$ on $R$, with $\mu$ symmetric and $\int x^2 d\mu(x) = 1$, then $(\gamma, \frac{1}{2}w''(0)t^2)$ also satisfies $(\tau)$. This shows the necessity of a subquadratic behavior at 0 for the function $w$.

**Corollary 4.** Let $\lambda_n$ denote the uniform probability measure on $[0, 1]^n$. There exists $a > 0$ such that $(\lambda_n, a\|x\|^2)$ satisfies $(\tau)$ for every integer $n$ (one can take $a = \pi/2$).

Proof: Using Lemma 3, this follows from Theorem 2, exactly like in Pisier [9].
IV. Convex property (τ)

Assume that \( w \) is a convex function on some topological vector space \( X \) and that \( \mu \) is a probability measure on \( X \). We say that the couple \( (\mu, w) \) satisfies the convex property (τ) provided

\[
(f \exp(w) d\mu)(f \exp(-w) d\mu) \leq 1
\]

for every convex measurable function \( \varphi \) on \( X \).

**Lemma 5.** If \((\mu_i, w_i)\) satisfies the convex property (τ) on \( X_i \) for \( i = 1, 2 \), then \((\mu_1 \otimes \mu_2, w)\) satisfies the convex property (τ) on \( X_1 \times X_2 \), with

\[
w(x_1, x_2) = w_1(x_1) + w_2(x_2).
\]

Proof: As in the proof of Lemma 1 we consider \( \varphi^\mu(x) = \varphi(x, y) \) for a convex function \( \varphi \) on \( X_1 \times X_2 \); we can apply the convex property (τ) to \( \psi(y) = \log(\int e^{\varphi \otimes w} d\mu_1) \) if we observe that \( \psi \) is a convex function.

We shall say that \( \mu \) has diameter \( \leq 1 \) as a short way to express that \( \mu \) is supported by a set of diameter \( \leq 1 \). The following Theorem is the equivalent in our language of a result of Talagrand [13] and its generalization by Johnson and Schechtman [5].

**Theorem 3.** Let \((X_i)\) be a family of normed spaces; for each \( i \), let \( \mu_i \) be a probability measure with diameter \( \leq 1 \) on \( X_i \), and \( w_i(x) = \frac{1}{4} \|x\|^2 \) for \( x \in X_i \). If \( \mu \) is the product of the family \((\mu_i)\), then \((\mu, w)\) satisfies the convex property (τ), with \( w(x) = \sum_i w_i(x_i) \).

Proof: According to Lemma 5, we only need to prove the result for a single probability measure \( \mu \) with diameter \( \leq 1 \) on a normed space \( X \). Let \( A \) be a set of diameter \( \leq 1 \) that supports \( \mu \), and let \( \varphi \) be a convex function on \( X \); assume without loss of generality that \( \inf \varphi(A) = 0 \). Define \( w(x) = \frac{1}{4} \|x\|^2 \) and \( \psi = \varphi \otimes w \).

Let \( x \in A \), \( \varepsilon > 0 \) and \( a \in A \) such that \( \varphi(a) \leq \varepsilon \). We have, if \( y = (1-\theta)x + \theta a \) and \( 0 \leq \theta \leq 1 \)

\[
\psi(x) \leq \varphi(y) + \frac{1}{4} \|x - y\|^2 \leq (1-\theta)\varphi(x) + \theta \varepsilon + \frac{1}{4} \theta^2.
\]

Choosing an optimal \( \theta \), we deduce from the above that \( \psi(x) \leq k(\varphi(x)) \) where \( k(u) \) is equal to \( u - u^2 \) if \( 0 \leq u \leq \frac{1}{2} \), and to \( \frac{1}{4} \) if \( u \geq \frac{1}{2} \). We claim now that \( e^{k(u)} \leq 2 - e^{-u} \). It follows that

\[
\int e^\psi d\mu \leq 2 - \int e^{-w} d\mu \leq (\int e^{-w} d\mu)^{-1}
\]

and this finishes the proof (the preceding computation was inspired by [5]).

Proof of the claim: For \( 0 \leq u \leq \frac{1}{2} \), we write

\[
\frac{1}{2} (e^{u-u^2} + e^{-u}) = e^{-u^2/2} \cosh(u-u^2/2) \leq e^{-u^2/2} \cosh(u) \leq 1.
\]

**Remark 5.** In the case of the probability \( \beta \) on \([0, 1]\) that gives measure \( \frac{1}{2} \) to \( \{0\} \) and \( \{1\} \), it is easy to improve the function \( w \) from \( \frac{1}{4} x^2 \) to \( \frac{1}{2} x^2 \).

**Corollary 5** [13], [5]. Let \( A \) be a measurable subset of \([0, 1]^n\) and \( B \) its convex hull. For every product probability measure \( \mu \) on \([0, 1]^n\) we have

\[
\int e^{\frac{1}{2} d_B^2} d\mu \leq (\mu(A))^{-1}
\]

where \( d_B \) denotes the Euclidean distance to the set \( B \).
References.
[1] C. Borell, The Brunn-Minkowski inequality in Gauss space, Inventiones Math. 30 (1975), 205–216.
[2] L. Chen, An inequality for the multivariate normal distribution, J. Multivariate Anal. 12 (1982) 306–315.
[3] A. Ehrhard, Symérisation dans l’espace de Gauss, Math. Scand. 53 (1983) 281–301.
[4] T. Figiel, J. Lindenstrauss, V. Milman, The dimension of almost spherical sections of convex bodies, Acta Math. 139 (1977) 53–94.
[5] W. Johnson, G. Schechtman, Remarks on Talagrand’s deviation inequality for Rademacher functions, Texas Functional Analysis Seminar 1988–89.
[6] L. Leindler, On a certain converse of Hölder’s inequality, Acta Sci. Math. 33 (1972) 217–223.
[7] V. Milman, A new proof of the theorem of A. Dvoretzky on sections of convex bodies, Func. Anal. Appl. 5 (1971) 28–37.
[8] V. Milman, G. Schechtman, Asymptotic theory of finite dimensional normed spaces, Springer Lecture Notes in Math. 1200 (1986).
[9] G. Pisier, Probabilistic methods in the geometry of Banach spaces, CIME Varenna 1985, Springer Lecture Notes in Math. 1206, 167–241.
[10] G. Pisier, Volume of convex bodies and Banach spaces geometry, Cambridge University Press.
[11] A. Prekopa, On logarithmically concave measures and functions, Acta Sci. Math. 34 (1973) 335–343.
[12] M. Talagrand, to appear.
[13] M. Talagrand, An isoperimetric theorem on the cube and the Khintchine-Kahane inequalities, Proc. Amer. Math. Soc. 104 (1988) 905–909.

Université Paris 7, U.F.R. de Mathématiques