DEVIATIONS LOWER BOUNDS AND CONDITIONAL PRINCIPLES.

PATRICK CATTIAUX AND NATHAEL GOZLAN

Ecole Polytechnique and Université Paris X

Abstract. The aim of this paper is to obtain non asymptotic bounds (mainly lower bounds) for the Probability of rare events in the Sanov theorem. These bounds are used to study the asymptotics in conditional limit theorems (Gibbs conditioning principle).

1. Introduction

Let $X_1, X_2, ...$ be i.i.d. random variables taking their values in some metrizable space (E, d). Set $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ the empirical mean (assuming here that E is a vector space) and $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ the empirical measure. In recent years new efforts have been made in order to understand the asymptotic behaviour of laws conditioned by some rare event.

The celebrated Gibbs conditioning principle is the corresponding meta principle for the empirical measure, namely

$$\lim_{n\to+\infty} \mathbb{P}^{\otimes n}((X_1, ..., X_k) \in B \,/\, L_n \in A) = (\nu_0)^{\otimes k}(B)\,,$$

where ν_0 minimizes the relative entropy $H(\nu_0|\mu)$ among the elements in A. When A is thin (i.e. $\mathbb{P}^{\otimes n}(L_n \in A) = 0$), such a statement is meaningless, so one can either try to look at regular desintegration (the so called "thin shell" case) or look at some enlargement of A. The first idea is also meaningless in general (see however O.Johnson ([17]) that extends previous work by Diaconis and Freedman). Therefore we shall focus on the second one.

An enlargement A_ε is then a non thin set containing A, and the previous statement becomes a double limit one i.e.

$$\lim_{\varepsilon\to 0}\ \lim_{n\to+\infty}\,.$$

Precise hypotheses are known for this meta principle ("thick shell" case) to become a rigorous result, and refinements (namely one can choose some increasing $k(n)$) are known (see e.g. [11] and the references therein). One possible way to prove this result is to identify relative entropy with the rate function in the Large Deviation Principle for empirical measures (Sanov's theorem).

In this paper we will introduce an intermediate "approximate thin shell" case, i.e. we will look at the case when the enlargement size depends on n, i.e. $\varepsilon_n \to 0$.

When A is some closed hyperplane (i.e. defined thanks to one linear constraint), such a program can be carried out by directly using well known results (see [12] Theorem 3.7.4, Exercise 3.7.10 and Corollary 7.3.34).

Date: June 14, 2002.

As soon as one considers conditioning, it clearly appears that it is important to get exact bounds for the probabilities of rare events. Furthermore, if one wants to improve Gibbs conditioning principle by considering only one limit (i.e. $\varepsilon = \varepsilon_n$), these bounds have to be as precise as possible. Our aim in this paper is to obtain such exact bounds (here by exact we mean non asymptotic ones) and to explore their use for conditioning results.

Exact upper bounds are known for general compact convex sets and in some related situations (see [12] Exercises 4.5.5 or 6.2.19 and [21]). We shall recall these results. In order to get exact lower bounds we will simply use this upper bound for the tilted probability and plug it into the classical proof of Sanov's theorem.

This method certainly does not furnish sharp estimates and will certainly appear as a stupid one in the classical cases. For the empirical mean, exact lower bounds are already known in some cases. Einmahl and Kuelbs (see [14] and [19]) obtained precise estimates (see (1.10) or (1.11) in [14], and Remark 1 p.1264 in [19]). But the constants appearing in their result are difficult to describe. The same holds with the full asymptotic expansion obtained by Iltis ([16]). Our rough method if it furnishes weaker asymptotic results, is more tractable for the control of the constants. In addition the geometric arguments used for the empirical mean are difficult to extend to the empirical measure (however look at [10] where such an extension is partly done), while the rough approach furnishes at least some result.

The second reason for trying to get as explicit bounds as possible is that we want to study super thin sets (i.e. sets on which the rate function is infinite). Such a situation appears in statistical mechanics for infinite systems, where the relative entropy has to be replaced by the specific entropy and the large deviations speed is no more n. It also appears in others problems like calibration of diffusion processes. These parts of our program however will not be discussed here.

What we tried to do in this paper is to push forward as far as possible the rough technology for elementary examples, and to compare it with more refined and clever arguments.

Before to describe the contents of the paper let us mention another conditional limit theorem, namely Nummelin's theorem. Roughly speaking, Nummelin's conditional weak law of large numbers is a statement like

$$\lim_{n\to+\infty} \mathbb{P}^{\otimes n}(d(M_n, a) < \varepsilon \,/\, M_n \in D) = 1\,,$$

where D is some rare set (i.e. does not contain the mean m of the $X_i's$) and a is some point in the closure of D that minimizes the rate function of the Large Deviations Principle. For convex open sets D of a Banach space E, a precise definition of a (the dominating point) has been introduced and studied by P.Ney, U.Einmahl and J.Kuelbs, and the previous meta principle (as well as some refinements) is proved (see e.g. [20] and the references therein).

In some important cases both Nummelin's law and the Gibbs conditioning principle are intimately connected. We shall also discuss these points.

The paper is organized as follows.

Section 2 deals with Gibbs conditioning principle for compact state spaces, in the "approximate thin shell" case. Our presentation is based on the classical example of an A defined through a finite number of linear constraints (for which controls via Sanov upper bound are clearly almost the worst ones) , in order to make the comparison easier with section 4. However a general statement is given in Theorem 2.24.

Convergence rate is precisely stated thanks to a non asymptotic Sanov lower bound. In section 3 we see how to extend these results when $E = \mathbb{R}^d$. In both sections we did not try to get the most general framework. Let us say that the exact Sanov lower bound obtained in [10] (that holds for n large enough) looks better than the one we are obtaining, but we do not know how to get precise controls for this bound for all n.

In section 4 we improve these results in the case when A is described through a finite number of linear constraints. Indeed in this case one can use concentration inequalities and (or) the arguments in the final section of [20]. They allow to skip the annoying covering number in the lower estimate, and this time provides a real improvement.

Examples of infinitely many constraints are studied in sections 5 and 6 when E is the space of continuous functions. The conditioning set A is defined by fixing some (or all) time marginal laws of the corresponding process. Such a problem is motivated by a statistical approach of Schrödinger bridges or Nelson processes. A short account of this physical interpretation is done at the end of the section. The reduction to the compact case involves some approximation results, while the explicit bound requires some calculations on metric entropy that are familiar to specialists in the theory of empirical processes.

Let us emphasize that in these last cases both the C.L.T. approach or the concentration results seem to be useless. This is essentially due to the fact that the conditioning set is defined through an infinite number of linear constraints. On one hand as we already said this turns the geometric description of dominating point delicate. On the other hand concentration results for infinite families are known but involve the calculation of the mean of some supremum that is not immediate. Actually some estimates are known (for instance for Donsker classes, see [13]) but also involve estimates for covering (or bracketing) numbers. A short discussion is done in section 7.

Though it is our main motivation, the case of "super thin" sets (i.e. sets for which relative entropy is infinite), in connection with statistical mechanics, or with calibration problems will be discussed elsewhere. We just discuss a purely academic example of such super thin set in connection with the usual Brownian bridge in the final section 9.

Let us just add one thing. One can of course ask the following question: does the Gibbs conditioning (in the approximate thin shell case) hold for any nested (convex) closed sets ? (This is of course true if one can prove some uniform version of the thin sell case.)

We did not find any counter example to this statement. Actually the limitations in our results are mainly (not only) due to the fact that we have to check that $\mathbb{P}^{\otimes n}(L_n \in A_{\varepsilon_n}) > 0$. If true the statement above will make our results useless except for one point: we can get some estimates for the rate of convergence in the Gibbs principle.

Let us recall the basic upper bound we shall use repeatedly.

Lemma 1.1. *Let $\mathcal{X}$ be a Hausdorff topological vector space with topological dual space $\mathcal{X}^*$. μ_n is a sequence of $\mathbb{M}_1(\mathcal{X})$ the set of Probability measures on $\mathcal{X}$, with logarithmic moment generating function denoted by Λ_n (i.e. for $\lambda \in \mathcal{X}^*$,*

$$\Lambda_n(\lambda) = \log \int_{\mathcal{X}} e^{<\lambda,x>}\, \mu_n(dx)\,)\,.$$

Λ_n^* *denotes the Fenchel-Legendre transform of the renormalized* Λ_n*, i.e.*

$$\Lambda_n^*(x) = \sup_{\lambda \in \mathcal{X}^*} \{< \lambda, x > - \frac{1}{n} \Lambda_n(n\,\lambda)\} .$$

Then for any convex compact subset K *of* $\mathcal{X}$ *and any* n *the following upper bound holds*

$$\mu_n(K) \leq \exp\big(- n \inf_{x \in K} \Lambda_n^*(x)\big) .$$

The proof lies on some version of Sion's min-max theorem. That is why compactness is required (see [12] Exercise 4.5.5).

Acknowledgements. We want to warmly acknowledge Christian Léonard for so many animated conversations on Large Deviations Problems, and for indicating to us various references on the topic.

2. Gibbs conditioning principle : the compact case.

In this section (E, d) will be a Polish space, and in order to simplify some arguments below, we shall assume in addition that it is a compact space. $M_1(E)$ (resp. $M(E)$) will denote the set of Probability measures (resp. bounded signed measures) on E equipped with its Borel σ-field. $M_1(E)$ is equipped with its weak topology and its natural Borel σ-field.
If $X_1, X_2, ...$ is a sequence of i.i.d. E valued random variables, we denote their empirical measure by $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. If $\alpha \in M_1(E)$ is the common law of the X_i's, Q_n^α will denote the law of L_n (i.e. $Q_n^\alpha \in M_1(M_1(E))$).
For $\alpha \in M_1(E)$ we denote by $E_n^\alpha(f) = \int f(\gamma)\, Q_n^\alpha(d\gamma)$ the expectation of any nonnegative measurable f defined on $M_1(E)$, and by

$$\Lambda_\alpha(\varphi) = \log\, \big(\int_E \exp(\varphi(x))\, \alpha(dx) ,$$

the logarithmic moment generating function defined for $\varphi \in C(E)$ the set of continuous functions.

Our aim is to study the asymptotic behaviour of the conditional law

$$\alpha^n_{A,k}(B) = \mathbb{P}_\alpha^{\otimes n}\Big((X_1, ..., X_k) \in B \,/\, L_n \in A_n\Big) \tag{2.1}$$

for some A_n going to some thin set A when n goes to ∞. In this section A will be a general convex set and we shall make the estimates as precise as possible for further extension. For the sake of simplicity again, k is fixed (i.e. does not depend on n). In order to understand the statement of our first Theorem we need more definitions.

The first tool we need is relative entropy. Recall that for α and γ in $M_1(E)$, the relative entropy $H(\alpha|\gamma)$ is defined by the two equivalent formulas

(2.2.1) $H(\alpha|\gamma) = \int \log\big(\frac{d\alpha}{d\gamma}\big)\, d\alpha$, if this quantity is well defined and finite, $+\infty$ otherwise,
(2.2.2) $H(\alpha|\gamma) = \sup\{< \alpha, \varphi > - \Lambda_\gamma(\varphi)\, , \, \varphi \in C_b(E)\}$.

If B is a measurable set of $M_1(E)$ we will write

$$H(B|\gamma) = \inf\{H(\alpha|\gamma)\,,\ \alpha \in B\}\,. \tag{2.3}$$

Lemma 1.1 can be used here with $\mathcal{X} = M(E)$. Since E is assumed to be compact, so does $M_1(E)$ and we thus have the following upper bound

Lemma 2.4. *For any convex closed $K \subseteq M_1(E)$,*

$$Q_n^\alpha(K) \ \leq\ \exp(-\,n\,H(K|\alpha))\,.$$

To draw the convexity assumption we have to introduce some distance on $M_1(E)$. We will choose the Fortet-Mourier distance (while [21] and [12] Exercise 6.2.19 are using respectively the Prohorov and the Lévy distances). Recall that the Fortet-Mourier distance β is defined as

$$\beta(\alpha,\gamma) = \sup\{|<\alpha-\gamma, g>|\,,\ g \in Lip(E)\,,\ \| g \|_{Lip} \leq 1\}\,, \tag{2.5}$$

where Lip is the set of Lipschitz functions on E equipped with its usual norm. Recall that relative entropy and distance are related by the well known Pinsker inequality

$$\beta(\alpha,\gamma)\ \leq \| \alpha-\gamma \|_{TV} \leq\ \sqrt{2\,H(\alpha|\gamma)}\,. \tag{2.6}$$

$(M_1(E),\beta)$ is then a compact metric space (whose topology is the weak topology), and all metric notions below (like balls) are understood to take place with this metric. In particular, closed balls are convex and compact. We may thus introduce the ξ metric entropy of any set. We shall use the following notation

(2.7.1) $m(A,\xi,d)$ will denote the ξ covering number of a subset A of E, $(m(\xi) = m(E,\xi,d))$, i.e. the minimal number of open balls with radius ξ needed to cover A ,

(2.7.2) $m(B,\xi,\beta)$ will denote the ξ covering number of a subset B of $M_1(E)$,

$$m_\xi = m(M_1(E),\xi,\beta)\,.$$

One can thus easily obtain a general upper bound

Lemma 2.8. Exact Sanov Upper Bound.

For any measurable $B \subseteq M_1(E)$,

$$Q_n^\alpha(B)\ \leq\ m_\xi\,\exp(-\,n\,H(B^\xi|\alpha))$$

where B^ξ denotes the closed ξ blowup of B.

We will now state and prove some result on an exact lower bound.

Proposition 2.9. Exact Sanov Lower Bound.

Let $\mu \in M_1(E)$. Then for all open subset G of $M_1(E)$, and all $\nu \in G$ such that $H(\nu|\mu) < +\infty$ and $\frac{d\nu}{d\mu} = f > 0$ the following properties hold with $h = \log(f)$:

1. $\mathbb{Q}_n^\mu(G) = \exp\left(-\,n\,H(\nu|\mu)\right) E_n^\nu[\mathbb{1}_G\,\exp(-n < . - \nu, h >)]$,
2. *if f is continuous, then for all $\varepsilon > 0$ one can find $\eta > 0$ such that for all $\eta > \xi > 0$*

$$\mathbb{Q}_n^\mu(G) \geq\ \exp\ \big(-\,n\,(H(\nu|\mu)+\varepsilon)\big)\Big(1 - m_\xi\,\exp\ \big(-\ n\,H(B^c(\nu,\eta-\xi)|\nu)\big)\Big)\,,$$

3. *if h is Lipschitz, for all $\eta < \beta(\nu, A^c)$ and all $\xi < \eta$,*

$$\mathbb{Q}_n^\mu(G) \geq\ \exp\ \big(-\,n\,(H(\nu|\mu)+\eta\ \| h \|_{Lip})\big)\Big(1 - m_\xi\,\exp\ \big(-\ n\,H(B^c(\nu,\eta-\xi)|\nu)\big)\Big)\,.$$

Proof. Denote by $L_n^x = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Then

$$\begin{aligned} Q_n^\mu(G) &= \int 1\!\!1_G(L_n^x)\mu(dx_1)...\mu(dx_n) \\ &= \int 1\!\!1_G(L_n^x)\frac{1}{f(x_1)}...\frac{1}{f(x_n)}\nu(dx_1)...\nu(dx_n) \\ &= \int 1\!\!1_G(L_n^x)\exp(-n < L_n^x, h >)\nu(dx_1)...\nu(dx_n) \\ &= \exp(-nH(\nu|\mu))\int 1\!\!1_G(L_n^x)\exp(-n < L_n^x - \nu, h >)\nu(dx_1)...\nu(dx_n) \\ &= \exp(-nH(\nu|\mu))E_n^\nu[1\!\!1_G\exp(-n < . - \nu, h >)] \end{aligned}$$

and 1. is proved.
If f is continuous and positive, it is bounded from below (since E is compact). Hence h is continuous and bounded. Thus the mapping :

$$M_1(E) \to \mathbb{R} : \gamma \mapsto < \gamma, h >$$

is continuous and for all $\varepsilon > 0$, one can find $\eta > 0$ such that

$$\beta(\alpha, \nu) < \eta \Rightarrow \alpha \in G \text{ and } | < \alpha - \nu, h > | < \varepsilon \, .$$

Applying 1. we get

$$\begin{aligned} Q_n^\mu(G) &= \exp(-nH(\nu|\mu))\int 1\!\!1_G(L_n^x)\exp(-n < L_n^x - \nu, h >)\nu(dx_1)...\nu(dx_n) \\ &\geq \exp(-nH(\nu|\mu))\int 1\!\!1_{B(\nu,\eta)}(L_n^x)\exp(-n < L_n^x - \nu, h >)\nu(dx_1)...\nu(dx_n) \\ &\geq \exp\big(-n\,(H(\nu|\mu)+\epsilon)\big)\,Q_n^\nu(B(\nu,\eta)) \\ &\geq \exp\big(-n\,(H(\nu|\mu)+\epsilon)\big)\,(1 - Q_n^\nu(B^c(\nu,\eta)))\, . \end{aligned}$$

The upper bound in Lemma 2.8 furnishes for all $\xi > 0$

$$Q_n^\nu(B^c(\nu,\eta)) \leq m_\xi \exp(-n\,H((B^c(\nu,\eta))^\xi|\nu))\, ,$$

and for $\xi < \eta$, $(B^c(\nu,\eta))^\xi = B^c(\nu, \eta - \xi)$. 2. is thus proved.
If $\eta < \beta(\nu, G^c)$, then $B(\nu,\eta) \subset G$. Furthermore $| < \gamma - \nu, h > | \leq \beta(\nu,\gamma) \parallel h \parallel_{Lip}$ so that we obtain 3. as above. □

Of course $\bar{G}$ is compact, hence if G is convex, $\nu \mapsto H(\nu|\mu)$ achieves its minimum on $\bar{G}$ at some ν_G. One can then approximate the corresponding density f_G by some f satisfying the hypotheses in Proposition 2.9. Controls on ε in 2.9.2 are possible in terms of the relative entropy $H(\bar{G}|\mu) = H(\nu_G|\mu)$. More delicate are controls of the Lipschitz norm in full generality. Instead of going further in this direction, we shall apply the previous results to the Gibbs conditioning principle (2.1).

We shall start with particular A's. These A's (energy levels of regular potentials) are of course one class of examples, and what follows can be done with much more general thin sets A (see Theorem 2.24) . Furthermore for this particular class, we shall see in section 4 than the results below can be greatly improved. However the method below (and some intermediate results) is used in sections 5 and 6 in cases where the ideas of section 4 seem to be useless.

Let us consider a finite number of Lipschitz functions $U_1, ..., U_j$ and define

$$A = \{\alpha \in M_1(E)\,,\, <\alpha, U_i>= a_i \text{ for all } i = 1, ..., j\}\,. \tag{2.10}$$

Without loss of generality, just modifying the U_i's we may and will assume that any U_i is non negative and that $a_i = 1$ for all $i = 1, ..., j$. We shall assume that

$$\text{there exists } \nu \in A \text{ such that } H(\nu|\mu) < +\infty \text{ and } \nu \sim \mu\,. \tag{2.11}$$

It is then well known that there exists an unique $\nu_0 \in A$ that minimizes relative entropy with respect to μ, and that ν_0 has the following form

$$\frac{d\nu_0}{d\mu} = f_0 = \exp\big(- \sum_{i=1}^{j} c_i U_i \, - \, \Lambda_\mu(- \sum_{i=1}^{j} c_i U_i)\big)\,, \tag{2.12}$$

for some c_i, $i = 1, ..., j$.
When $j = 1$ a sufficient condition for (2.11) to hold is (see e.g. [12] Lemma 7.3.6)

$$\mu(U > 1) > 0 \quad \text{ and } \quad \mu(U < 1) > 0\,. \tag{2.13}$$

One can formulate similar conditions when $j > 1$, but in general (2.11) is naturally satisfied. Consider now the enlarged

$$A_\varepsilon = \{\alpha \in M_1(E)\,,\, | <\alpha, U_i> -1| \leq \varepsilon \text{ for all } i = 1, ..., j\}\,. \tag{2.14}$$

If (2.11) holds, then $H(A_\varepsilon|\mu) < +\infty$ and since A_ε is closed (hence compact) the minimum is achieved at some ν_ε. The same argument as before tells that

$$\frac{d\nu_\varepsilon}{d\mu} = f_\varepsilon = \exp\big(- \sum_{i=1}^{j} c_{i\,\varepsilon} U_i \, - \, \Lambda_\mu(- \sum_{i=1}^{j} c_{i\,\varepsilon} U_i)\big)\,. \tag{2.15}$$

We shall simply write (recall (2.1))

$$\mu^n_{\varepsilon,k} = \mu^n_{A_\varepsilon,k}\,. \tag{2.16}$$

The first result of this section is the following Theorem

Theorem 2.17. *Assume that (2.11) holds, and that the $(U_i - 1)$'s are linearly independent in $\mathbb{L}^1(\mu)$. Define $C(A) = \min_{i=1,...,j} \frac{1}{\|U_i - 1\|_{Lip}}$. Then for all k,*

$$H(\mu^n_{\varepsilon_n,k}|\nu_0^{\otimes k}) \to 0$$

when n goes to ∞, provided ε_n goes to 0 and satisfies

$$m_{\frac{C(A)\varepsilon_n}{2}} \, e^{-\frac{n}{8}\,(C(A))^2\,\varepsilon_n^2} \to 0\,.$$

This result is an improvement of the classical Gibbs conditioning principle since we replace the double limit (first in n, then in ε see e.g. [12] Corollary 7.3.5) by a single one. Also notice that we have convergence in entropy which is stronger than weak convergence. Such a convergence (again with a fixed ε) lies on arguments due to I.Csiszar ([9] and [8]) recalled and used in [12] section 7.3.3. (adapted from [11]) in order to study the refined Gibbs conditioning principle (i.e. with $k = k(n)$). Contrary to what one can think, Theorem 2.17 is not a kind of "dual" version of the refined principle. We shall come back to this point later.

Of course the condition on ε_n in Theorem 2.17 looks difficult to check. Therefore, before to proceed with the proof of the Theorem we will state two lemmas that help to better understand this condition. The first one is concerned with covering number on the space of Probability measures and is adapted from [21]. The second one gives a more explicit description of ε_n for polynomial controls of the covering number on E (other description are of course possible for others controls)

Lemma 2.18. *For all $\xi > 0$,*

$$m_\xi \,\leq\, (\frac{16e}{\xi})^{m(\frac{\xi}{4})}\,.$$

Lemma 2.19. *If $m(\xi) \,\leq\, c\,\xi^{-d}$ then ε_n in Theorem 2.17 has to be such that*

$$\frac{n\,\varepsilon_n^{d+2}}{|\log(\varepsilon_n)|} \,\to\, +\infty\,.$$

In particular one may choose $\varepsilon_n = n^{-a}$ with $a < \frac{1}{d+2}$.

Let us proceed with the proof of the Theorem.

Proof. of Theorem 2.17
Since $\nu_0 \in A$ it is immediate that $B(\nu_0, C(A)\varepsilon) \subset A_\varepsilon$. Hence if

$$C = C(A)\;\parallel \log(f_0)\parallel_{Lip}$$

(recall (2.12)), Proposition 2.9.3 (applied to the interior of A_ε) yields for all $0 < \xi < C(A)\varepsilon$

$$Q_n^\mu(A_\varepsilon) \geq \exp\big(-n\,(H(\nu_0|\mu) + C\varepsilon)\big)\,\Big(1 - m_\xi \exp\big(-n\,H(B^c(\nu_0, C(A)\varepsilon - \xi)|\nu_0)\big)\Big)\,.$$

According to Pinsker's inequality

$$H(B^c(\nu_0, C(A)\varepsilon - \xi)|\nu_0) \geq \frac{1}{2}(C(A)\varepsilon - \xi)^2\,.$$

Hence if we choose $\xi = \frac{C(A)\varepsilon}{2}$ we obtain

$$Q_n^\mu(A_\varepsilon)\,\exp\big(n\,(H(\nu_0|\mu)\big) \geq \exp\big(-n\,C\varepsilon\big)\,\Big(1 - m_{\frac{C(A)\varepsilon}{2}}\exp\big(-\frac{n}{8}\,(C(A))^2\,\varepsilon^2\big)\Big)\,. \tag{2.20}$$

The condition on ε_n in the statement of the Theorem is thus nothing else but a sufficient condition ensuring that the constant in (2.20) is well behaved. In particular for n large enough the lower bound in (2.20) is strictly positive. We may now apply Theorem 7.3.21 in [12] to the set A_{ε_n}. It yields

$$\begin{aligned} H(\mu^n_{\varepsilon_n,k}|\nu_{\varepsilon_n}^{\otimes k}) \;&\leq\; \frac{-1}{[\frac{n}{k}]}\log\Big(Q_n^\mu(A_{\varepsilon_n})\exp(n\,H(A_{\varepsilon_n}|\mu))\Big) \\ &\leq\; \frac{-k}{n}\log\Big(Q_n^\mu(A_{\varepsilon_n})\exp(nH(A_0|\mu))\Big) + k\big(H(A_0|\mu) - H(A_{\varepsilon_n}|\mu)\big) \\ &\leq\; kC\varepsilon_n - \frac{k}{n}\log(1 - \exp(h_n)) + k\big(H(A_0|\mu) - H(A_{\varepsilon_n}|\mu)\big)\,, \end{aligned} \tag{2.21}$$

where $h_n = -\frac{n}{8}(C(A))^2\,\varepsilon_n^2 + \log(m_{\frac{C(A)\varepsilon_n}{2}})$.

But the sequence $H(A_{\varepsilon_n}|\mu) = H(\nu_{\varepsilon_n}|\mu)$ is non decreasing dominated by $H(A|\mu) = H(\nu_0|\mu)$, hence convergent. Recall Csiszar's inequality (see e.g. [12] Lemma 7.3.27)

$$H(\nu_{\varepsilon_p}|\mu) \geq H(\nu_{\varepsilon_p}|\nu_{\varepsilon_n}) + H(\nu_{\varepsilon_n}|\mu)\,, \tag{2.22}$$

that holds for all $p \geq n$ thanks to the minimality of relative entropy. Applying once again Pinsker's inequality, (2.22) shows that ν_{ε_n} is a Cauchy sequence for the Fortet-Mourier distance hence is convergent to some α. It is immediate that $\alpha \in A$ and

$$H(\nu_0|\mu) \,\geq\, \lim_n H(\nu_{\varepsilon_n}|\mu) \,\geq\, H(\alpha|\mu) \,\geq\, H(\nu_0|\mu)$$

inequalities being respectively consequences of monotonicity, lower semi continuity and minimality. Hence $\alpha = \nu_0$. We thus have shown that $H(A_{\varepsilon_n}|\mu) \to H(A_0|\mu)$ and ν_{ε_n} weakly converges to ν_0. Actually, replacing p by 0 in (2.22) we get the stronger

$$H(\nu_0|\nu_{\varepsilon_n}) \to 0\,.$$

Plugging these results into (2.21), we get

$$H(\mu^n_{\varepsilon_n,k}|\nu^{\otimes k}_{\varepsilon_n}) \to 0\,,$$

and thanks to Pinsker and the triangle inequalities

$$\beta(\mu^n_{\varepsilon_n,k}|\nu^{\otimes k}_0) \to 0\,.$$

In order to get the convergence in relative entropy, recall the classical entropy decomposition

$$\begin{aligned} H(\mu^n_{\varepsilon_n,k}|\nu^{\otimes k}_{\varepsilon_n}) &= H(\mu^n_{\varepsilon_n,k}|\nu^{\otimes k}_0) + k \int \log(\frac{d\nu_0}{d\nu_{\varepsilon_n}}) d\mu^n_{\varepsilon_n,1} \\ &= H(\mu^n_{\varepsilon_n,k}|\nu^{\otimes k}_0) + k\, H(\nu_0|\nu_{\varepsilon_n}) + k \int \log(\frac{d\nu_0}{d\nu_{\varepsilon_n}}) \Big(d\mu^n_{\varepsilon_n,1} - d\nu_0\Big)\,. \end{aligned}$$

According to what precedes, we only need to prove that the integral term goes to 0. But since

$$\beta(\mu^n_{\varepsilon_n,1}|\nu_0) \to 0\,,$$

it is enough to get an uniform bound (in n) for the Lipschitz norm of $\log(\frac{d\nu_0}{d\nu_{\varepsilon_n}})$, i.e. (recall (2.12) and (2.15)) of the $c_{i\,\varepsilon_n}$'s.
In order to prove the latter we proceed by contradiction. First

$$\log(\frac{d\nu_0}{d\nu_{\varepsilon_n}}) = \sum_{i=1}^{j} \theta^i_n\, U_i + d_n\,,$$

and

$$H(\nu_0|\nu_{\varepsilon_n}) = \sum_{i=1}^{j} \theta^i_n + d_n\,.$$

Hence $\lim_{n\to+\infty} (\sum_{i=1}^j \theta^i_n + d_n) = 0$ and any subsequence of ν_{ε_n} contains a subsequence such that

$$\sum_{i=1}^{j} \theta^i_n\, U_i + d_n \,\to\, 0 \quad , \quad \nu_0\, a.s.$$

It follows that $\sum_{i=1}^j \theta^i_n\,(U_i - 1) \,\to\, 0$ ν_0 a.s. and if the θ^i_n's are not bounded, dividing by the leading term, we will be able to build some linear combination of the $(U_i - 1)$'s that vanishes ν_0 a.s. Since ν_0 and μ are equivalent, this will contradict the assumption of linear independence. $\square$

Remark 2.23. The above proof shows that we can choose $k = k(n)$ provided in addition

$$k(n)\,\varepsilon_n \to 0 \quad \text{and} \quad k(n)\Big(H(A_0|\mu) - H(A_{\varepsilon_n}|\mu)\Big) \to 0\,.$$

In this case we get weak convergence. Recall that

$$H(\nu_0^{\otimes k(n)}|\nu_{\varepsilon_n}^{\otimes k(n)}) = k(n) H(\nu_0|\nu_{\varepsilon_n})\,,$$

and that we have shown during the proof that

$$H(\nu_0|\nu_{\varepsilon_n}) \leq H(A_0|\mu) - H(A_{\varepsilon_n}|\mu)\,.$$

Convergence in entropy is more delicate. But if we restrict the enlargement to one side (for example $a_i \leq <\alpha, U_i> \leq a_i + \varepsilon_n$) in such a way that $\nu_0 = \nu_{\varepsilon_n}$ (as in the case studied in [12] Corollary 7.3.34) the situation is much simplified.

Actually the method of proof is very close to the arguments in [11] in particular Proposition 2.8 and formula (2.10) therein. The main (and may be only) difference is that we get a (bad but) exact lower bound that allows to make ε varying with n. See section 4 for some improvements.

Also remark that (2.20) is still true if we replace A_ε by any bigger subset. In particular, for a fixed ε, and any sequence ε_n as in Theorem 2.17, this yields

$$H(\mu^n_{\varepsilon,k}|\nu_\varepsilon^{\otimes k}) \leq kC\varepsilon_n - \frac{k}{n}\log(1-\exp(h_n)) + k\big(H(A_0|\mu) - H(A_\varepsilon|\mu)\big)\,.$$

Taking first the lim sup in n, then the limit in ε (using the same arguments as the ones following (2.21)) we thus recover the classical double limit formulation of the Gibbs conditioning principle.

It is also possible to deal with general A considering a closed ε_n blowup of A. The proof of Theorem 2.17 immediately extends to the following

Theorem 2.24. *Let A be a closed convex set such that there exists $\nu \in A$ such that $H(\nu|\mu) < +\infty$. For $\varepsilon_n > 0$ let A_{ε_n} be the closed ε_n blowup of A. Assume in addition that the entropy minimizer $\nu_0 \in A$ has a Lipschitz positive density. Then the statement of Theorem 2.17 is still true with $C(A) = 1$.*

In particular, provided (2.11) holds and $C(A)$ is well defined, one expects to extend Theorem 2.17 to an infinite enumerable number of U_i's. This will require however additional technicalities that will be explained later.

Remark 2.25. In [10], A.Dembo and J.Kuelbs are studying the (refined) Gibbs conditioning principle in a more general framework (with Banach valued U_i's) for a non compact E. The proof of their Theorem 1 lies on an asymptotic Sanov lower bound replacing (2.20) i.e. (see Proposition 3 in [10])

$$Q_n^\mu(A_\varepsilon)\,\exp\big(n\,(H(\nu_0|\mu)\big) \geq \exp\big(-\,C_1(\varepsilon)\,n^{\frac{1}{2}}\big)$$

that holds under appropriate conditions, for n large enough. This asymptotic bound is of course much better than (2.20). However, when ε varies with n, we are not expert enough in

C.L.T. theory to get an explicit form (or an explicit control) for C_1. But as for the empirical mean, a careful study of this constant will certainly yield better results than ours.
Also notice that the final section in [20] contains some results on Gibbs conditioning in some different approximate thin shell case, for some very particular functions ($U(x) = x$). These results are obtained by combining the ones for Nummelin's law and the ones in [10]. We shall see in section 4 that they can be used in our framework.

Now we turn to the proofs of the Lemmas. Lemma (2.19) is almost immediate and left to the reader.

Proof. of Lemma 2.18
Let $\xi > 0$ and $p = m(E, \xi, d)$.
$B_1, B_2, ..., B_p$, are p balls with radius ξ covering E.
For $i = 1...p$ denote by $A_1 = B_1$ and define by induction

$$A_i = B_i - (A_1 \cup ... \cup A_{i-1}) .$$

We may assume that the A_i's are non empty.
Pick some x_i in each A_i, denote by $\delta_i = \delta_{x_i}$. For all n define

$$\mathcal{Y}_n = \{\alpha \in M_1(E) : \alpha = a_1\delta_1 + ... + a_p\delta_p, a_i \in \{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}\} .$$

It is easily seen that $Card(\mathcal{Y}_n) = C^{p-1}_{n+p-1}$. Since $n! > e(\frac{n-1}{e})^{n-1}$, it yields for $p \geq 2$ and $n \geq p$:

$$\begin{aligned} C^{p-1}_{n+p-1} &= \frac{(n+p-1)...(n+1)}{(p-1)!} \leq \frac{(n+p-1)^{p-1}}{(p-1)!} \\ &< \frac{(n+p-1)^{p-1}}{e(\frac{p-1}{e})^{p-1}} = e^{p-2}(\frac{n}{p} + \frac{p-1}{p})^{p-1}(\frac{p}{p-1})^{p-1} \\ &< e^{p-2}(2\frac{n}{p})^{p-1}(2)^{p-1} \leq (\frac{4en}{p})^p , \end{aligned}$$

that is still true for $p = 2$. Hence

$$Card(\mathcal{Y}_n) \leq (\frac{4en}{p})^p .$$

For $\gamma \in M_1(\mathbb{E})$ and all $i = 1...p-1$, there exists an unique $a_i \in \{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}$ such that

$$a_i \leq \gamma(A_i) \leq a_i + \frac{1}{n} .$$

Define

$$a_p = 1 - (a_1 + ... + a_{p-1}) \text{ and } \alpha = a_1\delta_1 + ... + a_p\delta_p .$$

Let g be a Lipschitz function with Lipschitz norm less than 1. It holds

$$|\int g(x)\gamma(dx) - \int g(x)\alpha(dx)| =$$

$$\begin{aligned}
&= |\sum_{i=1}^{p-1}\int_{A_i} g(x)\gamma(dx) - \int_{A_i} g(x)\alpha(dx) + \int_{A_p} g(x)\gamma(dx) - \int_{A_p} g(x)\alpha(dx)| \\
&= |\sum_{i=1}^{p-1}\int_{A_i} (g(x)-g(x_i))\gamma(dx) + g(x_i)(\gamma(A_i)-a_i) + \\
&\qquad + \int_{A_p} (g(x)-g(x_p))\gamma(dx) + g(x_p)(\gamma(A_p)-a_p)| \\
&\leq \sum_{i=1}^{p-1}\int_{A_i} |g(x)-g(x_i)|\gamma(dx) + \sum_{i=1}^{p-1}|g(x_i)|(\gamma(A_i)-a_i) + \\
&\qquad + \int_{A_p} |g(x)-g(x_p)|\gamma(dx) + |g(x_p)||\gamma(A_p)-a_p| \\
&\leq 2\xi\sum_{i=1}^{p}\gamma(A_i) + \sum_{i=1}^{p-1}(\gamma(A_i)-a_i) + |\gamma(A_p)-a_p| \\
&= 2\xi + 2\sum_{i=1}^{p-1}(\gamma(A_i)-a_i) \\
&\leq 2\xi + 2\frac{p-1}{n}
\end{aligned}$$

Choose $n=[\frac{p}{\xi}]$, it comes

$$\beta(\gamma,\alpha)\leq 4\xi\,,$$

and for $\xi\leq 1$

$$Card(\mathcal{Y}_n)\leq (\frac{4e}{\xi})^p\,.$$

We thus obtained

$$m_\xi \leq (\frac{16e}{\xi})^{m(E,\xi/4,d)}\,.$$

□

Remark 2.26. The above proof indicates why we have chosen Fortet-Mourier distance β, instead of the Total Variation distance. Indeed the latter does not furnish tractable estimate of m_ξ in terms of $m(\xi)$. Nevertheless, all the results up to (and including) Theorem 2.17 remain true with the total variation distance, for which we can even assume that the U_i's are only continuous.

Of course in some particular cases one can describe explicitly some sequence of Lipschitz functions that converges towards any continuous function, and so get precise estimates. This will be discussed in the next section, in a non compact framework. One can also use 2.9.2, but here again the decay of ε_n is not explicit.

3. Gibbs conditioning principle : the non compact $\mathbb{R}^d$ case.

The notations of section 2 are in force, except that we do no more assume that E is compact and thus replace the set *Lip* of Lipschitz functions by *BLip* the set of bounded Lipschitz functions equipped with its usual norm.
It is certainly out of reach to get similar results as in the compact case. Indeed, quantitative relationship between covering number of compact sets K and growth of $\mu(K)$ will clearly play some important role. Therefore we shall here restrict ourselves to $E = \mathbb{R}^d$ (other examples like path spaces will be discussed elsewhere), and will use two different methods. One is one point compactification, the other one (that can be extended to non locally compact spaces) uses the previously mentioned quantitative relationship.

3.1. **One point compactification.** Let $\mathbb{S}^d$ be the unit sphere in $\mathbb{R}^{d+1}$. N will denote the north pole and ψ the stereographic projection of $\mathbb{S}^d - \{N\}$ onto $\mathbb{R}^d$. $Lip(\mathbb{S}^d)$ is in one to one correspondence with $LLip(\mathbb{R}^d)$ the set of (bounded) Lipschitz functions that have a " nice limit at infinity". A measure α on $\mathbb{R}^d$ is lifted onto $\check{\alpha}$ by

$$\check{\alpha}(B) = \alpha(\psi(B))$$

such that $\check{\alpha}(\{N\}) = 0$, and conversely. If we denote by $\check{\beta}$ the Fortet-Mourier distance on $M_1(\mathbb{S}^d)$ it is easily seen that

$$\check{\beta}(\check{\alpha}, \check{\gamma}) = \beta(\alpha, \gamma)$$

for any pair of Probability measures on $\mathbb{R}^d$.
Consequently if the U_i's belong to *LLip* one can directly apply Theorem 2.17 (note that

$$m(\mathbb{S}^d, \xi) \leq c\,\xi^{-d}\,) .$$

More interesting is the case when the U_i's only belong to *BLip* (for instance trigonometric functions). Indeed we have to consider the lifted

$$\check{A}_{\varepsilon_n} = \{\check{\alpha}\,,\, \alpha \in A_{\varepsilon_n}\}$$

which has empty interior. So we first have to enlarge $\check{A}_{\varepsilon_n}$ by considering its closed $C(A)\varepsilon_n$ blowup in $(M_1(\mathbb{S}^d), \check{\beta})$. If we denote by $M_1^N(\mathbb{S}^d)$ the set of Probability measures that do not charge N, one sees that

$$\check{A}_{\varepsilon_n}^{C(A)\varepsilon_n} \cap M_1^N(\mathbb{S}^d) \,\subseteq\, \check{A}_{2\,\varepsilon_n} .$$

Hence, up to a factor 2 we can proceed as in the proof of Theorem 2.17, provided the Sanov exact lower bound is still true. But since μ and ν_0 do not charge N, the proof of Proposition 2.9 is still working, so that (2.20) remains true (just change A_ε into $\check{A}_{2\,\varepsilon}$).
The rest of the proof is unchanged (with the modification above). So

Theorem 3.1. *If $E = \mathbb{R}^d$ and the U_i's are in BLip, the statement of Theorem 2.17 is still available (with $m(\xi)$ replaced by $c\,\xi^{-d}$ and the bound of Lemma 2.18 for m_ξ).*
The same holds with the statement of Theorem 2.24.

Even more interesting is the case when the U_i's are only continuous (but still bounded). We then have the following

Theorem 3.2. *If A is given by*

$$A = \{\alpha \in M_1(E)\,,\, < \alpha, U_i >= a_i \text{ for all } i = 1, ..., j\}$$

with uniformly continuous and bounded U_i's. Assume that there exists $\nu \in A$ such that $H(\nu|\mu) < +\infty$ and $\nu \sim \mu$. Define $mod(x)$ as the maximum of the moduli of continuity of the U_i's, and for $\varepsilon > 0$ define

$$A_\varepsilon = \{\alpha \in M_1(E)\,,\, | < \alpha, U_i - a_i > | \leq \varepsilon \text{ for all } i = 1, ..., j\}\,.$$

Let

$$\mu^n_{A_{\varepsilon_n},k}(B) = \mu^{\otimes n}\Big((X_1, ..., X_k) \in B \,/\, L_n \in A_{\varepsilon_n}\Big)\,.$$

Then

$$\beta(\mu^n_{A_{\varepsilon_n},k}, \nu_0^{\otimes k}) \to 0$$

where ν_0 is the Gibbs measure defined in (2.12), provided

$$\varepsilon_n \gg \max\big(mod(n^{-b})\,,\, n^{-(a-b)}\big) \quad \text{for some } 0 < b < a < \frac{1}{d+2}\,.$$

A similar statement holds for Theorem 2.24 when the log-density of the minimizer is only assumed to be bounded and uniformly continuous.

Proof. That ν_0 given by (2.12) minimizes relative entropy is known. Define U_i^η as the convolution product $U_i * J_\eta$ where J_η is the gaussian kernel with covariance matrix $\eta^2\, Id$. Then

$$\sup_{x \in \mathbb{R}^d} |U_i^\eta(x) - U_i(x)| \leq mod(\eta) \quad \text{and} \quad \parallel U_i^\eta \parallel_{BLip} \leq K\eta^{-1}\,,$$

where K only depends on the sup norm of the U_i's. Define $A_{\varepsilon,\eta}$ as A_ε just replacing U_i by U_i^η. Then

$$\nu_0 \in A_{mod(\eta),\eta} \quad \text{and} \quad A_{\varepsilon - mod(\eta),\eta} \subseteq A_\varepsilon$$

provided $\varepsilon > mod(\eta)$. Define f_0^η as in (2.12) replacing U_i by U_i^η, and $\nu_0^\eta = f_0^\eta\, \mu$. Then we may find constants C_0 and $C_0' \geq 1$ depending on f_0 such that

$$\sup_{x \in \mathbb{R}^d} |f_0^\eta(x) - f_0(x)| \leq C_0\, mod(\eta) \quad \text{and} \quad \nu_0^\eta \in A_{C_0'\, mod(\eta),\eta}\,.$$

Hence

$$B(\nu_0^\eta, \alpha) \subset A_{\varepsilon - mod(\eta),\eta} \subseteq A_\varepsilon\,,$$

provided $\varepsilon > (1 + C_0')mod(\eta) + K\alpha\eta^{-1}$.
We thus can get the analogue of (2.21), by introducing the closed α blowup of $\check{A}_{\varepsilon - mod(\eta),\eta}$ in the set of measures on $\mathbb{S}^d$. We have

$$\check{A}^\alpha_{\varepsilon - mod(\eta),\eta} \subseteq \check{A}_{\varepsilon - mod(\eta) - K\alpha\eta^{-1},\eta}\,,$$

and may choose ε as above.
We thus obtain that for some C_0'' depending on f_0, ν_n, denoting the minimizer of relative entropy on A_{ε_n}

$$H(\mu^n_{A_{\varepsilon_n},k}|\nu_n^{\otimes k}) \leq C_0''\alpha_n\eta_n^{-1} - \frac{k}{n}\log\Big(1 - m_{\frac{\alpha_n}{2}}\exp(-\frac{n}{8}\,\alpha_n^2)\Big) + \tag{3.3}$$
$$+ k\big(H(\nu_0^{\eta_n}|\mu) - H(\nu_0|\mu)\big) + k\big(H(\nu_0|\mu) - H(\nu_n|\mu)\big)\,.$$

The rest of the proof is similar to what has been done before (note that $f_0^{\eta_n}$ goes to f_0 in sup norm) . Conditions on ε_n follow from what is said above and Lemma 2.19. □

3.2. **A direct study.** Since one point compactification has no chance to extend to non locally compact spaces, we shall now try to provide another viewpoint, by using a quantitative description of the approximation by compact sets. We still assume that $E = \mathbb{R}^d$ mainly to get a good description of covering number in terms of volumes, i.e.

$$m(B(0,R),\xi) \, \leq \, C(d) \Big(\frac{R}{\xi}\Big)^d . \tag{3.4}$$

For any probability measure ν on $\mathbb{R}^d$ we then introduce

$$\nu^R = \frac{\mathbb{I}_{B(0,R)} \, \nu}{\nu(B(0,R))} .$$

It is immediate that

$$\beta(\nu^R,\nu) \, \leq \, \nu(|x|>R) + \frac{1}{\nu(B(0,R)} - 1 \, \leq \, \nu(|x|>R)\Big(1+\frac{1}{\nu(|x|\leq R)}\Big) . \tag{3.5}$$

Hence if the U_i's are in $Blip$, $A = A_0$, ν_0 and f_0 being defined as before,

$$B(\nu_0^R, \varepsilon^R) \subset A_\varepsilon$$

for

$$\varepsilon^R = C(A)\varepsilon - \Big(\parallel f_0 \parallel_\infty \mu(|x|>R)\big(1+\frac{1}{\min f_0 \, \mu(|x|\leq R)}\big)\Big) .$$

So we can apply Sanov exact lower bound with ν^R, considering the induced topology on $M_1(B(0,2R))$ (which is a measurable (actually a closed) subset of $M_1(\mathbb{R}^d)$). Indeed

$$\begin{aligned} Q_n^\mu(G) \quad &= \quad \int \mathbb{I}_G(L_n^x)\mu(dx_1)...\mu(dx_n) \\ &\geq \quad \int \mathbb{I}_G(L_n^x)\, \Pi_{i=1}^n \mathbb{I}_{|x_i|\leq R} \; \mu(dx_1)...\mu(dx_n) \\ &\geq \quad (\nu_0(B(0,R)))^n \int \mathbb{I}_G(L_n^x)\exp(-n < L_n^x, h_0 >)\nu_0^R(dx_1)...\nu_0^R(dx_n) \\ &\geq \quad \exp\big(-nH(\nu_0^R|\mu^R) + n \, \log(\nu_0(B(0,R)))\big) \\ &\qquad\qquad \int \mathbb{I}_G(L_n^x)\exp(-n < L_n^x - \nu_0^R, h_0 >)\nu_0^R(dx_1)...\nu_0^R(dx_n) \end{aligned}$$

i.e.

$$Q_n^\mu(A_\varepsilon)\, \exp\big(n\,(H(\nu_0^R|\mu)\big) \geq \exp\big(-n\, C\varepsilon^R + n\, \log(\nu_0(B(0,R)))\big)\, \Big(1 - Q_n^{\nu_0^R}(B_R^c(\nu_0^R,\eta))\Big) ,$$

where the ball B_R is relative to this topology. It yields

$$Q_n^\mu(A_\varepsilon)\, \exp\big(n\,(H(\nu_0^R|\mu)\big) \geq \exp\big(-n\, C\varepsilon^R + n\, \log(\nu_0(B(0,R)))\big)\, \Big(1 - m_{\frac{\varepsilon^R}{2}} \exp\big(-\frac{n}{8}\,(\varepsilon^R)^2\big)\Big) ,$$

with

$$m_\xi \leq \Big(\frac{16e}{\xi}\Big)^{C'(d)\left(\frac{R}{\xi}\right)^d} .$$

Hence if we choose ε_n going to 0 and R_n going to $+\infty$ in such a way that

$$\varepsilon_n \gg \mu(|x| > R_n) \quad \text{and} \quad n\varepsilon_n^2 \gg \left(\frac{R_n}{\varepsilon_n}\right)^d \log(\frac{1}{\varepsilon_n})\,, \tag{3.6}$$

we will again obtain that

$$\beta(\mu^n_{A_{\varepsilon_n},k}, \nu_0^{\otimes\, k}) \to 0$$

provided $|H(\nu_0^{R_n}|\mu) - H(\nu_0|\mu)| \to 0$ which is easily seen thanks to Lebesgue bounded convergence theorem. (Actually it is enough to show that $|H(\nu_0^{R_n}|\mu) - H(\nu_n|\mu)| \to 0$, and this result follows by directly using (2.22)).

Condition (3.6) can be made explicit if μ admits some moments. In particular

$$\text{if } E^\mu(|x|^p) < +\infty\,, \text{ one can choose } \varepsilon_n \sim n^{-a} \quad \text{with } 0 < a < \frac{1}{d(1+\frac{1}{p})+2}\,, \tag{3.7}$$

thanks to Markov inequality, and if for some $\lambda > 0$

$$E^\mu(e^{\lambda|x|}) < +\infty\,, \text{ one can choose } \varepsilon_n \sim n^{-a} \quad \text{with } 0 < a < \frac{1}{d+2}\,. \tag{3.8}$$

Remark 3.9. The behaviour of ε_n is not as good as what is obtained via one point compactification, but the method here extend to more general state spaces.

4. **Improved lower bounds, applications.**

What we shall do in this section is see how the estimates in the previous two sections can be improved in various directions. However these improvements are only possible for some particular choices of A i.e. the ones given by a finite number of linear constraints as in (2.10). As we said in the Introduction, what follows is more or less part of the folklore in Large Deviations theory.

The framework is even more general. E is any Polish space, A is defined as in (2.10) but we only assume that the U_i's have all their exponential moments that is

$$\int e^{t\,U_i}\, d\mu \,<\, +\infty \quad \text{ for all } t \in \mathbb{R}\,. \tag{4.1}$$

This assumption is enough for applying all the results (2.12)-(2.15) as soon as (2.11) is satisfied (the situation when only some exponential moments are finite is more delicate, see [12] Lemma 7.3.6 , [22] and the discussion in subsection 4.2).

Now recall (2.21)

$$\begin{aligned} H(\mu^n_{\varepsilon_n,k}|\nu_{\varepsilon_n}^{\otimes k}) \;&\leq\; \frac{-1}{[\frac{n}{k}]} \log\Big(Q^\mu_n(A_{\varepsilon_n}) \exp(n\,H(A_{\varepsilon_n}|\mu))\Big) \\ &\leq\; \frac{-k}{n} \log\Big(Q^\mu_n(A_{\varepsilon_n}) \exp(nH(A_0|\mu))\Big) + k\big(H(A_0|\mu) - H(A_{\varepsilon_n}|\mu)\big)\,. \end{aligned} \tag{4.2}$$

Actually we called upon the results in section 7.3.3 of [12] to use (2.21). These results lie on some assumption (A-5) ([12] p.335) which is not satisfied here. But this assumption is only used in order to check that $\mu^n_{\varepsilon_n,1}$ belongs to A_{ε_n} (see the proof of Theorem 7.3.21 p.337 in [12]) which is immediate here due to exchangeability (see (7.3.1) p.323 in [12]). Hence (4.2) holds.

We may write

$$Q^\mu_n(A_{\varepsilon_n}) = \exp(-nH(\nu_0|\mu)) \int \mathbb{I}_{A_{\varepsilon_n}}(L^x_n) \exp(-n < L^x_n - \nu_0, h_0 >)\nu_0(dx_1)...\nu_0(dx_n)\,, \tag{4.3}$$

where $h_0 = \log\left(\frac{d\nu_0}{d\mu}\right)$. Define for $l = 1, ..., j$

$$Y^l_i = U_l(X_i) - < \nu_0, U_l >= U_l(X_i) - a_l\,.$$

Then $E^{\nu_0}(Y^l_i) = 0$ and

$$\int \mathbb{I}_{A_{\varepsilon_n}}(L^x_n) \exp(-n < L^x_n - \nu_0, h_0 >)\nu_0(dx_1)...\nu_0(dx_n) \ \geq \tag{4.4}$$

$$\geq\ e^{-\,c(\nu_0)\,n\,\varepsilon_n}\,\mathbb{P}^{\nu_0}\Big(|\sum_{i=1}^n Y^l_i| \leq\ n\,\varepsilon_n\,,\, l = 1, ..., j\Big)\,,$$

where $c(\nu_0)$ only depends on the coefficients c_l in (2.12) (actually one can choose $c(\nu_0) = j\,\max_{l=1,...,j} |c_l|$).

4.1. **Improved lower bounds via concentration inequalities.** In view of (4.4) one can of course call upon Bernstein's inequality (see [13] 2.2.11 p.103). Remark that f_0 belongs to all the $\mathbb{L}^p(\mu)$ since the U_i's have all their exponential moments (apply Hölder's inequality repeatedly). Hence the variables Y^l_i have all their exponential ν_0 moments, once again thanks to Hölder. So we may use Bernstein's inequality in order to get

$$Q^\mu_n(A_{\varepsilon_n})\ \geq\ \exp(-nH(\nu_0|\mu))\,e^{-\,c(\nu_0)\,n\,\varepsilon_n}\,\Big(1\ -\ 2j\exp(-\,C(\nu_0)\,n\varepsilon^2)\Big)\,, \tag{4.5}$$

where $C(\nu_0)$ depends one more time on the c_l's for $l = 1, ..., j$.

One can then proceed as in the proof of Theorem 2.17. Let us state the result

Theorem 4.6. *Let U_l $(l = 1, ..., j)$ satisfying (4.1). If (2.11) holds and $1 \gg \varepsilon_n \gg n^{-\frac{1}{2}}$,*

$$\beta(\mu^n_{\varepsilon_n,k(n)}, \nu_0^{\otimes k(n)})\ \to\ 0\,,$$

provided

$$k(n)\,\varepsilon_n\ \to\ 0 \quad \text{ and } \quad k(n)\Big(H(A_0|\mu) - H(A_{\varepsilon_n}|\mu)\Big)\ \to\ 0\,.$$

When $k(n) = k$ is fixed the above convergence holds in the stronger relative entropy sense.

If convergence in Fortet-Mourier distance immediately follows from the proof of Theorem 2.17, convergence in entropy requires one word of explanation. Looking at the end of the proof of 2.17, one sees that one can push the argument further and show that the θ^i_n and d_n are going to 0. Hence for a fixed k convergence in entropy follows from the entropy decomposition we used in the proof of 2.17.

This result is of course much better than all the results we have proved in the previous two sections. As it was expected at least for a fixed k, some good size of enlargement is (a little bit greater than) $\sqrt{\frac{1}{n}}$. But we saw in various places that the particular form of A is crucial.

4.2. **The Hyperplane case.** In this subsection we assume that $j = 1$, i.e. A is an hyperplane $< \alpha, U >= 1$, and we assume that U is nonnegative. To save place we shall use in this subsection (and only here) the notations of [12] section 7.3. In particular $-U$ is allowed to only have some (not all) exponential moments up to some $\beta_\infty \geq -\infty$. We also assume that the hypotheses of Lemma 7.3.6 in [12] are in force. Then there exists an unique Gibbs measure in A_ε for ε small enough. One can then use the same arguments as in subsection 4.1. However these arguments may be improved by using Berry-Eessen bound.

Indeed, according to (4.4) it is enough to get an estimate for

$$\mathbb{P}^{\nu_0}\Big(0 \leq \frac{1}{\sqrt{n}}\,(\sum_{i=1}^n Y_i) \leq \sqrt{n}\,\varepsilon_n\Big),$$

for centered real valued i.i.d random variables Y_i having all their moments. According to the Berry-Eessen bound (we do not need any expansion contrary to the much more delicate statement in Theorem 3.7.4 in [12]), provided the above quantity is much greater than $n^{-\frac{1}{2}}$ when the Y_i's are standard normal variables, the same holds for any random variables Y_i. Hence what is required is $\varepsilon_n \gg n^{-1}$. We thus have

Theorem 4.7. *The statement of Theorem 4.6 is still true for $j = 1$ and $\varepsilon_n \gg n^{-1}$.*

4.3. **Connection with Nummelin's conditional law of large numbers.** Let us formulate the problem in a may be more illuminating way (if one wants to connect Gibbs and Nummelin principles). What we are looking at is

$$(\mathbb{P}^\mu)^{\otimes n}\Big((X_1, ..., X_{k(n)}) \in . \,/\, |\sum_{i=1}^n Y_i^l| \leq n\,\varepsilon_n \,,\, l = 1, ..., j\Big)\,.$$

Hence the conditioning set can be viewed as a ball for a j dimensional empirical mean. Thus one can mimic the arguments used by J.Kuelbs and A.Meda in the proof of their Theorem 7 ([20]) (that deals with the particular case $U(x) = x$ in a much more general framework) to relate the correct rate ε_n with the rate for Nummelin's theorem.

Let us briefly describe the argument.

Let a be the μ mean of $Y_1, ..., Y_j$. D will be a half space in $\mathbb{R}^j$ delimited by the tangent hyperplane at the origin (which is the ν_0 mean of $Y_1, ..., Y_j$) to the level surface of the Cramer transform of the μ law of $Y_1, ..., Y_j$. Of course D is the half space that does not contain a. In most cases 0 will be a dominating point of D. Nummelin's weak law of large numbers shows that for correct rates

$$\mathbb{P}^{\otimes n}(\sum_{i=1}^n Y_i^l \in B(0, \varepsilon_n) \,/\, \sum_{i=1}^n Y_i^l \in D)$$

goes to 1. Hence we may replace the conditioning set A_{ε_n} by D, and thus apply the usual Gibbs conditioning with the no more thin D.

In particular for $j \geq 2$ one find the same $\gg \sqrt{\frac{1}{n}}$, and for $j = 1$ the rate is improved in $\varepsilon_n \gg n^{-1}$. In addition a more precise description of $k(n)$ is given (see the discussion in section 8 of [20]). The only advantage of what we did in the previous subsection is that one get some control for $\beta(\mu^n_{\varepsilon_n, k(n)}, \nu_0^{\otimes k(n)})$.

It is thus natural to ask whether we did not loose time in obtaining the (worse) results in the previous sections. If A is given by an infinite number of linear constraints, both the concentration method and the use of Nummelin's law are useless. For the first one it is easily seen just looking at the proof of Theorem 4.6 (however see section 7 for a more precise discussion). For the second one, even if one considers an enumerable collection of linear constraints, the natural space for the Y_l's is $\mathbb{R}^{\otimes \mathbb{N}}$ and not a Banach space. Examples of such problems are discussed in the next two sections. A short discussion of an intermediate problem (taking an increasing number of constraints) will be done in section 7.

5. **Gibbs conditioning principle : Schrödinger bridges.**

Let D be either $\mathbb{R}^d$ or a d dimensional connected and compact smooth Riemanian manifold (for instance $\mathbb{S}^d$) equipped with its natural measure dv (we shall call it improperly its Lebesgue measure). We also assume that the ξ covering number of D is less than $C(D)\,\xi^{-d}$ in the compact case. In this section $E = C([0,1], D)$ with generic element $x = (x(t))_{t\in[0,1]}$, and we shall discuss Gibbs conditioning in a physically meaningful framework. That is we shall provide a statistical interpretation of Schrödinger bridges.

$\mathcal{W}$ will denote the law of a nice D valued diffusion process, and for simplicity we shall proceed with the Brownian measure (associated to the Laplace Beltrami operator) on E with initial law μ_0. Hence, under $\mathcal{W}$ the X_i's are supposed to be i.i.d. standard Brownian motions on D. The notation in the previous sections are still in force, except that $\mu = \mathcal{W}$ and the probability measures on E will be generically denoted by $\mathcal{V}$ instead of ν. We are interested in subsets $A \in M_1(E)$ of the following form

$$A = \{\mathcal{V} \in M_1(E)\,, < \mathcal{V}, F_i >= a_i \text{ for all } i \in I\}\,, \tag{5.1}$$

where I is a finite or infinite set of indices, and the F_i's have the following form

$$F_i(x) = f_i(x(t_i)) \quad \text{for some } t_i \in [0,1] \quad \text{and some } f_i \in BLip(\mathbb{R}^d)\,. \tag{5.2}$$

We shall also assume that

$$\sup_{i\in I} \parallel f_i \parallel_{BLip} \leq K\,. \tag{5.3}$$

Since A is closed and convex, the natural assumption is as before

$$\text{there exists some } \mathcal{V} \in A \quad \text{such that} \quad H(\mathcal{V}|\mathcal{W}) < +\infty\,. \tag{5.4}$$

Let us briefly discuss this existence problem.
When I is finite, the problem reduces to a finite dimensional one as studied in the previous section. One first solves the existence problem in D^I with μ the appropriate measure, and then build $\mathcal{V}$ as the Brownian bridge with appropriate marginal laws.

The next interesting case is the one where I is infinite, but for all $i \in I$, $t_i = 0$ or $t_i = 1$ while the f_i's are a determining class of bounded Lipschitz functions (for each t_i). In other words, one tries to build $\mathcal{V}$ with given marginal laws ν_t at time $t = 0$ and $t = 1$ with finite relative entropy with respect to the Brownian measure. This problem is known as the construction of Schrödinger bridges. It is solved by first building the joint law at times 0 and 1, next by considering the Brownian bridge associated to this joint law. We refer to [15] p.161-164 for the description of the construction and to [3] for sufficient conditions for the existence of the joint law (as well as its product structure). These processes are related to the so called euclidean version of Stochastic Mechanics. Some references will be given below.

The enlargement A_ε is then given as before when I is finite or by

$$A_\varepsilon = \{\mathcal{V} \in M_1(E)\,,\, \beta(\nu_t, \nu_t^0) \leq \varepsilon \ \text{ for } t \in T\}\,, \tag{5.5}$$

where T is $\{0,1\}$, ν_t^0 denoting the flow of marginal laws induced by A.

A slightly different version of this study was done by Aebi and Nagasawa (see the related chapter in [1]) in order to prove some version of the (classical) Gibbs conditioning principle for Schrödinger bridges. Let us recall their result.

Choose some partition of $\mathbb{R}^d$ into 2^J measurable sets B_j^J in such a way that the partition at level $J+1$ is a refinement of the one at level J. Now define the 2^{-J} blow up of A as

$$A_J = \{\mathcal{V}\,,\, |\mathcal{V}(x(t)) \in B_j^J) - \nu_t(B_j^J)| \leq 2^{-J} \text{ for all } t \in T \text{ and all } j\}\,.$$

Then the following holds

Theorem 5.6. *The conditional law*

$$\mathcal{W}^n_{J,k} = \mathcal{W}^{\otimes n}((X_1, ..., X_k) \in ./L_n \in A_J)$$

satisfies

$$\lim_J \lim_n \mathcal{W}^n_{J,k} = \mathcal{V}_0^{\otimes k}\,,$$

where $\mathcal{V}_0$ is the entropy minimizer in A.

Of course if we replace $(\mathbb{1}_{B_j^J})_j$ by some smooth partition of unity (f_j) the same result holds. Hence one can also ask whether one can refine this statement in our approximate thin case by taking J_n going to infinity with n.

The main advantage of the present framework (with bridges) is that it reduces to a finite dimensional situation (on D^2). To understand this and get the results we have in mind we need first recall some results on the construction we briefly discussed at the beginning of the section.

Denote by

$$\mu_{0,1} = \text{ the } \mathcal{W} \text{ joint law of } (x(0), x(1)) = p_1(u,v)\, \mu_0(du)\, dv\,, \tag{5.7}$$

where $p_1(.,.)$ is the Brownian kernel at time 1 (i.e. $p_1(.,.)$ is either a gaussian random variable in case $D = \mathbb{R}^d$ or a smooth bounded positive function in case D is compact). If ν is a probability measure on D^2 we denote by

$$\mathcal{V}(\nu) = \int \mathcal{W}_{u,v}\, \nu(du, dv)\,, \tag{5.8}$$

where $\mathcal{W}_{u,v}$ denotes the law of the Brownian bridge from u to v. The key point is the immediate

$$H(\mathcal{V}(\nu)|\mathcal{W}) = H(\nu|\mu_{0,1})\,. \tag{5.9}$$

In particular if $A = A_0$ is as in (5.5), the entropy minimizer $\mathcal{V}_0$ in A is given by $\mathcal{V}(\nu^0)$ for ν^0 minimizing $H(\nu|\mu_{0,1})$ among all $\nu \in M_1(D^2)$ with marginal laws (ν_0, ν_1).

In addition one knows that provided (5.4) holds,

$$\frac{d\mathcal{V}_0}{d\mathcal{W}} = \frac{d\nu^0}{d\mu_{0,1}}(x(0), x(1)) = f(x(0))\, g(x(1))\,, \tag{5.10}$$

for some measurable nonnegative f and g that satisfy the pair of equations

$$\begin{aligned} \frac{d\nu_0^0}{d\mu_0}(u) &= f(u) \int p_1(u,v)\, g(v)\, dv\,, \\ \frac{d\nu_1^0}{dv}(v) &= g(v) \int p_1(u,v)\, f(u)\, d\mu_0(u)\,. \end{aligned} \tag{5.11}$$

This representation is shown in [15] when μ_0 is equivalent to Lebesgue measure, and extended in a more general situation including the one we are interested in in [3] section 6.

In the sequel we denote by $\mathcal{W}^n_{\varepsilon,k}$ the analogue of $\mu^n_{\varepsilon,k}$ (i.e. the conditional law) for A_ε given by (5.5), and by $\mathcal{V}_\varepsilon$ the minimizer of relative entropy on A_ε .

Note that a similar decomposition holds for $\mathcal{V}_\varepsilon$ for suitable f_ε and g_ε. Also note that the first inequality in (2.21) is still hold i.e.

$$H(\mathcal{W}^n_{\varepsilon_n,k}|\mathcal{V}^{\otimes k}_{\varepsilon_n}) \leq \frac{-1}{[\frac{n}{k}]} \log\Big(Q^{\mathcal{W}}_n(A_{\varepsilon_n}) \exp(n\, H(A_{\varepsilon_n}|\mathcal{W}))\Big)\,. \tag{5.12}$$

According to (5.9) the calculation in the right hand side of (5.12) reduces to finite dimensional estimates in D^2. Hence the following theorem is a direct application of section 2 and subsection 3.2

Theorem 5.13. *Assume that (5.4) holds. Assume in addition that f and g in (5.10) belong to BLip. Then if ε_n goes to 0 when n goes to infinity and satisfies*

$$\varepsilon_n \gg \mu_{0,1}(|x| > R_n) \quad \textit{and} \quad n\varepsilon_n^2 \gg \left(\frac{R_n}{\varepsilon_n}\right)^{2d} \log(\frac{1}{\varepsilon_n})\,,$$

one has

$$\beta(\mathcal{W}^n_{k,\varepsilon_n}, \mathcal{V}_0^{\otimes k}) \to 0\,.$$

Of course if D is compact one can choose $R_n = R$ for a large enough R ($|x|$ will then denote the distance between x and an arbitrary point x_0 in D).

A similar statement holds in Theorem 5.6 for $2^{-J_n} = \varepsilon_n$.

One should improve the Theorem just assuming that f and g are bounded and continuous as in Theorem 3.2. But the main problem is still to prove such a regularity. The following is one step in this direction.

Corollary 5.14. *Assume that D is compact, $\mu_0 = \nu_0^0$ and $\frac{d\nu_1^0}{dv}$ is Lipschitz and everywhere positive (hence bounded from below by some positive constant). Then (5.4) is satisfied and the conclusion in Theorem 5.13 holds with $\varepsilon_n > n^{-a}$ for some $a < \frac{1}{2d+2}$.*

Indeed in this case $\nabla \log(p_1)$ is bounded. It implies that, provided $\frac{d\nu_1^0}{dv}$ is Lipschitz and everywhere positive, one can find some versions of f and g that are Lipschitz too (just apply Lebesgue differentiation theorem under the integral sign in (5.10) and recall $\mu_0 = \nu_0^0$). That (5.4) is satisfied follows from [3] Proposition 6.3 and our hypotheses on the marginal laws.

A similar statement in the $\mathbb{R}^d$ case is much more delicate since differentiation under the integral sign is not easy to justify in general. Note that we should try to approximate f and g by Lipschitz function f_n and g_n at least in $\mathbb{L}^1(\mu_{0,1})$. The main problem is that the approximation rate η_n not only depends on f and g but will introduce an additional problem. Indeed, one has to choose $\varepsilon_n \gg \eta_n$ for the approximate law to be in the interior of A_{ε_n} but at the same time we must have

$$\varepsilon_n \parallel f_n g_n \parallel_{BLip} \to 0\,,$$

in order to get some convergence (recall that the Sanov exact lower bound involves the $BLip$ norm). This competition does not seem to be tractable.

However if we assume that the initial law is a Dirac mass, we may choose f as a constant and the regularity of g only depends on $\frac{d\nu_1^0}{dv}$. Hence we get

Corollary 5.15. *Let $D = \mathbb{R}^d$ and $\mu_0 = \nu_0^0 = \delta_{x_0}$. Denote by $\mathcal{N}(x_0)$ the gaussian law with mean x_0 and variance 1. Then if $\log(\frac{d\nu_1^0}{d\mathcal{N}(x_0)})$ is in BLip, the conclusion of Corollary 5.14 is still hold. If it is only uniformly continuous and bounded, then a similar statement holds but for ε_n as in Theorem 3.2.*

6. Gibbs conditioning principle : Nelson processes.

We shall continue to use the framework in the previous section but this time we will choose $T = [0, 1]$ and for each $t \in T$ a determining class of $BLip$. That is one is led to build $\mathcal{V}$ with given marginal laws ν_t at each time t. That this problem is connected with the existence of Nelson's diffusion processes (see [23] and [2]) was first remarked by H.Föllmer (see [15] p.165-167). This point of view was further developed by C.Léonard and the first named author (see [4], [5] and [6] for the existence problem). We shall recall the results in these papers when necessary.

The enlargement of A will be given either by A_ε as in (5.5) or defining A_J as an approximation using bridges. More precisely considering T_J the set of dyadic numbers of level J in $[0, 1]$, we define

$$A_J = \{\mathcal{V} \in M_1(E)\,,\, \beta(\nu_t, \nu_t^0) \leq 2^{-l_n} \text{ for } t \in T_J\}\,. \tag{6.1}$$

A similar statement as Theorem 5.6 can be shown exactly as in [1]. If one wants to study the approximate thin case we certainly will need some regularity properties for densities. In this case we have the following

Theorem 6.2. *Assume that (5.4) holds and that $\frac{d\mathcal{V}_0}{d\mathcal{W}} \in BLip(E)$, where $\mathcal{V}_0$ is the entropy minimizer on A. Then*

1. *if D is compact, ε_n goes to 0 when n goes to infinity and satisfies for some $0 < a < 2d$,*
$$\varepsilon_n \gg (\log(n))^{-\frac{1}{2d-a}}$$
one has
$$\beta(\mathcal{W}^n_{k,\varepsilon_n}, \mathcal{V}_0^{\otimes k}) \to 0\,,$$
2. *the same holds if $D = \mathbb{R}^d$ provided $\mu_0(|x| > R) \leq C\,R^{-p}$ for some positive p,*

3. *if D is compact, J_n and l_n are going to $+\infty$ when n goes to infinity and satisfy*

$$\log_2(n^{\frac{1}{2d}}) \gg l_n\, J_n$$

one has

$$\beta(\mathcal{W}^n_{k,J_n}, \mathcal{V}_0^{\otimes k}) \to 0\,,$$

4. *if $D = \mathbb{R}^d$, $\mu_0(|x| > R) \leq C\,R^{-p}$ for some positive p , J_n and l_n are going to $+\infty$ when n goes to infinity and satisfy*

$$\frac{1}{2d(1+\frac{2}{p})}\,\log_2(n) \gg l_n\, J_n$$

one has

$$\beta(\mathcal{W}^n_{k,J_n}, \mathcal{V}_0^{\otimes k}) \to 0\,.$$

Proof. We start with (6.2.3) and (6.2.4). What we need is an estimate for

$$Q_n^{\mathcal{W}}(A_{J_n})\exp(n\,H(A_0|\mathcal{W}))\,.$$

One can proceed as for the proof of Theorem 5.13 just replacing $\mu_{0,1}$ by μ_{T_J}. So what we need is

$$2^{-l_n} \gg \mu_{T_{J_n}}(|x(T_{J_n})| > R_n) \quad \text{and} \quad n\,2^{-2l_n} \gg l_n\,\Big(R_n\,2^{l_n}\Big)^{d(2J_n+1)}\,.$$

In the compact case one can choose $R_n = R$ large enough. When $D = \mathbb{R}^d$ one has

$$\begin{aligned}\mu_{T_{J_n}}(|x(T_{J_n})| > R_n) &\leq \mu_0(|x(0)| > \frac{R_n}{2}) + \mathcal{W}(\sup_{0\leq t\leq 1}|x(t)-x(0)| > \frac{R_n}{2}) \\ &\leq \mu_0(|x(0)| > \frac{R_n}{2}) + e^{-c(d)R_n^2}\,.\end{aligned}$$

Hence we may choose $R_n = 2^{\frac{2}{p}\,l_n}$ and we get the result.

When dealing with A_{ε_n} one really has to face the infinite dimension problem. The new point is thus to choose some appropriate compact subset in infinite dimension. This is done by using well known paths properties, namely the law of the Brownian motion is almost supported by some big ball of some regular functions space. We will choose the space of Hölder functions of order α (usually denoted by $Lip(\alpha)$) for some $\alpha < \frac{1}{2}$ for two main reasons. First it is certainly more familiar to the prospective reader than more sophisticated spaces, and second the arguments immediately extend to general smooth diffusion processes thanks to Kolmogorov continuity criterion. Slight improvements are possible if we consider instead of Hölder some Besov spaces and we shall shortly explain this below. But extension to more general diffusion processes is not immediate since the corresponding results do not exist (up to our knowledge) in the literature. Of course one can strongly suspect that these arguments are true. In both cases the calculation of the covering number is known. Actually this improvement does not really yield better results except for the computation of constants.

For $\alpha < \frac{1}{2}$ introduce

(6.3) $$K(R,\alpha,M) = \{x\,,\ |x(0)| \leq R \ \text{and}\ \Pi_\alpha(x) \leq M\}\,,$$

where

$$\Pi_\alpha(x) = \sup_{0\leq a<b\leq 1} \frac{d(x(a),x(b))}{|a-b|^\alpha}\,.$$

It is well known that

$$\mathcal{W}(K^c(R,\alpha,M)) \,\leq\, \mu_0(|x|>R) + C(p,\alpha)\,M^{-p}\,, \tag{6.4}$$

for all $1\leq p<+\infty$ for some constant $C(p,\alpha)$ (of course in the compact case the first term in (6.4) vanishes for R large enough) . Furthermore $K(R,\alpha,M)$ is a compact subset of E whose covering number $m(K(R,\alpha,M),\xi)$ satisfies

$$m(K(R,\alpha,M),\xi) \,\leq\, c(\alpha,d)\left(\frac{2R}{\xi}\right)^d \exp\left(k(\alpha,d)\,(\frac{M}{\xi})^{\frac{d}{\alpha}}\right). \tag{6.5}$$

To get (6.5) one uses covering by balls of radius $\frac{\xi}{2}$ for both the initial condition and the paths increments, for which the metric entropy is known (see e.g. [13] Theorem 2.7.1).

At least in the usual Brownian case (in $\mathbb{R}^d$) one can replace the balls in $Lip(\alpha)$ by the balls in the Besov space $\mathcal{B}^{\frac{1}{2}}_{p,\infty}$ (see e.g. [7]). This yields an improvement of (6.5) allowing to take $\alpha=\frac{1}{2}$ according to a well known result of Birman and Solomyak (see e.g. [18] for a nice modern proof).

Hence we can follow what is done in subsection 3.2 replacing the balls therein by the $K(R,\alpha,M_n)$ when D is compact, provided

$$\varepsilon_n \gg M_n^{-p}\,.$$

Recall that according to Lemma 2.18 and (6.5)

$$m_\xi \leq \left(\frac{16e}{\xi}\right)^{m(\frac{\xi}{4})}\,,$$

and

$$m(\frac{\xi}{4}) \,\leq\, c(\alpha,d)\left(\frac{2R}{\xi}\right)^d \exp\left(k(\alpha,d)\,(\frac{M_n}{\xi})^{\frac{d}{\alpha}}\right).$$

So we get the result by choosing p and α in an appropriate way.

When $D=\mathbb{R}^d$, the condition on μ_0 allows to choose R_n as a negative power of ε_n, so that the previous proof is still available. □

The condition $\frac{d\mathcal{V}_0}{d\mathcal{W}} \in BLip(E)$ in the previous Theorem looks a very strong one. Indeed in general one only knows that this density is given by some Girsanov density (written below in the $\mathbb{R}^d$ case, for the more general manifold value case see [6])

$$\frac{d\mathcal{V}_0}{d\mathcal{W}} = \frac{d\nu_0^0}{d\mu_0}\,G\,, \tag{6.6}$$

with

$$G=\exp\Big(\int_0^1 B(t,x(t)).dx(t) \,-\, \int_0^1 |B(t,x(t))|^2dt\Big)$$

and G is not even continuous. The second problem is that (6.6) is not correct in full generality. Indeed one has to replace the final time 1 in the definition of G by the stopping time

$$\tau = \inf\{t\geq 0\,,\,\int_0^t |B(s,x(s))|^2ds = +\infty\}\,, \tag{6.7}$$

which is not necessarily equal to 1 $\mathcal{W}$ a.s., but is equal to 1 $\mathcal{V}_0$ a.s. (see [4]).

However it is shown in [4] or [6] that, thanks to the minimality property of $\mathcal{V}_0$ one can find some sequence φ_j of smooth functions such that

$$\int_0^1 \int_D |B(s,x) - \nabla_x \varphi_j(s,x)|^2 \, \nu_s^0(dx)\, ds \;\to\; 0\,. \tag{6.8}$$

Here ∇ is the gradient operator associated with the riemanian metric. We thus get some approximation of G by the corresponding G_j which can be rewritten after integration by parts

$$G_j = \exp\Big(\varphi_j(1,x(1)) - \varphi_j(0,x(0)) - \int_0^1 \big((\frac{\partial}{\partial s} + \frac{1}{2}\,\Delta)\varphi_j + \frac{1}{2}\,|\nabla_x \varphi_j|^2\big)(s,x(s))\, ds\Big)\,, \tag{6.9}$$

which belongs to $BLip$. Unfortunately in general one has to face the same competition problem we discussed at the end of the previous section. Nevertheless if we start from the beginning with a smooth gradient diffusion process, the hypotheses in Theorem 6.2 are fulfilled and we get

Theorem 6.10. *If $(\nu_t)_{t\in[0,1]}$ is the flow of time marginal laws of the $\frac{1}{2}\,\Delta + \nabla\,\varphi$ diffusion process, for some smooth φ, then the conclusion of Theorem 6.2 holds.*

In particular if

$$\nu_t^0 = \psi^2(x)\, dx$$

for all t, the previous property holds with

$$B = \nabla\, \log(\psi)$$

which is not smooth in general. But for (5.4) to hold a necessary condition is that $\psi \in H^1(dx)$ the usual Sobolev space. Hence if D is compact the smoothness assumption above is not too strong in the stationary case. Furthermore one can approximate ψ in H^1 by

$$(\psi \vee \frac{1}{L} \wedge L) * \eta_j$$

for some mollifier η_j that furnishes an explicit control for $\Delta\, \log(\psi)$. It can be shown that the corresponding $\mathcal{V}(L,j)$ satisfies

$$H(\mathcal{V}_0 | \mathcal{V}(j,L)) \to 0$$

with a rate corresponding to the above rate of approximation. Hence we are in a situation that is very close to the one of Theorem 3.2, and a similar proof yields the analogous of Theorem 6.2.1 but for a non explicit ε_n (actually an explicit one that is not easy to describe).

We shall conclude this section by an analogous discussion but for A_J. Assume that D is compact, $\| \frac{d\nu_t^0}{dv} \|_{Lip} \leq L$ for all $t \in [0,1]$, $\frac{1}{L} \leq \frac{d\mu_0}{dv} \leq L$ and $\frac{1}{L} \leq \frac{d\nu_t^0}{dv}$ for some positive L and all $t \in [0,1]$. Assume finally that (5.4) holds. Our assumptions allow to build for each J the minimal Bridge $\mathcal{V}(J)$ with time marginal laws ν_t^0 for all $t \in T_J$ as in Corollary 5.14. One can then evaluate

$$Q_n^{\mathcal{W}}(A_{J_n}) \exp(n\, H(\mathcal{V}_{J_n} | \mathcal{W}))\,,$$

by using the exact Sanov lower bound. The only things we have to do are thus first choose l_n in such a way that

$$2^{-l_n} \parallel \frac{d\mathcal{V}(J_n)}{d\mathcal{W}} \parallel_{Lip} \to 0\,,$$

and then to check that

$$H(\mathcal{V}_{J_n}|\mathcal{V}_0) \to 0\,.$$

The second point is shown by using Csiszar's argument. But again a precise description of the Lip norm is not easy.

7. Is it possible to improve the results in section 5 and section 6 ?

We claimed in various places that we cannot use concentration inequalities for an infinite number of constraints. Let us see why.

Let $(f_j)_{j\in\mathbb{N}}$ be an infinite collection of functions. We may for instance choose an enumerable collection of determining bounded Lipschitz functions such that $\parallel f_j \parallel_{BLip}= 1$ for all j in the case of Schrödinger bridges. Concentration inequalities for

$$Z = \sup_j |\sum_{i=1}^{n} f_j(X_i)|$$

are known. They are mainly due to Talagrand (see [24]) and have been intensively studied since that time by many authors.

Such a concentration inequality holds for

$$\mathbb{P}(Z \geq \mathbb{E}(Z) + x)\,.$$

In particular a good knowledge of $\mathbb{E}(Z)$ is necessary for its use in our problem. A lot of works have been devoted to the problem of getting bounds for this mean in the theory of empirical processes (we refer to [13] and the references therein). It turns out that, in general, bounds involve intricate calculation of bracketing number (instead of covering number). So this concentration inequality does not seem to be useful for us.

Nevertheless one can ask whether the observation of a time marginal law is practically meaningful. It is not. Such an observation is performed through the observation of a large but finite linear moment constraints. What is possible to do in this framework is to let this number $j(n)$ increase with n. One may thus apply the method in subsection 4.2. Due to our assumption on the f_j's the random variables $f_j(X) - < f_j, \nu_0 >$ are equibounded. Hence Bernstein's inequality furnishes an uniform bound for each of them and what we have to choose is

$$\log(j(n)) \ll n\,\varepsilon_n^2\,.$$

Doing so the rate for ε_n in this new framework should be $\gg n^{-\frac{1}{2}}$ in the situation of Schrödinger bridges as in section 4. Unfortunately we have to face another problem, namely the control of $e^{-c(\nu_0)n\varepsilon_n}$ in (4.4). Indeed nothing ensures that $c(\nu_0) < +\infty$. This is due to the fact that the log-density h_0 is no more a simple linear combination of the f_j's. If we use an approximation (for example replace ν_0 by some ν_{ε_k} for $k > n$) what has to be done is to obtain some explicit control on the coefficients c_j^k defining the corresponding Gibbs measure. Uniform controls are unexpected and the problem becomes a very difficult one.

8. **Additional examples and Remarks.**

As we already said in various places, the strategy we used was partly developed in previous works dealing with Gibbs conditioning. For instance in the framework of Corollary 7.3.34 in [12], the minimizer ν_0 of relative entropy on the thin set A is still minimizing relative entropy on a non thin enlargement of A, so that one gets some conditional limit theorem for this enlargement. On one hand such a result should bee seen as a more interesting result since it requires less a priori information. On the other hand, our framework is physically relevant since it allows to approximate the desired energy level. In particular without any extra work (just use the estimates in section 4), one easily sees that

$$\beta(\mu^n_{\varepsilon_n,k(n)}, \nu_0^{\otimes k(n)}) \to 0$$

in Corollary 7.3.34 of [12], if we replace A by

$$A_{\varepsilon_n} = \{\nu\,,\, 1-\varepsilon_n \leq < U, \nu > \leq 1\}$$

in the statement of the Corollary, provided $\varepsilon_n \ll k(n)^{-1}$ and $\varepsilon_n \gg n^{-1}$.

The rate we have obtained for ε_n is clearly not optimal. The examples below will illustrate what can happen.

Example 1. In some cases weak convergence will hold for any ε_n. It is the case when one can prove the result for conditional densities in $\mathbb{R}^d$ (thin shell case).

For example the conditional law of X_1 knowing that the empirical mean is $a > 0$ goes to the exponential (or gaussian) law with mean a, when $(X_n)_{n\in\mathbb{N}}$ is a sample of the exponential (or gaussian) law with mean 1 (or 0). This limit is the relative entropy minimizer under the considered constraint.

As soon as the joint law of $(X_1, \sum_{i=1}^n U(X_i))$ has a density, it is natural to conjecture that this is the general situation, i.e. that a conditional convergence holds for the regular desintegration knowing $\frac{1}{n}\sum_{i=1}^n U(X_i) = 1$.

Example 2. What should happen is that for too small ε_n the conditioning is meaningless. This will hold for example when $U(X)$ is lattice. Here is a trivial example (written in a particular form for later use).

Let $E = [0,1]$, μ be the uniform measure on E, $U = a\,\mathbb{1}_{x\leq\frac{1}{2}}$ for some nonnegative a, and define $A = \{\nu \in M_1(E)\,,\, < U, \nu >= 1\}$. Then $< U, \mu >= \frac{a}{2}$ is smaller than 1 if and only if $a < 2$, and in this case $\mu \notin A$. Define $A_\varepsilon = \{\nu \in M_1(E)\,,\, | < U, \nu > -1| < \varepsilon\}$. Remark that $| < U, L_n > -1| < \varepsilon$ if and only if the number $N(n)$ of the X_i's belonging to $[0, \frac{1}{2}]$ satisfies $\frac{N(n)}{n} \in]\frac{1-\varepsilon}{a}, \frac{1+\varepsilon}{a}[$ (that is we are dealing with Binomial laws).

For $1 < a < 2$ there exists an unique Gibbs measure in A, which is then the entropy minimizer. Its density is given by

$$f_a(x) = \frac{2}{a}\,\mathbb{1}_{[0,\frac{1}{2}]}(x) + 2(1-\frac{1}{a})\,\mathbb{1}_{[\frac{1}{2},1]}(x)\,. \tag{8.1}$$

A similar discussion can be made in the case $a = 1$. But this time the minimal ν_0 is the uniform law on $[0, \frac{1}{2}]$ and is no more a Gibbs measure. This is not in contradiction with what we said in section 2 since there is no ν in A which is equivalent to μ.

Let us choose

$$\varepsilon_n = (\frac{a}{2n})\,\varepsilon\,.$$

Then

$$L_n \in A_{\varepsilon_n} = \{|N(n) - \frac{n}{a}| < \frac{\varepsilon}{2}\}\,. \tag{8.2}$$

If we choose a as an irrational number, for $\varepsilon = 1$ there exists one and only one integer number $N(n)$ satisfying (8.2). An easy binomial calculation shows that the conditioned law of X_1 goes to the Gibbs measure described above. But if $\varepsilon < 1$ it is easily seen that $L_n \in A_{\varepsilon_n}$ may be empty. Nevertheless for any subsequence such that it is not empty the conditioned law still goes to the Gibbs measure.

So we may think that as soon as $L_n \in A_{\varepsilon_n}$ is non (a.s.) empty (as least for some subsequence) the conditioned law goes to the minimizing ν_0. We are unable to give a proof or find a counter example.

9. **A statistical description of the Brownian bridge.**

We started the paper by what could be viewed as an exercise on Nummelin's theorem. We shall close it with another exercise. Our aim is to provide an elementary example where the previous ideas yield some result in a "super thin shell case", that is by considering some conditioning set A for which $H(A|\mu) = +\infty$. The framework will be as simple as possible.

Hence we consider $E = C([0,1],\mathbb{R})$ equipped with the standard Wiener measure $\mathcal{W}$. In particular $\mathcal{W}(x(0) = 0) = 1$. We shall consider the following

$$A = \{\mathcal{V} \in M_1(E)\,,\, \mathcal{V}(x(0) = x(1) = 0) = 1\}\,. \tag{9.1}$$

A can be viewed as a limit case for the Schrödinger bridges we discussed in section 4, but of course $H(A|\mathcal{W}) = +\infty$.

A natural enlargement of A is given by

$$A_\varepsilon = \{\mathcal{V} \in M_1(E)\,,\, \mathcal{V}(x(0) = 0\,,\, |x(1)| \leq \varepsilon) = 1\}\,.$$

For such an enlargement however the problem is trivial. Indeed if $L_n \in A_\varepsilon$, all particles $X_i(1)$ belong to $[-\varepsilon, \varepsilon]$, and using independence we immediately see that the conditional limit is the usual Brownian bridge $\mathcal{W}_{0,0}$ (for any sequence ε_n).

More interesting is the case when

$$A_\varepsilon = \{\mathcal{V} \in M_1(E)\,,\, \mathcal{V}(x(0) = 0) = 1\,,\, \beta(\mathcal{V} \circ x(1)^{-1}, \delta_0) \leq \varepsilon\}\,. \tag{9.2}$$

That is one can test the final law through an infinite number of linear filters as in section 5. But as we saw, we need to have a good knowledge of the minimizing law at time 1 , ν_ε, in particular we have to control $H(\nu_\varepsilon|\mathcal{N}(0,1))$ and the $BLip$ norm of its log-density if this one is in $BLip$ or a neighboring element in A_ε.

Our first remark is the following: introduce

$$f(x) = |x| \wedge 1$$

and the set

$$D_\varepsilon = \{\mathcal{V} \in M_1(E)\,,\, \mathcal{V}(x(0) = 0) = 1\,,\, < \mathcal{V}, f(x(1)) > \leq \varepsilon\}\,. \tag{9.3}$$

Then it is easily seen that

$$D_\varepsilon \subset A_{2\varepsilon} \subset D_{4\varepsilon} \,. \tag{9.4}$$

Accordingly one can expect that the asymptotic behaviour conditioning by A_{ε_n} or by D_{ε_n} is the same. We do not know whether this is true or not. Nevertheless we will replace one by the other (which is much more simple), and will prove

Theorem 9.5. *Let*

$$\mathcal{W}^n_{\varepsilon_n,k} = \mathcal{W}^{\otimes n}\big((X_1, ..., X_k) \in . \,/\, L_n \in D_{\varepsilon_n}\big) \,.$$

Then

$$\beta(\mathcal{W}^n_{\varepsilon_n,k}, \mathcal{W}^{\otimes k}_{0,0}) \to 0$$

when n goes to $+\infty$ provided

$$\varepsilon_n \gg n^{-1} \,.$$

Proof. As in section 5 we know that the minimizing $\mathcal{V}_{\varepsilon_n}$ in D_{ε_n} is the Brownian bridge with initial law δ_0 and final law

$$\nu_{\varepsilon_n}(dx) = a_{\varepsilon_n} \, e^{-c_{\varepsilon_n} f(x)} \, \mu(dx)$$

where μ is the standard normal law. Furthermore (recall (2.21))

$$H(\mathcal{W}^n_{\varepsilon_n,k} | \mathcal{V}^{\otimes k}_{\varepsilon_n}) \leq \frac{-1}{[\frac{n}{k}]} \log \Big(Q^{\mathcal{W}}_n(D_{\varepsilon_n}) \exp(n \, H(D_{\varepsilon_n}|\mu)) \Big) \,. \tag{9.6}$$

One can then rewrite for $\eta_n < \varepsilon_n$

$$Q^{\mathcal{W}}_n(D_{\varepsilon_n}) \exp(n \, H(D_{\varepsilon_n}|\mu)) = \tag{9.7}$$

$$\begin{aligned} &= \int 1\!\!1_{D_{\varepsilon_n}}(L^{x(1)}_n) \, \exp\big(n \, c_{\varepsilon_n} \, (< L^{x(1)}_n, f > - \varepsilon_n)\big) \nu^{\otimes n}_{\varepsilon_n}(dx(1)) \\ &\geq \exp(- n \, c_{\varepsilon_n} \, \eta_n) \, \mathbb{P}^{\otimes n}_{\nu_{\varepsilon_n}} \Big(- \eta_n \leq \frac{1}{n} \sum_{i=1}^n (f(X_i(1)) - \varepsilon_n) \leq 0 \Big) \,. \end{aligned}$$

Now according to the Berry-Eessen bound as in subsection 4.2 we get

$$H(\mathcal{W}^n_{\varepsilon_n,k} | \mathcal{V}^{\otimes k}_{\varepsilon_n}) \to 0 \,,$$

provided

$$\eta_n \gg n^{-1} \quad \text{and} \quad c_{\varepsilon_n} \, \eta_n \to 0 \,.$$

In addition, since $\beta(\nu_{\varepsilon_n}, \delta_0) \to 0$, a simple time reversal argument shows that

$$\beta(\mathcal{V}_{\varepsilon_n}, \mathcal{W}_{0,0}) \to 0 \,,$$

hence using the fact that $(BLip)^{\otimes k}$ is a determining class for the convergence in law, the same holds for the k tensor product.

In order to finish the proof we thus just have to estimate c_{ε_n}, that of course goes to $+\infty$. Just using Laplace method for both integrals

$$2 \, a_{\varepsilon_n} \int_0^{+\infty} e^{- \, c_{\varepsilon_n} x} \, e^{- \frac{x^2}{2}} \, dx = \sqrt{2\pi} \,,$$

and

$$2 \, a_{\varepsilon_n} \int_0^{+\infty} (|x| \wedge 1) \, e^{- \, c_{\varepsilon_n} x} \, e^{- \frac{x^2}{2}} \, dx = \sqrt{2\pi} \, \varepsilon_n \,,$$

we get successively

$$a_{\varepsilon_n} \sim \sqrt{\frac{\pi}{2}}\, c_{\varepsilon_n} \quad \text{and} \quad c_{\varepsilon_n} \sim \varepsilon_n^{-1}\,.$$

Hence it is enough to choose $\varepsilon_n \gg \eta_n$, and the proof is completed. □

The previous proof works for more general (academic) examples. But the main examples we have in mind are more intricate.

References

[1] R. Aebi. *Schrödinger diffusion processes.* Birkhäuser, Basel-Berlin-Boston, 1996.
[2] E. Carlen. Conservative diffusions. *Comm. Math. Phys.*, 94:293–316, 1984.
[3] P. Cattiaux and F. Gamboa. Large deviations and variational theorems for marginal problems. *Bernoulli*, 5:81–108, 1999.
[4] P. Cattiaux and C. Léonard. Minimization of the Kullback information of diffusion processes. *Ann. Inst. Henri Poincaré. Prob. Stat.*, 30(1):83–132, 1994. and correction in *Ann. Inst. Henri Poincaré* vol.31, p.705-707, 1995.
[5] P. Cattiaux and C. Léonard. Large deviations and Nelson processes. *Forum Mathematicum*, 7:95–115, 1995.
[6] P. Cattiaux and C. Léonard. Minimization of the Kullback information for general Markov processes. Séminaire de Probas XXX. *Lect. Notes Math.*, 1626:283–311, 1996.
[7] Z. Ciesielski, G. Kerkyacharian, and B. Roynette. Quelques espaces fonctionnels associés à des processus gaussiens. *Studia Math.*, 107:171–204, 1993.
[8] I. Csiszar. I-divergence geometry of probability distributions and minimization problems. *Ann. Prob.*, 3:146–158, 1975.
[9] I. Csiszar. Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Prob.*, 12:768–793, 1984.
[10] A. Dembo and J. Kuelbs. Refined Gibbs conditioning principle for certain infinite dimensional statistics. *Studia Sci. Math. Hung.*, 34:107–126, 1998.
[11] A. Dembo and O. Zeitouni. Refinements of the Gibbs conditioning principle. *Probab. Theory Relat. Fields*, 104:1–14, 1996.
[12] A. Dembo and O. Zeitouni. *Large deviations techniques and applications. Second edition.* Springer Verlag, 1998.
[13] A. Van der Vaart and J. Wellner. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer, 1995.
[14] U. Einmahl and J. Kuelbs. Dominating points and large deviations for random vectors. *Probab. Theory Relat. Fields*, 105:529–544, 1996.
[15] H. Föllmer. Random fields and diffusion processes, Ecole d'été de probabilités de Saint-Flour. *Lect. Notes Math.*, 1362:101–204, 1988.
[16] M. Iltis. Sharp asymptotics of large deviations in $\mathbb{R}^d$. *J. of Theo. Proba.*, 8:501–522, 1995.
[17] O. Johnson. Entropy and conditional limit theorems: based on a paper of Diaconis and Freedman. Preprint, 2000.
[18] G. Kerkyacharian and D. Picard. Replicant coding in Besov spaces. to appear in ESAIM P&S, 2002.
[19] J. Kuelbs. Large deviation probabilities and dominating points for open convex sets : nonlogarithmic behavior. *Ann. Prob.*, 28:1259–1279, 2000.
[20] J. Kuelbs and A. Meda. Rates of convergence for the Nummelin conditional weak law of large numbers. *Stoch. Proc. Appl.*, 98:229–252, 2002.
[21] S. R. Kulkarni and O. Zeitouni. A general classification rule for probability measures. *Annals of Stat.*, 23:1393–1407, 1995.
[22] C. Léonard and J. Najim. An extension of Sanov's theorem. Application to the Gibbs conditioning principle. To appear in Bernoulli.
[23] E. Nelson. Stochastic mechanics and random fields, Ecole d' été de probabilités de Saint-Flour. *Lect. Notes Math.*, 1362:429–450, 1988.

[24] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126:505–563, 1996.

Patrick CATTIAUX,, Ecole Polytechnique, CMAP, F- 91128 Palaiseau cedex, CNRS 756, and Université Paris X Nanterre, équipe MODAL'X, UFR SEGMI, 200 avenue de la République, F-92001 Nanterre, Cedex.

E-mail address: cattiaux@cmapx.polytechnique.fr

Nathael GOZLAN,, Université Paris X Nanterre, équipe MODAL'X, UFR SEGMI, 200 avenue de la République, F- 92001 Nanterre, Cedex.