The Annals of Probability
1996, Vol. 24, No. 3, 1049–1103

SPECIAL INVITED PAPER

MAJORIZING MEASURES: THE GENERIC CHAINING

BY MICHEL TALAGRAND

University Paris VI

Majorizing measures provide bounds for the supremum of stochastic processes. They represent the most general possible form of the chaining argument going back to Kolmogorov. Majorizing measures arose from the theory of Gaussian processes, but they now have applications far beyond this setting. The fundamental question is the construction of these measures. This paper focuses on the tools that have been developed for this purpose and, in particular, the use of geometric ideas. Applications are given to several natural problems where entropy methods are powerless.

CONTENTS

1. Introduction. The purpose of majorizing measures is to give sharp bounds for the supremum of a process $(X_t)_{t\in T}$ when T is provided with a distance d that controls the increments of the process. Typically, we will have a condition of the type

$$\forall\, s,t\in T,\ \forall\, u>0,\qquad P(|X_s-X_t|>u)\le 2\exp\left(-\frac{u^2}{d^2(s,t)}\right). \tag{1.1}$$

The most important case is when $(X_t)_{t\in T}$ is a Gaussian process, in which case (1.1) holds for the distance

$$d(s,t)=(E(X_s-X_t)^2)^{1/2}. \tag{1.2}$$

In that case we can even replace the right-hand side of (1.1) by $2\exp(-u^2/2d^2(s,t))$. The factor 2 is irrelevant and has purposely been omitted in (1.1).

Received January 1995; revised February 1996.
AMS 1991 *subject classifications*. Primary 60G05, 60G15; secondary 47A40.
Key words and phrases. Chaining, increment condition, boundedness of trajectories, Gaussian properties, majorization measure, matchings, random restrictions of operators.

Majorizing measures allow us to find correct orders of magnitude, but are not adapted to the pursuit of sharp numerical constants.

Majorizing measures have long been considered as an obscure and exotic topic. The first few sections of this paper, written in considerable detail and self-contained, attempt to show that actually the theory of majorizing measures relies on a few simple ideas.

Majorizing measures originated in the theory of Gaussian processes (where they play an important role). They have, however, been successfully applied in settings increasingly far from the Gaussian case. We will present two such applications (based on a common geometric idea), which appear today as the core of the theory, and which fortunately now have simple proofs. Due to space limitations, it has not been possible to make the last sections of the paper completely self-contained. Yet we have included a complete treatment of (almost) all the ingredients pertaining to the theory of majorizing measures, and the reader having penetrated the present paper should find no difficulty in accessing the more specialized literature.

Majorizing measures are simply an elaboration of the traditional chaining argument. Unfortunately, both their name and their definition are rather misleading (we feel that the name "generic chaining" would be more appropriate, since majorizing measures represent the most general possible form of chaining). For this reason we have decided not to run the risk of discouraging the reader at such an early stage, and to give the definition only after its meaning has been made clear. This will be done in Section 2, where we discuss chaining at length, we prove the majorizing measure bound and we compare it with Dudley's entropy integral. This choice prevents the precise statement of any theorem in this introduction, which therefore constitutes only a high-level description of the contents of the paper.

We certainly wish to dispel the myth that majorizing measures are "fancy stuff" that can safely be ignored for all practical purposes. As a first step in that direction, we will explain in Section 3 why entropy methods fail to explain on which ellipsoids of Hilbert space the canonical Gaussian process is bounded.

The first major success of majorizing measures was to characterize sample boundedness and sample continuity of Gaussian processes (a problem going back to Kolmogorov). The meaning of this result is simply that, for any Gaussian process $(X_t)_{t\in T}$, there exists a "generic chaining" (i.e., majorizing measure) that provides a bound for $E\sup_{t\in T} X_t$ which is of the correct order. This is nice to know, because we no longer have to seek other methods to bound Gaussian processes. This, however, does not address the more important question of how to construct such chainings, and how to determine in practical situations the order of $E\sup_{t\in T} X_t$. Very much remains to be done in that direction. There has, however, been one successful idea, of a geometric nature, relying on proper use of convexity. This idea is closely connected to the correct understanding of the ellipsoids of Section 3. It is quite interesting that a single construction allows one to treat both the "geometric" case and the case of general Gaussian processes. This construction is presented in Section 4, the characterization of boundedness of Gaussian processes is given in Section 5, and in Section 6 we explain the structure of ellipsoids with respect to majorizing measures. As an application we sketch how to derive a theorem of Ajtai, Komlòs and Tusnàdy, with an approach rather different from the original one.

Majorizing measures have recently made it possible to clarify and generalize an extremely difficult result of Bourgain [3] in harmonic analysis, in a setting very different from the Gaussian one. Due to a lack of space, we will not discuss the entire problem, but rather we will concentrate on a result (of independent interest) which is the main step of the new approach. This result (Section 7) involves the construction of a majorizing measure using geometric ideas in the spirit of Section 6.

As we have mentioned, much remains to be understood concerning the construction of majorizing measures. In Section 8 we will describe a few open questions in that direction, some of which have deceptively simple statements.

For many processes of interest, the increments are not controlled as simply as (1.1), by one single distance, but rather by several distances. Yet in several cases the lower bounds of Section 5 have been extended. In Section 9 we will provide motivation for the necessary adaptations in this direction.

This paper will not discuss the history of the topic. This history, up to 1985, is sketched in [14]; the subsequent progress discussed here is the author's work, and can be tracked down in the author's publications given in the References ([14]–[24]).

2. Chaining, generic chaining, majorizing measures. In this section we consider a metric space (T, d) and a process $(X_t)_{t\in T}$. The word "process" means here only a collection of r.v.'s. The notation T (= time) for the index set has historical reasons, but does not mean that T is a subset of $\mathbb{R}$ or $\mathbb{R}^n$. For specificity we assume that the increment condition (1.1) holds, although there is certainly nothing magical about this special condition. Our objective is to find upper bounds for the "size" of the r.v. $\sup_{t\in T} X_t$. For the purposes of this paper a good measure of this size is $E\sup_{t\in T} X_t$. To avoid problems with the supremum of possibly uncountably many r.v.'s (each of them being defined a.e.), we define

$$E\sup_{t\in T} X_t = \sup\left\{E\sup_{t\in F} X_t \colon\ F\subset T,\ F \text{ finite}\right\}.$$

We will assume that $EX_t = 0$ for each $t\in T$. Then, given any point t_0 in T, we have

$$E\sup_{t\in T} X_t = E\sup_{t\in T}(X_t - X_{t_0}).$$

The latter form has the advantage that we seek estimates for the expectation of the nonnegative random variable $Y = \sup_{t\in F}(X_t - X_{t_0})$ (where F is a finite subset of T).

Then

$$EY = \int_0^\infty P(Y\geq u)\,du.$$

Thus we look for bounds of

$$P\left(\sup_{t\in T}(X_t - X_{t_0}) \geq u\right).$$

The first bound that comes to mind is

$$P\Big(\sup_{t\in F}(X_t - X_{t_0}) \geq u\Big) \leq \sum_{t\in F} P(X_t - X_{t_0} \geq u). \tag{2.1}$$

This bound is not going to be so bad if the variables $X_t - X_{t_0}$ are rather uncorrelated. However, it is a disaster if the variables $(X_t)_{t\in F}$ are nearly identical. Thus it seems a good idea to regroup those variables X_t that are nearly identical. To do this, we consider a finite subset T_1 of T, and for t in T we find $p_1(t)$ in T_1, which we can think of as a (first) approximation of t. Those values of T to which correspond the same $p_1(t)$ are, at this first level of approximation, considered as identical. We then write

$$X_t - X_{t_0} = X_t - X_{p_1(t)} + X_{p_1(t)} - X_{t_0}. \tag{2.2}$$

The idea is that the terms $X_{p_1(t)} - X_{t_0}$ will be handled through (2.1), since there are not too many of them, and since the variables $(X_v)_{v\in T_1}$ are really different. On the other hand, the variables $X_t - X_{p_1(t)}$ are "smaller" than the original variables $X_t - X_{t_0}$, so their supremum should be easier to handle. The procedure will then be iterated.

How do we measure that $p_1(t)$ is close to t? One could require something like

$$\forall\, t, \qquad d(t, p_1(t)) \leq \tfrac{1}{10}\operatorname{diam} T, \tag{2.3}$$

where the diameter $\operatorname{diam} T$ of T is $\sup\{d(x,y);\ x, y \in T\}$.

Writing (2.3) means that we measure distances with the size of T as unit scale. This is notationally not a good idea; rather we measure distances by comparing them with numbers r^{-i}, $i \in \mathbb{Z}$. For the purposes of the present section, we could take $r = 2$. It will, however, be convenient for the sequel to allow larger values of r. (Thus we assume $r \geq 2$.) To measure the size of T, we consider the largest $i \in \mathbb{Z}$ such that $\operatorname{diam} T \leq 2r^{-i}$. The factor 2 is convenient but unimportant. For $j \geq i$, we consider a finite set Π_j of T, and for $t \in T$, we consider points $\pi_j(t) \in \Pi_j$. The idea is that the points $\pi_j(t)$ are successive approximations of t. The fundamental relation is

$$X_t - X_{t_0} = \sum_{j>i} X_{\pi_j(t)} - X_{\pi_{j-1}(t)}, \tag{2.4}$$

which decomposes the increments of the processes as one moves from t_0 to t along the "chain" $\pi_j(t)$. To make (2.4) true, we require that $\pi_i(t) = t_0$ for every t in T. The potentially infinite series in (2.4) could also create problems; but in fact (2.4) holds a.s. under the mild condition

$$\lim_{j\to\infty} d(t, \pi_j(t)) = 0. \tag{2.5}$$

To express that $\pi_j(t)$ approximates t, we could control the distance of t and $\pi_j(t)$. From (2.4), we see that what matters is the distance between $\pi_j(t)$ and $\pi_{j-1}(t)$. It turns out to be convenient to assume

$$\forall\, t \in T,\ \forall\, j > i, \qquad d(\pi_j(t), \pi_{j-1}(t)) \le 2r^{-j+1}. \tag{2.6}$$

The factor 2 is convenient but unimportant.

The traditional way to use (2.4) is as follows. Assume that for certain positive numbers a_j, we have, for some $u > 0$,

$$\forall\, t \in T,\ \forall\, j > i, \qquad X_{\pi_j(t)} - X_{\pi_{j-1}(t)} \le u a_j. \tag{2.7}$$

Then, if F is a finite subset of T on which (2.4) holds, we have, setting $S = \sum_{j>i} a_j$, that

$$\forall\, t \in F, \qquad X_t - X_{t_0} \le Su. \tag{2.8}$$

Consider the number M_j of all possible pairs $(\pi_j(t), \pi_{j-1}(t))$ as t varies through T. Then, using (2.6) and (1.1), we get

$$P(\exists\, t \in T,\ X_{\pi_j(t)} - X_{\pi_{j-1}(t)} > u a_j) \le 2M_j \exp - \frac{u^2 a_j^2}{(2r^{-j+1})^2}$$

and thus

$$P\Big(\sup_{t\in F} X_t - X_{t_0} \ge uS\Big) \le \sum_{j>i} 2M_j \exp - \frac{u^2 a_j^2}{(2r^{-j+1})^2}. \tag{2.9}$$

We see then that it is a good choice to take

$$a_j = 2r^{-j+1}\sqrt{\log 2^{j-i} M_j}.$$

In that case, the right-hand side of (2.9) becomes

$$\sum_{j>i} 2M_j (2^{j-i} M_j)^{-u^2}.$$

For $u^2 \ge 1$, this is at most

$$\begin{aligned} \sum_{j>i} 2(2^{j-i})^{-u^2} &\le 2 \cdot 2^{-u^2}\Big(\sum_{j>i} 2^{-j+i+1}\Big) \\ &\le 4 \cdot 2^{-u^2} \end{aligned}$$

and thus, in particular,

$$E \sup_{t\in T} X_t \le KS. \tag{2.10}$$

Here, as well as in the rest of the paper, K denotes a universal constant, not necessarily the same at each occurrence.

How, then, does one construct the points $\pi_j(t)$? A simple method is to choose Π_j such that

$$\forall\, t \in T,\ \forall\, j > i,\ \exists\, u \in \Pi_j, \qquad d(t,u) \le r^{-j}.$$

One then picks $\pi_j(t) \in \Pi_j$ such that $d(t, \pi_j(t)) \le r^{-j}$. This implies (2.5) and, by the triangle inequality, this implies (2.6).

How does one control the number M_j? We see that, if we set $N_j = \operatorname{card} \Pi_j$, we have $M_j \le N_j N_{j-1}$, so that

$$S \le \sum_{j>i} 2r^{-j+1} \sqrt{\log 2^{j-i} N_j N_{j-1}}.$$

It is more elegant to express this bound in a simpler manner. Using that $\sqrt{ab} \le \sqrt{a} + \sqrt{b}$, we have

$$\begin{aligned} S &\le \sum_{j>i} 2r^{-j+1} \left(\sqrt{j-i}\sqrt{\log 2} + \sqrt{\log N_j} + \sqrt{\log N_{j-1}} \right) \\ &\le K(r)\left(r^{-i} + \sum_{j\ge i} r^{-j} \sqrt{\log N_j} \right). \end{aligned} \tag{2.11}$$

Here, as well as in the rest of the paper, $K(r)$ denotes a constant that depends on r only. Let us now make a simple observation: the definition of i shows that

$$2r^{-i-1} < \operatorname{diam} T.$$

Thus we cannot have $N_{i+1} = 1$, since then we would have $\operatorname{diam} T \le 2r^{-i-1}$. Thus $N_{i+1} \ge 2$, and thus

$$r^{-i} \le K(r) r^{-i-1} \sqrt{\log N_{i+1}}$$

and (2.11) becomes

$$S \le K(r) \sum_{j \ge i} r^{-j} \sqrt{\log N_j}. \tag{2.12}$$

Certainly it is to our benefit to take N_j as small as possible. Let us recall the definition of covering numbers. Given a metric space (T, d), we denote by $N(T, d, \varepsilon)$ the smallest number N such that for a certain subset U of T with $\operatorname{card} U \le N$ we have

$$\forall\, t \in T,\ \exists\, u \in U, \qquad d(t, u) \le \varepsilon.$$

Equivalently, N is the smallest integer such that T can be covered by N closed balls of radius ε (for the distance d). Then we can take $N_j = N(T, d, r^{-j})$. Certainly the numbers $N(T, d, \varepsilon)$ increase when ε decreases. Thus

$$\varepsilon \le r^{-j} \Rightarrow N(T, d, \varepsilon) \ge N_j.$$

Thus

$$\int_0^{r^{-i}} \sqrt{\log N(T, d, \varepsilon)}\, d\varepsilon \ge \sum_{j \ge i} (r^{-j} - r^{-j-1}) \sqrt{\log N_j},$$

so that (2.12) becomes

$$S \le K(r) \int_0^{r^{-i}} \sqrt{\log N(T, d, \varepsilon)}\, d\varepsilon.$$

We observe that $N(T, d, \varepsilon) = 1$ for $\varepsilon \geq 2r^{-i} \geq \operatorname{diam} T$, so that there is no loss in writing $\int_0^\infty$ rather than $\int_0^{r^{-i}}$. Taking $r = 2$, we have proved the following result.

PROPOSITION 2.1 (Dudley's entropy bound). *Under* (1.1) *we have*

$$E \sup_{t \in T} X_t \leq K \int_0^\infty \sqrt{\log N(T, d, \varepsilon)}\, d\varepsilon. \tag{2.13}$$

When we look back at the above proof, we realize that we have not been very cautious. A first observation is that, since by the chaining equation (2.4) we try to control the increments $X_{\pi_j(t)} - X_{\pi_{j-1}(t)}$, it is to our advantage to arrange that

$$\forall\, s,\ t \in T, \qquad \pi_j(t) = \pi_j(s) \Rightarrow \pi_{j-1}(t) = \pi_{j-1}(s), \tag{2.14}$$

because this ensures that there will be fewer such increments to control. It seems also a very minor restriction to assume

$$\forall\, v \in \Pi_j, \qquad \pi_j(v) = v, \tag{2.15}$$

so that, by (2.14),

$$\pi_{j-1}(t) = \pi_{j-1}(\pi_j(t)), \tag{2.16}$$

since $\pi_j(\pi_j(t)) = \pi_j(t)$ by (2.15).

Thus, by (2.14), controlling the increments $X_{\pi_j(t)} - X_{\pi_{j-1}(t)}$ means controlling the increments $X_v - X_{\pi_{j-1}(v)}$ for $v \in \Pi_j$.

We should observe in passing that there are now at most $N_j = \operatorname{card} \Pi_j$ such increments. This is better than our previous estimate $N_j N_{j-1}$. This observation does not allow us to improve upon (2.11). (It is, however, essential when one attempts to extend (2.11) to cases where (1.1) is replaced by a moment condition $\|X_t - X_s\|_p \leq d(s, t)$; see [16] and [13].)

We observe that, under (2.16), (2.7) becomes

$$\forall\, v \in \Pi_j, \qquad X_v - X_{\pi_{j-1}(v)} \leq u a_j. \tag{2.17}$$

The crucial new idea of the majorizing measure bound is to replace (2.17) by

$$\forall\, v \in \Pi_j, \qquad X_v - X_{\pi_{j-1}(v)} \leq u a_j(v), \tag{2.18}$$

where the number $a_j(v) \geq 0$ depends on j, v. We observe under (2.18) that, if (2.5) holds for each t in F, we have

$$\forall\, t \in F, \qquad X_t - X_{t_0} \leq uS,$$

where

$$S = \sup_{t \in T} \sum_{j > i} a_j(\pi_j(t)).$$

Thus

$$P\Big(\sup_{t\in F}(X_t - X_{t_0}) \geq uS\Big) \leq \sum_{j>i}\sum_{v\in\Pi_j} P(X_v - X_{\pi_{j-1}(v)} \geq ua_j(v))$$

(2.19)

$$\leq \sum_{j>i}\sum_{v\in\Pi_j} 2\exp\Big(-\frac{u^2 a_j^2(v)}{(2r^{-j+1})^2}\Big).$$

This bound brings information only when the right-hand side is less than or equal to 1. Thus it is natural to require that this is the case for $u = 1$, and thus

$$\sum_{j>i}\sum_{v\in\Pi_j} 2\exp\Big(-\frac{a_j^2(v)}{(2r^{-j+1})^2}\Big) \leq 1.$$

In other words, setting

$$w_j(v) = 2\exp\Big(-\frac{a_j^2(v)}{(2r^{-j+1})^2}\Big), \tag{2.20}$$

we want to have $\sum w_j(v) \leq 1$. This condition reminds us of a "total weight at most 1," and this is where the idea of probability comes in. Consider a probability μ on T, and set

$$\forall\ j > 1,\ \forall\ v \in \Pi_j, \qquad w_j(v) = \mu(\{v\}).$$

Thus, by (2.20),

$$a_j(v) = 2r^{-j+1}\sqrt{\log\frac{2}{\mu(\{v\})}}.$$

Thus, by the definition of S,

$$S = 2\sup_{t\in T}\sum_{j>i} r^{-j+1}\sqrt{\log\frac{2}{\mu(\{\pi_j(t)\})}}. \tag{2.21}$$

For $u \geq 1$, the right-hand side of (2.19) is

$$\sum_{j>i}\sum_{v\in\Pi_j} 2\Big(\frac{w_j(v)}{2}\Big)^{u^2} \leq 2^{1-u^2},$$

since

$$2\Big(\frac{w_j(v)}{2}\Big)^{u^2} \leq w_j(v)2^{1-u^2},$$

and this ensures that

$$E\sup_{t\in F}(X_t - X_{t_0}) \leq KS.$$

Thus we have proved the following result.

PROPOSITION 2.2. *If μ is a probability measure on T, then, under* (1.1), (2.5), (2.6) *and* (2.15), *we have*

$$E \sup_{t \in T} X_t \leq K \sup_{t \in T} \sum_{j > i} r^{-j+1} \sqrt{\log \frac{2}{\mu(\{\pi_j(t)\})}}. \tag{2.22}$$

This is, in essence, the majorizing measure bound. The formulation is, however, neither the most useful nor the simplest possible. It thus remains to reformulate this result.

We observe that each map π_j defines a partition $\mathscr{A}_j$ of T. The partition $\mathscr{A}_j$ is the collection of sets $(A_v)_{v \in \Pi_j}$, where

$$A_v = \{t \in T;\ \pi_j(t) = v\}.$$

We have $\mathscr{A}_i = \{T\}$, and condition (2.16) implies that the sequence $(\mathscr{A}_j)_{j \geq i}$ is increasing.

We note also that, by (2.5), (2.6) and the triangle inequality,

$$\forall\ t \in T, \qquad d(t, \pi_j(t)) \leq 2 \sum_{\ell > j} r^{-\ell+1} \leq 4r^{-j}$$

since $r \geq 2$. Thus

$$\forall\ A \in \mathscr{A}_j, \qquad \operatorname{diam} A \leq 8r^{-j}. \tag{2.23}$$

This explains why increasing sequences of partitions that satisfy (2.23) will play such an important role in the sequel. Rather than (2.23) we will use (for unimportant reasons)

$$\forall\ A \in \mathscr{A}_j, \qquad \operatorname{diam} A \leq 2r^{-j}. \tag{2.24}$$

Given an increasing sequence of partitions $(\mathscr{A}_j)_{j \geq i}$, we will, throughout the paper, denote by $A_j(t)$ the unique element of $\mathscr{A}_j$ that contains t. The following is a simple reformulation of Proposition 2.2.

PROPOSITION 2.3. *Consider an increasing sequence of partitions $(\mathscr{A}_j)_{j \geq i}$ of T, such that $\mathscr{A}_i = \{T\}$. Assume* (1.1) *and* (2.24), *and consider a probability measure μ on T. Then*

$$E \sup_{t \in T} X_t \leq K(r) \sup_{t \in T} \sum_{j > i} r^{-j} \sqrt{\log \frac{1}{\mu(A_j(t))}}. \tag{2.25}$$

In the statement we, of course, implicitly assume that the sets of the partitions are μ-measurable.

PROOF. In order to apply Proposition 2.2, we choose, for each $j \geq i$ and each $A \in \mathscr{A}_j$, an arbitrary point $x_A \in A$. For each $t \in T$, we define

$$\pi_j(t) = x_{A_j(t)}. \tag{2.26}$$

Thus (2.5) and (2.6) hold and (2.15) holds since the sequence $(\mathscr{A}_j)_{j \geq i}$ increases.

Since $\sum_{j>i}\sum_{A\in\mathscr{A}_j} 2^{-j+i}\mu(A) \le 1$, it should be obvious that there exists a probability μ' on T such that

$$\forall\ j > i,\ \forall\ A \in \mathscr{A}_j, \qquad \mu'(\{x_A\}) \ge 2^{-j+i}\mu(A).$$

We now apply (2.22) to μ' rather than μ to get

$$E \sup_{t\in T} X_t \le K \sup_{t\in T} \sum_{j>i} r^{-j+1}\sqrt{\log \frac{2^{j-i+1}}{\mu(A_j(t))}}. \tag{2.27}$$

This is not quite (2.25). However, we observe that

$$\sqrt{\log \frac{2^{j-i+1}}{\mu(A_j(t))}} \le \sqrt{j-i+1}\sqrt{\log 2} + \sqrt{\log \frac{1}{\mu(A_j(t))}}.$$

Thus the right-hand side of (2.27) is at most [using $\sum_{j>i} r^{-j}\sqrt{j-i+1} \le K(r)r^{-i}$]

$$K(r)\left(r^{-i} + \sup_{t\in T} \sum_{j>i} \sqrt{\log \frac{1}{\mu(A_j(t))}}\right).$$

To complete the proof, it suffices to show that

$$r^{-i} \le K(r) \sup_{t\in T} \sum_{j>i} \sqrt{\log \frac{1}{\mu(A_j(t))}}.$$

This is a variation of the argument used at the same place in the proof of Proposition 2.1. By (2.4) and since diam $T > 2r^{-i-1}$, we must have card $\mathscr{A}_{i+1} > 1$. Thus there is $A \in \mathscr{A}_{i+1}$ with $\mu(A) \le 1/2$. If $t \in A$, then

$$r^{-i} \le K(r)\left(r^{-i-1}\sqrt{\log \frac{1}{\mu(A_{i+1}(t))}}\right). \qquad \square$$

Thus we have shown that increasing sequences of partitions of T are closely linked to the "generic chaining" argument we have presented in this section. Moreover, the main result of this paper will be an efficient method to construct such sequences of partitions. Nonetheless, there is a stronger form of Proposition 2.3 that is sometimes useful, and is more elegant, in the sense that it forgets the partitions and remembers only μ. It is due to Fernique [5]. To formulate it, we denote by $B(t, \varepsilon)$ the ball of radius ε centered at t.

PROPOSITION 2.4. *Consider a probability μ on T. Then, under* (1.1), *we have*

$$E \sup_{t\in T} X_t \le K \sup_{t\in T} \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t,\varepsilon))}}\, d\varepsilon. \tag{2.28}$$

Thus the measure μ provides a bound for $E \sup_{t\in X} X_t$, and hence is called a majorizing measure (although the name is implicitly reserved for those μ that give a decent bound!). We observe that if $\varepsilon > \operatorname{diam} T$, then $B(t, \varepsilon) = T$, so that $\mu(B(t, \varepsilon)) = 1$ and the integrand is 0. Thus the integral is in fact between 0 and $\operatorname{diam} T$. The link with (2.25) is that

$$A_j(t) \subset B(t, 2r^{-j}).$$

Thus, after comparison of the integral with a series, (2.28) appears as an improvement upon (2.25). It must be stressed, however, that, the historical and elegant formulation of (2.28) notwithstanding, it is the formulation of majorizing measures through a sequence of partitions and a family of weights that now appears as the important concept.

To prove that (2.28) follows from (2.5), we will have, given μ, to construct an appropriate sequence of partitions of T. This is not a triviality. The possibility of this construction was observed around 1985, independently in [1] and [15] (although (2.28) was known much earlier [6]). Quite remarkably, a similar construction is now at the center of the theory (Section 4). In order to avoid repetition, we will delay the proof of Proposition 2.4 until Section 4.

Certainly in (2.28) we are interested in the case where the right-hand side is as small as possible. We define the quantity $\gamma_2(T, d)$ as

$$\gamma_2(T, d) = \inf \sup_{t\in T} \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t, \varepsilon))}}\, d\varepsilon,$$

where the infimum is taken over all probability measures, and we reformulate Proposition 2.4 as follows.

THEOREM 2.5 (The majorizing measure bound). *Under* (1.1), *we have*

$$E \sup_{t\in T} X_t \le K\gamma_2(T, d).$$

It should be pointed out that if we consider another metric space (T', d'), such that there exists a contraction φ from T onto T', it is obvious that $\gamma_2(T', d') \le \gamma_2(T, d)$. (This is about the only result where the formuation of Proposition 2.4 is easier to use than the formulation of Proposition 2.3.)

3. A first look at ellipsoids. In this section we explain a natural way to look at Gaussian processes, as indexed by a subset T of the space ℓ^2 of sequences $t = (t_n)$ for which $\|t\| = (\sum_{n\ge1} t_n^2)^{1/2} < \infty$. We then demonstrate that when T is an ellipsoid, Proposition 2.1 is not sharp.

Consider a sequence $(g_n)_{n\ge1}$ of independent standard normal r.v.'s. For $t = (t_n)_{n\ge1}$, we can define

$$X_t = \sum_{n\ge1} t_n g_n \tag{3.1}$$

(convergence in L^2 and a.e.). Formula (3.1) defines a Gaussian process $(X_t)_{t\in\ell^2}$. We observe that $EX_t^2 = \sum_{n\ge1} t_n^2$. Since $X_t - X_u = X_{t-u}$, the distance induced by the process on ℓ^2 following (1.2) coincides with the distance induced by the norm of ℓ^2.

To each subset T of ℓ^2 we associate a Gaussian process $(X_t)_{t\in T}$ and we would like to understand, as a function of the "geometry" of T, when $E\sup_{t\in T} X_t$ is finite. The reason this is an important question is that this amounts in fact to studying $E\sup_{t\in U} Y_t$ for *any* Gaussian process $(Y_t)_{t\in U}$. Indeed, it can be shown that if $(h_n)_{n\geq 1}$ is an orthonormal basis of $L^2(\Omega)$, where Ω is the basic probability space, and if, for $t\in U$, we set $\varphi(t) = (E(X_t h_n))_{n\geq 1}$, then

$$E\sup_{t\in U} Y_t = E\sup_{t\in U} X_{\varphi(t)} = E\sup_{t\in\varphi(U)} X_t.$$

The traditional example used to demonstrate the superiority of Proposition 2.4 over Proposition 2.1 is the case where

$$T = \left\{\frac{e_n}{\sqrt{\log n}};\ n\geq 2\right\},$$

where (e_n) denotes the canonical basis of ℓ^2. It is left to the reader to check that

$$\sqrt{\log N(T,d,\varepsilon)} \geq \frac{1}{K\varepsilon}$$

(where d denotes, of course, the distance induced by H), so that the integral in Proposition 2.1 is infinite. On the other hand, $\gamma_2(T,d) < \infty$, as is witnessed by a measure μ such that

$$\mu\left(\left\{\frac{e_n}{\sqrt{\log n}}\right\}\right) \geq \frac{1}{Kn(\log n)^2}$$

(through a simple calculation). This example is somewhat canonical, but it is also somewhat artificial. We will rather consider here the case of the ellipsoids

$$\mathscr{E} = \left\{t = (t_n);\ \sum_{n\geq 1}\frac{t_n^2}{a_n^2} \leq 1\right\} \tag{3.2}$$

(where $a_n > 0$). Certainly it is difficult to argue that ellipsoids are unnatural. Using (3.1), we have, by the Cauchy–Schwarz inequality,

$$\sup_{t\in\mathscr{E}} X_t = \left(\sum_{n\geq 1} a_n^2 g_n^2\right)^{1/2},$$

so that

$$\begin{aligned} E\sup_{t\in\mathscr{E}} X_t &= E\left(\sum_{n\geq 1} a_n^2 g_n^2\right)^{1/2} \leq \left(E\sum_{n\geq 1} a_n^2 g_n^2\right)^{1/2} \\ &= \left(\sum_{n\geq 1} a_n^2\right)^{1/2}. \end{aligned} \tag{3.3}$$

It is not difficult to show (using the tail properties of Gaussian processes) that (3.2) is sharp in the sense that

$$\left(\sum_{n\geq 1} a_n^2\right)^{1/2} \leq KE \sup_{t\in\mathscr{E}} X_t. \tag{3.4}$$

Thus we may feel that we understand the behavior of ellipsoids with respect to the process (X_t). This is a very dangerous illusion. We actually benefited from a kind of coincidence that allows the computation of (3.3). We will show that Proposition 2.1 does *not* explain (3.3). It turns out that for the ellipsoid (3.2) we have

$$\gamma_2(\mathscr{E}) \leq K\left(\sum_{n\geq 1} a_n^2\right)^{1/2}, \tag{3.5}$$

where we make the convention (valid throughout the paper) that we omit the mention of the distance when we deal with subsets of Hilbert space and the distance induced by the norm. Only after (3.5), which is not trivial, is proved, can we say that we really understand (3.3) through Theorem 2.4. The rest of this section is devoted to estimates of entropy numbers for ellipsoids (which were discovered around 1960—see [12] and [9]) and a discussion of the results.

Consider the ellipsoid (3.1). For $k \in \mathbb{Z}$, we consider

$$I_k = \{n;\ 2^{-k} \leq a_n < 2^{-k+1}\},$$
$$J_k = \{n;\ a_n \geq 2^{-k}\}.$$

We set $n_k = \operatorname{card} I_k$, $m_k = \operatorname{card} J_k = \sum_{\ell\leq k} n_\ell$.

LEMMA 3.1. *We have* $N(\mathscr{E}, 2^{-k-1}) \geq 2^{m_k}$.

PROOF. Consider the space H_k of sequences $(t_n)_{n\in J_k}$ provided with the Euclidean norm. The map φ: $\ell^2 \to H_k$ that sends $(t_n)_{n\geq 1}$ to $(t_n)_{n\in J_k}$ satisfies

$$\|\varphi(t) - \varphi(u)\| \leq \|t - u\|.$$

Thus

$$N(\mathscr{E}, 2^{-k-1}) \geq N(\varphi(\mathscr{E}), 2^{-k-1}).$$

Now, if we denote by B the unit ball of H_k, we see that $2^{-k}B \subset \varphi(\mathscr{E})$. Thus

$$N(\varphi(\mathscr{E}), 2^{-k-1}) \geq N(2^{-k}B, 2^{-k-1}).$$

Setting $N = N(2^{-k}B, 2^{-k-1})$, by definition we can find points $t_1, \dots, t_N$ such that

$$2^{-k}B \subset \bigcup_{\ell\leq N} (2^{-k-1}B + t_\ell).$$

Thus, denoting by Vol the m_k-dimensional volume, we get

$$\operatorname{Vol}(2^{-k}B) \leq N \operatorname{Vol}(2^{-k-1}B).$$

Since $\operatorname{Vol}(aB) = a^{m_k} \operatorname{Vol} B$, we get $N \geq 2^{m_k}$. □

It follows from Lemma 3.1 that

$$2^{-k}\sqrt{\log\, N(\mathscr{E}, 2^{-k-1})} \geq 2^{-k}\sqrt{m_k} \geq 2^{-k}\sqrt{n_k}.$$

In particular, we get

$$\sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{n_k} \leq K\int_0^\infty \sqrt{\log\, N(\mathscr{E}, \varepsilon)}\, d\varepsilon.$$

On the other hand,

$$\left(\sum_{n\geq 1} a_n^2\right)^{1/2} \leq 2\left(\sum_{k\in\mathbb{Z}} 2^{-2k} n_k\right)^{1/2}.$$

In view of (3.3), this shows that Proposition 2.1 does not give a correct bound in the case of ellipsoids, as is seen in the case $\sum 2^{-k}\sqrt{n_k} = \infty$, $\sum 2^{-2k} n_k < \infty$.

We now turn to upper bounds for $N(\mathscr{E}, \varepsilon)$.

LEMMA 3.2. $\log\, N(\mathscr{E}, 2^{-k+1}) \leq K\sum_{\ell\leq k}(k-\ell+3)n_\ell$.

PROOF.

Step 1. We observe that, if $t \in \mathscr{E}$, then

$$1 \geq \sum_{n\notin J_k} \frac{t_n^2}{a_n^2} \geq 2^{2k}\sum_{n\notin J_k} t_n^2,$$

so that $(\sum_{n\notin J_k} t_n^2)^{1/2} \leq 2^{-k}$. It should then be clear that

$$N(\mathscr{E}, 2^{-k+1}) \leq N(\mathscr{E}', 2^{-k}),$$

where

$$\mathscr{E}' = \left\{t \in H_k,\ \sum_{n\in J_k} \frac{t_n^2}{a_n^2} \leq 1\right\}$$

and where H_k has been defined in the proof of Lemma 3.1.

Step 2. Consider a subset Z of $\mathscr{E}'$ with the following properties:

(3.6) Any two distinct points of Z are at mutual distance greater than 2^{-k}.

(3.7) The cardinality of Z is as large as possible.

The last condition implies that the balls of radius 2^{-k} centered at Z cover $\mathscr{E}'$. Thus

$$N(\mathscr{E}', 2^{-k}) \leq \operatorname{card} Z.$$

On the other hand, by (3.6) the balls of radius 2^{-k-1} centered at the points of Z are disjoint. These balls are contained in $\mathscr{E}' + 2^{-k-1}B$, where B is the unit ball of H_k. Thus, if we denote again by Vol the m_k-dimensional volume,

$$\operatorname{card} Z\, \mathrm{Vol}(2^{-k-1}B) \leq \mathrm{Vol}(\mathscr{E}' + 2^{-k-1}B). \tag{3.8}$$

Since $a_n > 2^{-k}$ for $n \in J_k$, we have $2^{-k}B \subset \mathscr{E}'$, so that (crudely) $\mathscr{E}' + 2^{-k-1}B \subset 2\mathscr{E}'$ and

$$\mathrm{Vol}(\mathscr{E}' + 2^{-k-1}B) \le 2^{m_k}\mathrm{Vol}(\mathscr{E}').$$

Now

$$\mathrm{Vol}(\mathscr{E}') = \Big(\prod_{n \in J_k} a_n\Big)\mathrm{Vol}(B)$$
$$\le \prod_{\ell \le k} 2^{(-\ell+1)n_\ell}\mathrm{Vol}(B),$$

since $a_n \le 2^{-\ell+1}$ for $n \in I_\ell$.

Combining with (3.8), we get

$$\mathrm{card}\, Z \le 4^{m_k}(2^{k+1})^{m_k} \prod_{\ell \le k} 2^{-\ell n_\ell} = \prod_{\ell \le k} 2^{(k-\ell+3)n_\ell}.$$

Thus

$$\log N(\mathscr{E}, 2^{-k+1}) \le \log \mathrm{card}\, Z \le \log 2\Big(\sum_{\ell \le k}(k-\ell+3)n_\ell\Big). \qquad \square$$

It turns out that ellipsoids for which $\sum a_n^2 < \infty$ do satisfy a certain condition on their entropy numbers.

PROPOSITION 3.3. *For the ellipsoid* $\mathscr{E}$ *of* (3.2), *we have*

$$\int_0^\infty \varepsilon \log N(\mathscr{E}, \varepsilon)\, d\varepsilon \le K \sum_{n \ge 1} a_n^2. \tag{3.9}$$

REMARK 1. We leave it to the reader to show, using Lemma 3.1, that

$$\sum_{n \ge 1} a_n^2 \le K \int_0^\infty \varepsilon \log N(\mathscr{E}, \varepsilon)\, d\varepsilon.$$

REMARK 2. In Section 6, we will prove, using the geometry of ellipsoids, that

$$\gamma_2(\mathscr{E}) \le \Big(\int_0^\infty \varepsilon \log N(\mathscr{E}, \varepsilon)\, d\varepsilon\Big)^{1/2}.$$

Combined with (3.9) and Theorem 2.4, this provides a considerably longer, but, as will be demonstrated, considerably more instructive proof of (3.3) (with a worse constant).

PROOF. It is standard to show that

$$\int_0^\infty \varepsilon \log N(\mathscr{E}, \varepsilon)\, d\varepsilon \le K \sum_{k \in \mathbb{Z}} 2^{-2k} \log N(\mathscr{E}, 2^{-k}).$$

Using Lemma 3.2 and interverting summation, we get

$$\begin{aligned}\sum_{k\in\mathbb{Z}} 2^{-2k}\log N(\mathscr{E}, 2^{-k}) &\le K\sum_{k\in\mathbb{Z}}\sum_{\ell<k} 2^{-2k}(k-\ell+3)n_\ell \\ &= K\sum_{\ell\in\mathbb{Z}}\Bigl(\sum_{k>\ell} 2^{-2k}(k-\ell+3)\Bigr)n_\ell \\ &\le K\sum_{\ell\in\mathbb{Z}} 2^{-2\ell}n_\ell \le K\sum_{n\in\mathbb{N}} a_n^2.\end{aligned}$$

□

4. The partitioning scheme. The ancestor of the scheme we present in this section was invented to study Gaussian processes, that is, as explained before, subspaces of Hilbert space provided with the induced distance. Fortunately, the method is valid in any metric space. It allows one, under general conditions, to construct increasing sequences of partitions such as those of Section 2.

Consider a metric space (T, d). We denote by $B(t, a)$ the ball centered at t of radius a.

We assume that, for $j \in \mathbb{Z}$, we are given a map $\varphi_j\colon T \to \mathbb{R}^+$. We assume that

(4.1) $$S = \sup\{\varphi_j(t);\ j \in \mathbb{Z},\ t \in T\} < \infty.$$

Consider a function $\theta\colon \mathbb{N} \to \mathbb{R}^+$, and assume that

(4.2) $$\lim_{n\to\infty} \theta(n) = \infty.$$

We assume that, for certain numbers $r \ge 4$, $\beta > 0$, the following holds, for any point s of T, any $j \in \mathbb{Z}$ and any $n \in \mathbb{N}$:

(4.3) Given any points $t_1, \dots, t_n$ of $B(s, r^{-j})$ such that

$$\forall\ p, q \le n, \qquad p \ne q \Rightarrow d(t_p, t_q) \ge r^{-j-1},$$

we have

(4.4) $$\varphi_j(s) \ge r^{-\beta j}\theta(n) + \min_{\ell\le n} \varphi_{j+2}(t_\ell).$$

The parameter β will be used for later purposes; for the moment the reader should assume $\beta = 1$. The most important case is when $\theta(n) = \sqrt{\log n}$ [or $\theta(n) = K^{-1}\sqrt{\log n}$]. Before we comment on the more subtle aspects of condition (4.4), we state a typical result.

THEOREM 4.1. *Assume that condition* (4.4) *holds for* $\beta = 1$, $\theta(n) = \sqrt{\log n}$. *Consider the largest* $i \in \mathbb{Z}$ *such that* $\operatorname{diam} T \le 2r^{-i}$. *Then one can find an* *increasing sequence of partitions* $(\mathscr{A}_j)_{j\geq i}$ *of* T *and a probability measure* μ *on* T *such that*

$$\sup_{t\in T}\sum_{j>i} r^{-j}\log(1/\mu(A_j(t)))^{1/2} \leq K(r)S. \tag{4.5}$$

To make the case that this is useful, we must demonstrate that (4.4) is a weak condition, and explain what these mysterious maps φ_j are and how to find them. First, (4.4) is weak because there is a min rather than a max on the right-hand side. The second (and crucial) reason is that on the right-hand side of (4.4) there is the index $j+2$ rather than $j+1$. The point here is that in all the cases we will consider $\varphi_j(s)$ will be a kind of measure of the size of $B(s, 3r^{-j})$. The number $\varphi_{j+2}(t_\ell)$ then depends only on the ball $B(t_\ell, 3r^{-j-2})$. Since $d(t_\ell, t_{\ell'}) \geq r^{-j-1}$ for $\ell \neq \ell'$, these balls are well separated from each other, at least for $r \geq 8$, and lie well inside $B(s, 3r^{-j})$. How, then, does one choose the maps φ_j? In the setting of Theorem 4.1, the choice

$$\varphi_j(s) = \gamma_2(T, d) - \gamma_2(B(s, 2r^{-j})) \tag{4.6}$$

always yields an essentially optimal result (and is essentially the choice made in Section 5). However, in many cases, we want to use Theorem 4.1 to get an upper bound on $\gamma_2(T, d)$, because this quantity is so elusive; for this reason, we cannot use the choice (4.6). Rather, one has to make a guess (based on the geometry of the situation) of useful functions φ_j, a task that I could carry out in several situations of interest. Other situations of interest, where I did not succeed, are presented in Section 8.

Before proving Theorem 4.1, we feel it appropriate to provide a natural example of a situation where condition (4.4) holds. This will help to illustrate why it is rather weak.

PROPOSITION 4.2. *Assume that* μ *is a probability measure on* T, *and that*

$$S = \sup_{t\in T}\int_0^\infty \sqrt{\log\frac{1}{\mu(B(t,\varepsilon))}}\,d\varepsilon < \infty.$$

We take $r = 8$, *and we define*

$$\varphi_j(t) = \sup\left\{\int_0^{r^{-j}} \sqrt{\log\frac{1}{\mu(B(u,\varepsilon))}}\,d\varepsilon;\ d(t,u) \leq 2r^{-j}\right\}.$$

Then conditions (4.1)–(4.4) *hold with* $\theta(n) = (1/r^2)\sqrt{\log n}$.

Together with Proposition 2.3 and Theorem 4.1, this proves Proposition 2.4.

PROOF. Conditions (4.1) and (4.2) are obvious. We prove (4.4). Consider $s, t_1, \ldots, t_n$ as in (4.3). For $\ell \leq n$, consider $s_\ell \in B(t_\ell, 2r^{-j-2})$.

We observe that

$$d(s, s_\ell) \leq d(s, t_\ell) + d(t_\ell, s_\ell) \leq r^{-j} + 2r^{-j-2} \leq 2r^{-j}.$$

We also observe that, if $\ell \neq \ell'$,

$$d(s_\ell, s_{\ell'}) \geq d(t_\ell, t_{\ell'}) - 4r^{-j-2} \geq r^{-j-1} - 4r^{-j-2} \geq 4r^{-j-2}$$

since $r = 8$. Thus the (open) balls $B(s_\ell, 2r^{-j-2})$ are disjoint for $\ell \leq n$. Thus we can find ℓ for which $\mu(B(s_\ell, 2r^{-j-2})) \leq 1/n$.

Now, since $s_\ell \in B(s, 2r^{-j})$, we have

$$\varphi_j(s) \geq \int_0^{r^{-j}} \sqrt{\log \frac{1}{\mu(B(s_\ell, \varepsilon))}}\, d\varepsilon$$

$$\geq \int_0^{r^{-j-2}} \sqrt{\log \frac{1}{\mu(B(s_\ell, \varepsilon))}}\, d\varepsilon + r^{-j-2}\sqrt{\log n},$$

since the integrand is at least $\sqrt{\log n}$ for $r^{-j-2} \leq \varepsilon < 2r^{-j-2}$.

Thus

$$\varphi_j(s) \geq \frac{1}{r^2}(r^{-j}\sqrt{\log n}) + \min_{\ell \leq n} \int_0^{r^{-j-2}} \sqrt{\log \frac{1}{\mu(B(s_\ell, \varepsilon))}}\, d\varepsilon.$$

Taking the supremum over all possible choices of s_ℓ finishes the proof. □

We now perform the main construction using conditions (4.1)–(4.4) to produce a sequence of partitions, which, in some sense, is not too large.

PROPOSITION 4.3. *Assume conditions* (4.1)–(4.4). *Assume that T has finite diameter, and consider the largest $i \in \mathbb{Z}$ such that* $\operatorname{diam} T \leq 2r^{-i}$. *Then one can find an increasing sequence of partitions $(\mathscr{A}_j)_{j \geq i}$ of T, and, for each $A \in \mathscr{A}_j$, one can find a natural number $\ell_j(A)$ such that the following properties hold*:

(4.7) *Each set of $\mathscr{A}_j$ has diameter at most $2r^{-j}$.*

(4.8) *Given $j \geq i$, any two sets A, B of $\mathscr{A}_{j+1}$ that are contained in the same element of $\mathscr{A}_j$, we have $\ell_{j+1}(A) \neq \ell_{j+1}(B)$.*

(4.9) $$\forall\, t \in T, \qquad \sum_{j \geq i} r^{-\beta j} \theta(\ell_{j+1}(A_{j+1}(t))) \leq 4S.$$

REMARK. One should observe that (4.3), (4.8) and (4.9) imply that each partition $\mathscr{A}_j$ is finite.

PROOF. Together with each set $A \in \mathscr{A}_j$, we will also construct a distinguished point $u_j(A) \in A$ such that

(4.10) $$\forall\, t \in A, \qquad d(t, u_j(A)) \leq r^{-j}.$$

This condition implies (4.7).

The construction proceeds by induction over j. For $j = i$, we set $\mathscr{A}_i = \{T\}$, $\ell_i(T) = 1$, and we choose $u_i(T)$ such that

$$\varphi_{i+2}(u_i(T)) \geq \sup\{\varphi_{i+2}(t);\ t \in T\} - \frac{S}{2}.$$

We now assume that the partition $\mathscr{A}_j$ has been constructed, as well as the points $u_j(A)$, $A \in \mathscr{A}_j$. To construct $\mathscr{A}_{j+1}$, it suffices to partition any given element A of $\mathscr{A}_j$ into elements of $\mathscr{A}_{j+1}$. This will be done by a kind of exhaustion argument that will produce pieces of A one at a time. The index of any piece will simply be the rank at which it is constructed.

At the first step, we pick $t_1 \in A$ such that

$$\varphi_{j+2}(t_1) \geq \sup\{\varphi_{j+2}(t);\ t \in A\} - 2^{i-j-1}S.$$

Our first piece of A is then

$$D_1 = A \cap B(t_1, r^{-j-1}).$$

We then repeat this procedure replacing A by $A \backslash D_1$, and we continue until A is exhausted. More formally, we construct $t_1, \dots, t_p$ in A such that

$$t_p \in A \backslash \bigcup_{\ell < p} B(t_\ell, r^{-j-1})$$

and

$$\varphi_{j+2}(t_p) \geq \sup\left\{\varphi_{j+2}(t);\ t \in A \backslash \bigcup_{\ell < p} B(t_\ell, r^{-j-1})\right\} - 2^{i-j-1}S, \tag{4.11}$$

and we set

$$D_p = B(t_p, r^{-j-1}) \cap \left(A \backslash \bigcup_{\ell < p} B(t_\ell, r^{-j-1})\right).$$

The sets $(D_p)_{p \geq 1}$ are disjoint by construction, and form the partition of A we wanted to construct. [It follows from (4.3) and (4.4) that the construction eventually stops.]

We set

$$u_{j+1}(D_p) = t_p, \qquad \ell_{j+1}(D_p) = p. \tag{4.12}$$

In particular, (4.8) holds. Thus, it remains to prove (4.9).

We observe that, by (4.10), we have $d(u_j(A), t_p) \leq r^{-j}$ for each p. Also, by construction, for $\ell < \ell'$ we have $d(t_\ell, t_{\ell'}) \geq r^{-j-1}$. Thus, by (4.4), for each p we have

$$\varphi_j(u_j(A)) \geq r^{-\beta j}\theta(p) + \min_{\ell \leq p} \varphi_{j+2}(t_\ell). \tag{4.13}$$

On the other hand, we have

$$t_p \in A \backslash \bigcup_{\ell < p} B(t_\ell, r^{-j-1}),$$

so that, by (4.11) (used for t_ℓ rather than t_p where $\ell \leq p$), we have

$$\varphi_{j+2}(t_\ell) \geq \varphi_{j+2}(t_p) - 2^{i-j-1}S$$

and thus, by (4.13),

$$\varphi_j(u_j(A)) \geq \varphi_{j+2}(t_p) + r^{-\beta j}\theta(p) - 2^{i-j-1}S. \tag{4.14}$$

Consider now $t \in D_p$. Then

$$A = A_j(t), \qquad D_p = A_{j+1}(t), \qquad \ell_{j+1}(A_{j+1}(t)) = \ell_{j+1}(D_p) = p,$$

so that (4.14) can be rewritten

$$\varphi_j(u_j(A_j(t))) \geq \varphi_{j+2}(t_p) + r^{-\beta j}\theta(\ell_{j+1}(A_{j+1}(t))) - 2^{i-j-1}S. \tag{4.15}$$

We now observe that

$$u := u_{j+2}(A_{j+2}(t)) \in A_{j+2}(t) \subset A_{j+1}(t) = D_p,$$

so that, by (4.10),

$$\varphi_{j+2}(t_p) \geq \varphi_{j+2}(u) - 2^{i-j-1}S$$

and combining with (4.15) gives

$$\varphi_j(u_j(A_j(t))) \geq \varphi_{j+2}(u_{j+2}(A_{j+2}(t))) + r^{-\beta j}\theta(\ell_{j+1}(A_{j+1}(t))) - 2^{i-j}S.$$

This holds for any t in T. We sum this relation for $j \geq i$ to get

$$\sum_{j\geq i} r^{-\beta j}\theta(\ell_{j+1}(A_{j+1}(t))) \leq 2S + \varphi_i(u_i(A_i(t))) + \varphi_{i+1}(u_{i+1}(A_{i+1}(t))) \leq 4S. \quad \square$$

Here is another version of Proposition 4.3.

PROPOSITION 4.4. *Consider a metric space (T, d). Assume that for $j \in \mathbb{Z}$ we are given a map ψ_j from T to $\mathbb{R}^+$. Assume that $\psi_j(t) \leq S$ for $j \in \mathbb{Z}$, $t \in T$, and that under* (4.3) *we have*

$$\max_{\ell\leq n} \psi_{j+2}(t_\ell) \geq \psi_j(s) + r^{-\beta j}\theta(n). \tag{4.4bis}$$

Then the conclusion of Proposition 4.3 *holds.*

PROOF. Use Proposition 4.3 for $\varphi_j(t) = S - \psi_j(t)$. $\square$

We now relate condition (4.9) with majorizing measures.

PROPOSITION 4.5. *Consider a metric space (T, d) and an increasing sequence $(\mathscr{A}_j)_{j\geq i}$ of partitions of T, and assume that, to each $A \in \mathscr{A}_j$, $j \geq i$,* *we associate a number $\ell_j(A) \in \mathbb{N}$ that satisfies* (4.8). *Then there is a probability measure μ on T such that, given α, $\beta > 0$, we have*

$$\sup_{t\in T}\sum_{j>i} r^{-\beta j}(\log(1/\mu(A_j(t))))^{1/\alpha} \le K(\alpha,\beta,r)\left(r^{-\beta i} + \sup_{t\in T}\sum_{j>i} r^{-\beta j}(\log \ell_j(A_j(t)))^{1/\alpha}\right), \tag{4.16}$$

where $K(\alpha,\beta,r)$ depends on α, β, r only.

REMARK. For now, we care only about the case $\alpha = 2$, $\beta = 1$.

PROOF. For $j \ge i$, $A \in \mathscr{A}_j$, we construct numbers $w_j(A)$ as follows.

We start the construction by setting $w_i(T) = 1$. Assuming that the numbers $w_{j-1}(A)$ have been constructed for $A \in \mathscr{A}_{j-1}$, if $B \in \mathscr{A}_j$, we set

$$w_j(B) = \frac{1}{4\ell_j(B)^2} w_{j-1}(A), \tag{4.17}$$

where A is the element of $\mathscr{A}_{j-1}$ that contains B. Since

$$\sum_{\ell\ge 1}\frac{1}{\ell^2} \le 2,$$

we see from (4.17) that

$$\sum\{w_j(B);\ B \in \mathscr{A}_j,\ B \subset A\} \le \tfrac{1}{2} w_{j-1}(A).$$

Thus, by induction over j, we see that

$$\sum_{A\in\mathscr{A}_j} w_j(A) \le 2^{-j+i}$$

and thus

$$\sum_{j>i}\sum_{A\in\mathscr{A}_j} w_j(A) \le 1.$$

It follows that there exists a probability μ on T such that

$$\forall\ j \ge i,\ \forall\ A \in \mathscr{A}_j, \qquad \mu(A) \ge w_j(A).$$

Consider now $t \in T$, $j > i$. Then, by (4.17), we have

$$w_j(A_j(t)) = 4^{i-j}\prod_{i<k\le j} \ell_k(A_k(t))^{-2},$$

so that

$$\log\frac{1}{\mu(A_j(t))} \le (j-i)\log 4 + 2\sum_{i<k\le j}\log \ell_k(A_k(t)). \tag{4.18}$$

We first consider the case $\alpha \geq 1$. This is the most important and also the easiest, because in this case, setting $\delta = 1/\alpha$, we have $(x+y)^\delta \leq x^\delta + y^\delta$, so that

$$\left(\log \frac{1}{\mu(A_j(t))}\right)^\delta \leq K(j-i)^\delta + 2^\delta \sum_{i<k\leq j} (\log \ell_k(A_k(t)))^\delta. \tag{4.19}$$

Thus, by changing the order of summations,

$$\sum_{j>i} r^{-\beta j}\left(\log \frac{1}{\mu(A_j(t))}\right)^\delta \leq K \sum_{j>i} r^{-j\beta}(j-i)^\delta + 2^\delta \sum_{k>i}\left(\sum_{j\geq k} r^{-j\beta}\right)(\log \ell_k(A_k(t)))^\delta,$$

which finishes the proof.

In the case $\alpha < 1$ (which should be omitted at a first reading), we need a substitute for (4.19). We observe that the function $x \to x^\delta$ is convex, so that, for numbers $y_k \geq 0$ and numbers $a_k \geq 0$, with $\sum a_k = 1$, we have

$$\left(\sum a_k y_k\right)^\delta \leq \sum a_k y_k^\delta.$$

Thus, taking $a_k = a r^{\beta(k-j)/2\delta}$, where $a^{-1} = \sum_{\ell \geq 1} r^{-\beta\ell/2\delta}$, we see that (setting $y_k = x_k/a_k$)

$$\left(\sum_{i<k\leq j} x_k\right)^\delta \leq \sum_{i<k\leq j} a_k a_k^{-\delta} y_k^\delta$$

and thus

$$\left(\log \frac{1}{\mu(A_j(t))}\right)^\delta \leq K(\alpha,\beta,r)\left[(j-i)^\delta + \sum_{i<k\leq j} r^{\beta(j-k)/2}\left(\log \frac{1}{\mu(A_j(t))}\right)^\delta\right]$$

and the proof is then finished as before. □

PROOF OF PROPOSITION 2.4. Combining Propositions 4.2 and 4.5, we obtain a bound

$$K(r)(r^{-i} + S)$$

for the left-hand side of (4.5). To prove that $r^{-i} \leq K(r)S$, one then uses an argument similar to the one given after (2.11).

5. Gaussian processes. In this section we go back to general Gaussian processes $(X_t)_{t\in T}$. There is no special advantage in considering the "canonical" setting where T is a subset of ℓ^2, so here $(X_t)_{t\in T}$ is simply a jointly Gaussian family of centered r.v.'s indexed by T. We provide the index set T with the distance (1.1).

THEOREM 5.1 (The majorizing measure theorem). *For some universal constant K, we have*

$$\frac{1}{K}\gamma_2(T,d) \leq E \sup_{t\in T} X_t \leq K\gamma_2(T,d). \tag{5.1}$$

This statement characterizes the sample boundedness of the process as a function of the geometry of (T, d) (and results about continuity follow easily).

We observe that Theorem 5.1 does not say how to evaluate the quantity (5.1) in concrete examples. Rather, the content of Theorem 5.1 is that there is no other way to bound $E\sup_{t\in T} X_t$ than to find a good majorizing measure on T.

PROOF OF THEOREM 5.1. The right-hand inequality is a special case of Proposition 2.4. To prove the left-hand inequality, by Theorem 5.1, it suffices to prove that the functions $\varphi_j(t)$ on T given by

$$\varphi_j(t) = E\sup\{X_u;\ d(u,t) \leq 2r^{-j}\}$$

satisfy (4.4) for

$$\theta(n) = \frac{1}{K}\sqrt{\log n}, \qquad \beta = 1,$$

provided r is large enough.

There are two key ingredients to this proof.

LEMMA 5.2 (Sudakov minoration). *Assume that*

$$\ell, \ell' \leq n, \qquad \ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a.$$

Then we have, for some universal constant K_1,

$$E\sup_{\ell\leq n} X_{t_\ell} \geq \frac{1}{K_1} a\sqrt{\log n}. \tag{5.2}$$

We refer to [10], page 83, for a proof. In what follows, the suprema of collections of r.v.'s are essential suprema.

LEMMA 5.3. *Consider a Gaussian process* $(Z_t)_{t\in V}$. *Let* $\sigma = \sup_{t\in V}\|Z_t\|_2$. *Then, for some universal constant* K_2, *we have*

$$P\left(\left|\sup_{t\in V} Z_t - E\sup_{t\in V} Z_t\right| \geq K_2 u\sigma\right) \leq 2e^{-u^2}. \tag{5.3}$$

This is a very important property of Gaussian processes. We purposely do not give a sharp form of (5.3), but rather what makes the proof work; this is important for further extensions of the method, where sharp forms will not be available. This property follows either from the deviation inequality of Ibragimov, Sudakov and Tsirelson [8] or the Gaussian isoperimetric inequality.

LEMMA 5.4. *Consider points* $(t_\ell)_{\ell\leq n}$. *Assume that* $\ell, \ell' \leq n$, $\ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a$. *Consider* $\sigma > 0$ *and for* $\ell \leq n$ *consider a set* $A_\ell \subset B(t_\ell, \sigma)$. *Then, setting* $A = \cup_{\ell\leq n} A_\ell$, *we have*

$$E\sup_{t\in A} X_t \geq \frac{a}{K_1}\sqrt{\log n} - \sigma K_3\sqrt{\log n} + \min_{\ell\leq n} E\sup_{t\in A_\ell} X_t.$$

PROOF. For $\ell \le n$, we consider

$$Y_\ell = \sup_{t\in A_\ell}(X_t - X_{t_\ell}) = \Big(\sup_{t\in A_\ell} X_t\Big) - X_{t_\ell}. \tag{5.4}$$

We now apply (5.3) to $Z_t = X_t - X_{t_\ell}$, $V = A_\ell$, so that

$$P(|Y_\ell - EY_\ell| \ge K_2 u\sigma) \le 2e^{-u^2}.$$

Setting

$$h = \max_{\ell \le n} |Y_\ell - EY_\ell|,$$

we then have, by (5.3),

$$P(h \ge K_2 u\sigma) \le 2ne^{-u^2}.$$

Using the fact that $Eh = \int_0^\infty P(h \ge v)\,dv$ and a routine computation, we see that

$$Eh \le K_3 \sigma\sqrt{\log n}, \tag{5.5}$$

where K_3 is universal. Now, for each $\ell \le n$,

$$Y_\ell \ge EY_\ell - h \ge \min_{\ell\le n} EY_\ell - h.$$

Thus

$$\sup_{t\in A_\ell} X_t = Y_\ell + X_{t_\ell} \ge X_{t_\ell} + \min_{\ell \le n} EY_\ell - h$$

and thus

$$\sup_{t\in A} X_t \ge \max_{\ell\le n} X_{t_\ell} + \min_{\ell\le n} EY_\ell - h.$$

Taking expectation, using Lemma 5.2 and (5.5) yield the result. □

It is possible to give a proof of Lemma 5.4 based on the comparison properties of Gaussian processes [7]. This is, however, missing the main point. Comparison properties are quite specific to Gaussian processes, while general principles used in the present approach are much more general, and consequently, allow for considerable extensions of Theorem 5.1 (see [18] and [20]).

PROOF OF (4.4). We use Lemma 5.4 with

$$A_\ell = \{u \in T;\ d(u, t_\ell) \le 2r^{-j-2}\}.$$

Since $d(s, t_\ell) \le r^{-j}$ and since $r \ge 2$, we have $A_\ell \subset B(s, 2r^{-j})$, so that, by the definition of φ_j and Lemma 5.3, we have

$$\varphi_j(A) \ge \frac{r^{-j-1}}{K_1}\sqrt{\log n} - 2r^{-j-2}K_3\sqrt{\log n} + \min_{\ell \ge n}\varphi_j(t_\ell).$$

Thus (4.4) holds for

$$\theta(n) = \frac{1}{2rK_1}\sqrt{\log n},$$

provided $r \geq 4K_1K_2$. □

6. The ellipsoid theorem and matchings. In this section we will use the partitioning scheme of Section 4 and geometry to explain the structure of ellipsoids in Hilbert space with respect to majorizing measures. The geometry will occur in the form of convexity properties, an idea that will be used again in the next section. We will then outline why the structure of ellipsoids is the key to deep matching theorems.

For purposes that will become apparent later, it is of importance to consider functionals related to $\gamma_2(T, d)$ but where the integral condition is replaced by a more general one.

Consider a metric space (T, d), and $\alpha, \beta > 0$. We define $\gamma_{\alpha,\beta}(T, d)$ as the infimum over all choices of the probability measure μ on T of the quantity

$$\sup_{t\in T}\left(\int_0^\infty \varepsilon^\beta\left(\log\frac{1}{\mu(B(t,\varepsilon))}\right)^{\beta/\alpha}\frac{d\varepsilon}{\varepsilon}\right)^{1/\beta}. \tag{6.1}$$

Thus $\gamma_{2,1}(T, d)$ coincides with the functional $\gamma_2(T, d)$. Consider the ellipsoid of ℓ^2 defined by (3.2). The purpose of the present section is to show how to compute $\gamma_{\alpha,\beta}(\mathscr{E})$. (We do not mention the distance when it is naturally induced by Hilbert space.) In the case $\alpha = 2$, $\beta = 1$, it follows from (3.5) and Theorem 5.1 that

$$\gamma_2(\mathscr{E}) \leq K\left(\sum_{n\geq 1} a_n^2\right)^{1/2}. \tag{6.2}$$

We are, however, mostly interested in the case $\beta = 2$, for which we must find an entirely different proof. This proof will be mostly geometric. The geometric proof will also yield (6.2), and thereby a satisfactory understanding of ellipsoids.

For $x \in \ell^2$, we set $\|x\|_{\mathscr{E}} = (\sum_{n\geq 1} a_n^2 x_n^2)^{1/2}$.

LEMMA 6.1. *We have*

$$\|x\|_{\mathscr{E}}, \|y\|_{\mathscr{E}} \leq 1 \Rightarrow \frac{\|x+y\|_{\mathscr{E}}}{2} \leq 1 - \frac{\|x-y\|_{\mathscr{E}}^2}{8}. \tag{6.3}$$

PROOF. By the parallelogram identity we have

$$\|x-y\|_{\mathscr{E}}^2 + \|x+y\|_{\mathscr{E}}^2 = 2\|x\|_{\mathscr{E}}^2 + 2\|y\|_{\mathscr{E}}^2 \leq 4,$$

so that

$$\|x+y\|_{\mathscr{E}}^2 \leq 4 - \|x-y\|_{\mathscr{E}}^2$$

and

$$\|x+y\|_{\mathscr{E}} \le 2\left(1-\frac{1}{4}\|x-y\|_{\mathscr{E}}^2\right)^{1/2} \le 2\left(1-\frac{\|x-y\|_{\mathscr{E}}^2}{8}\right). \qquad \square$$

As it turns out, (6.3) is the *only* property of ellipsoids that will be important to us. This motivates the following definition.

DEFINITION 6.2. A norm $\|\cdot\|$ in a Banach space X is called 2-convex if, for a certain number γ, it satisfies

$$\|x\|, \|y\| \le 1 \Rightarrow \left\|\frac{x+y}{2}\right\| \le 1-\gamma\|x-y\|^2. \tag{6.4}$$

We now consider the following setting. The set T is the unit ball of a Banach space W for a 2-convex norm $\|\cdot\|$. The distance d on T is induced by another norm $\|\cdot\|_V$, that is, $d(x, y) = \|x-y\|_V$, where V is the unit ball of $\|\cdot\|_V$.

THEOREM 6.3. *In the setting above, for any $\alpha > 0$,*

$$\gamma_{\alpha,2}(T, d) \le K(\alpha, \gamma) \sup_{\varepsilon>0} \varepsilon(\log N(T, d, \varepsilon))^{1/\alpha}.$$

To understand this result, one should observe that it is always true that

$$\gamma_{\alpha,\beta}(T, d) \le K(\alpha, \beta)\left(\int_0^\infty \varepsilon^\beta \log N(T, d, \varepsilon)^{\beta/\alpha}\frac{d\varepsilon}{\varepsilon}\right)^{1/\beta}.$$

This fact is an extension of the well-known fact (which simple proof we will not give) that the integral of (2.13) dominates $\gamma_2(T, d)$. However, the bound of Theorem 6.3 is quite a bit smaller than the previous integral. In fact, it is always true that

$$\sup_{\varepsilon>0} \varepsilon(\log N(T, d, \varepsilon))^{1/\alpha} \le K(\beta)\gamma_{\alpha,\beta}(T, d).$$

To see this, we note that if for some number $\eta > 0$ the balls $B(t_\ell, \eta)$ are disjoint ($\ell \le N$) and if μ is a probability measure on T, then for some $\ell \le N$ we have $\mu(B(t_\ell, \eta)) \le 1/N$.

Then

$$\begin{aligned}\left(\int_0^\infty \varepsilon^\beta \left(\log\frac{1}{\mu(B(t_\ell, \varepsilon))}\right)^{\beta/\alpha}\frac{d\varepsilon}{\varepsilon}\right)^{1/\beta} &\ge \left(\int_0^\eta \varepsilon^\beta(\log N)^{\beta/\alpha}\frac{d\varepsilon}{\varepsilon}\right)^{1/\beta}\\ &= K(\beta)\eta(\log N)^{1/\alpha}.\end{aligned}$$

The content of Theorem 6.3 is that this inequality can be reversed under the conditions of this theorem.

COROLLARY 6.4. *If $\mathscr{E}$ is the ellipsoid* (3.2), *then*

$$\gamma_{\alpha,2}(\mathscr{E}) \le K(\alpha) \sup_n a_n n^{1/\alpha}.$$

PROOF. We use Theorem 6.3 with

$$\|x\| = \|x\|_{\mathscr{E}} = \left(\sum_{n\geq 1} a_n x_n^2\right)^{1/2}, \qquad V = \{x \in \ell^2;\ \|x\| \leq 1\},$$

so that $\|x\|_V$ is the norm of x in ℓ^2. We observe that, if $B = \sup_n a_n n^{1/\alpha}$, then $a_n \leq Bn^{-1/\alpha}$, so that, with the notation of Lemma 3.2, we have $2^{-k} \leq Bn_k^{-1/\alpha}$, so that $n_k \leq (2^k B)^\alpha$, and, by Lemma 3.2,

$$(\log N(\mathscr{E}, 2^{-k+1}))^{1/\alpha} \leq K(\alpha) 2^k B$$

and thus

$$\sup_{\varepsilon>0} \varepsilon(\log N(T, d, \varepsilon))^{1/\alpha} \leq K(\alpha)B. \qquad \square$$

To prove Theorem 6.3, we will use Proposition 4.4 with the functionals

$$\psi_j(t) = \inf\{\|v\|;\ d(t, v) \leq 2r^{-j}\}.$$

It is obvious that (4.1) holds for $S = 1$.

For $n \geq 2$, we set

$$\varepsilon(n) = \sup\{\varepsilon > 0;\ \exists\, t_1, \dots, t_n \in T,\ \forall\, \ell, \ell',\ 1 \leq \ell < \ell' \leq n,\ d(t_\ell, t_{\ell'}) > \varepsilon\}.$$

We observe the following simple relations:

$$\varepsilon < \frac{\varepsilon(n)}{2} \Rightarrow N(T, d, \varepsilon) \geq n, \tag{6.5}$$

$$\varepsilon > \varepsilon(n) \Rightarrow N(T, d, \varepsilon) \leq n. \tag{6.6}$$

The center of the proof is to establish condition (4.4bis) with an appropriate choice of θ. This is the object of the following lemma.

LEMMA 6.5. *Assume $r = 8$. Consider $s \in T$, $n \geq 2$, points $t_1, \dots, t_n$ in $B(s, r^{-j})$ such that*

$$1 \leq \ell < \ell' \leq n \Rightarrow d(t_\ell, t_{\ell'}) \geq r^{-j-1}.$$

Then

$$\sup_{\ell \leq n} \psi_{j+2}(t_\ell) \geq \psi_j(s) + \frac{\gamma r^{-2j}}{(2r\varepsilon(n))^2}. \tag{6.7}$$

PROOF. Consider $u > \sup_{\ell \leq n} \psi_{j+2}(t_\ell)$. Then, for $\ell \leq n$, by the definition of ψ_{j+2}, we can find a point $w_\ell \in uT$ with $d(w_\ell, t_\ell) \leq 2r^{-j-2}$. Thus, if $\ell < \ell'$ and since $r = 8$,

$$d(w_\ell, w_{\ell'}) \geq d(t_\ell, t_{\ell'}) - 4r^{-j-2} \geq r^{-j-1} - 4r^{-j-2} = r^{-j-1}/2. \tag{6.8}$$

Also,

$$d(s, w_\ell) \leq d(s, t_\ell) + d(t_\ell, w_\ell) = r^{-j} + 2r^{-j-2} \leq 2r^{-j}. \tag{6.9}$$

We now use (6.4) for $x = w_\ell/u,\ \ y = w_{\ell'}/u$, to get

$$\left\|\frac{w_\ell + w_{\ell'}}{2}\right\| \le u\left(1 - \frac{\gamma}{u^2}\|w_\ell - w_{\ell'}\|^2\right). \tag{6.10}$$

Now (6.9) means that for each ℓ we have $w_\ell \in s + 2r^{-j}V$. The convexity of V implies

$$\frac{w_\ell + w_{\ell'}}{2} \in s + 2r^{-j}V$$

for any $\ell, \ell' \le n$, so that, by definition,

$$\psi_j(s) \le \left\|\frac{w_\ell + w_{\ell'}}{2}\right\|.$$

Combining with (6.10), we see that, setting

$$R^2 = \frac{u}{\gamma}(u - \psi_j(s)),$$

we have

$$\|w_\ell - w_{\ell'}\|^2 \le R^2.$$

Thus, in particular, the points $x_\ell = (w_\ell - w_{\ell'})/R$ belong to T.

If we recall that d comes from a norm, by (6.8) we have

$$1 \le \ell < \ell' \le n \Rightarrow d(x_\ell, x_{\ell'}) \ge r^{-j-1}/2R.$$

Thus, by definition, we have $r^{-j-1}/2R \le \varepsilon(n)$, that is, by the definition of R,

$$\gamma\left(\frac{r^{-j-1}}{2\varepsilon(n)}\right)^2 \le u(u - \psi_j(s)).$$

Taking the infimum over u and realizing that $\psi_{j+2} \le 1$, we get the result. □

COROLLARY 6.6. *There exist an increasing sequence of partitions* $(\mathscr{A}_j)_{j\ge 0}$ *of* T *that satisfies* (4.7) *and indexes* $\ell_j(A)$ *that satisfy* (4.8) *such that*

$$\forall\, t \in T, \qquad \sum_{j\ge 0} \frac{r^{-2j}}{\varepsilon(\ell_{j+1}(A_{j+1}(t)))^2} \le \frac{16r^2}{\gamma}, \tag{6.11}$$

where we make the convention that $1/\varepsilon(1) = 0$.

PROOF. We apply Proposition 4.4, noticing that $i = 0$.

PROOF OF THEOREM 6.3. If $\varepsilon(\log N(T, d, \varepsilon))^{1/\alpha} \le S$ for each $\varepsilon > 0$, then by (6.5) we have $\varepsilon(n)(\log n)^{1/\alpha} \le 2S$, so that (6.11) implies

$$\forall\, t \in T, \qquad \sum_{j\ge 0} r^{-2j}(\log \ell_{j+1}(A_{j+1}(t)))^{2/\alpha} \le \frac{32S^2r^2}{\gamma}.$$

To conclude the proof, we appeal to Proposition 4.5. □

We have singled out Theorem 6.3 because it has significant applications. However, we can prove much more. Here is another statement.

THEOREM 6.7. *Under the conditions of Theorem* 6.3, *we have*

$$\gamma_2(T,d) \le K(\gamma)\left(\int_0^\infty \varepsilon \log N(T,d,\varepsilon)\, d\varepsilon\right)^2.$$

REMARK 1. When specialized to ellipsoids, this proves (3.9).

REMARK 2. Theorems 6.3 and 6.7 are special cases of a general result (with essentially the same proof). See [21].

PROOF. Consider a sequence of partitions such as in Corollary 6.5. In order to apply Proposition 4.5, we try to bound, for a given $t \in T$, the quantity

$$S = \sum_{j\ge 0} r^{-j}\sqrt{\log \ell_{j+1}(A_{j+1}(t))}.$$

Consider, for $k \ge 0$, the smallest integer m_k such that $\varepsilon(m_k) < 2^{-k}$. Consider, for $k \ge 1$, the set

$$I_k = \{j \ge 0;\ m_k \le \ell_{j+1}(A_{j+1}(t)) < m_{k+1}\}.$$

When I_k is not empty, denote by $j(k)$ its smallest element. Then

$$\begin{aligned}\sum_{j\in I_k} r^{-j}\sqrt{\log \ell_{j+1}(A_{j+1}(t))} &\le \sqrt{\log m_{k+1}}\sum_{j\in I_k} r^{-j}\\ &\le 2r^{-j(k)}\sqrt{\log m_{k+1}}\end{aligned}$$

and thus

$$S \le 2\sum_{k\ge 0} r^{-j(k)}\sqrt{\log m_{k+1}},$$

where we make the convention that only the terms for which $I_k \ne \varnothing$ appear. Now, by Cauchy–Schwarz,

$$S \le 2\left(\sum_{k\ge 0}\frac{r^{-2j(k)}}{2^{-2k}}\right)^{1/2}\left(\sum_{k\ge 0} 2^{-2k}\log m_{k+1}\right)^{1/2}.$$

We have, since $j(k) \in I_k$,

$$\varepsilon(\ell_{j(k)+1}(A_{j(k)+1}(t))) \le \varepsilon(m_k) \le 2^{-k},$$

so that, by (6.11),

$$\sum_{k>0}\frac{r^{-2j(k)}}{2^{-2k}} \le \sum_{j\ge 0}\frac{r^{-2j}}{\varepsilon(\ell_{j+1}(A_{j+1}(t)))^2} \le \frac{16r^2}{\gamma}.$$

Also, by the definition of m_k, we have $\varepsilon(m_k - 1) \geq 2^{-k}$, so that by (6.5) we have

$$N(T, d, 2^{-k-1}) \geq m_k - 1,$$

so that

$$\log m_{k+1} \leq \log(1 + N(T, d, 2^{-k-2})).$$

It then follows easily that

$$\sum_{k \geq 0} 2^{-2k} \log m_{k+1} \leq K \int_0^\infty \varepsilon \log N(T, d, \varepsilon)\, d\varepsilon.$$

This finishes the proof. □

We now turn to the application of Theorem 6.3 to the Ajtai–Komlòs–Tusnàdy matching theorem. Consider points $(X_i)_{i \leq n}$ that are independently uniformly distributed over $[0, 1]^2$. The problem is to understand the "transportation cost" C_n from the empirical measure $\delta_X = (1/n)\sum_{i \leq n} \delta_{X_i}$ to the uniform measure on $[0, 1]^2$, when the cost of moving a unit mass is simply the distance by which it travels. Through duality (i.e., the Hahn–Banach theorem) this cost is

$$C_n = \sup_{f \in \mathscr{L}} \left| \frac{1}{n} \sum_{i \leq n} f(X_i) - Ef \right|, \tag{6.12}$$

where $\mathscr{L}$ is the class of Lipschitz functions on $[0, 1]^2$, that is, functions that satisfy

$$|f(x) - f(y)| \leq d(x, y)$$

for all x, y in $[0, 1]^2$, and where Ef is the average of f over the unit square. (For a simple proof of (6.12), one can, e.g., see [22], where the link between transportation cost and matching is explained.)

THEOREM 6.8 (Ajtai, Komlòs and Tusnàdy [2]). $EC_n \leq K\sqrt{\log n / n}$.

We will only sketch the proof, and we refer the reader to [19], Section 4, for the missing details. Consider a sequence $(\varepsilon_i)_{i \leq n}$ of independent Bernoulli r.v.'s $[P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = \frac{1}{2}]$. The first step of the proof is to establish that

$$\begin{aligned} EC_n &\leq 2E \sup_{f \in \mathscr{L}_0} \left| \frac{1}{n} \sum_{i \leq n} \varepsilon_i f(X_i) \right|. \\ &= 2E\left(E_\varepsilon \sup_{f \in \mathscr{L}_0} \left| \frac{1}{n} \sum_{i \leq n} \varepsilon_i f(X_i) \right| \right), \end{aligned}$$

where E_ε is conditional expectation given $(X_i)_{i \leq n}$ and $\mathscr{L}_0$ is $\{f \in \mathscr{L};\ Ef = 0\}$.

Next, we recall the "sub-Gaussian inequality" ([10], page 90)

$$P\left(\left|\sum_{i\le n}\varepsilon_i a_i\right|\ge u\right)\le 2\exp\left(-\frac{u^2}{2\sum_{i\le n}a_i^2}\right). \tag{6.13}$$

Consider the random distance d_U on $\mathscr{L}_0$ given by

$$d_U(f,g)^2=\frac{1}{n}\sum_{i\le n}(f(X_i)-g(X_i))^2.$$

We see that if we set

$$X_f=\sum_{i\le n}\varepsilon_i f(X_i),$$

we have to estimate

$$\frac{1}{n}E_\varepsilon\sup_{f\in\mathscr{L}_0}|X_f|=\frac{1}{n}E_\varepsilon\sup_{f\in\mathscr{L}_0}X_f$$

by symmetry.

Now, by (6.13),

$$P(|X_f-X_g|\ge u)\le 2\exp\left(-\frac{u^2}{2nd_U^2(f,g)}\right),$$

so that we are essentially in the situation of Section 1. Before using Proposition 2.4, we have to take care of a few details. First, we need to control the random distance d_U. This is done through the following easy lemma.

LEMMA 6.9 (Talagrand [19]). *There exists a random variable R with $ER\le K$ such that*

$$\forall f,g\in\mathscr{L},\qquad d_U(f,g)\le R\left(\|f-g\|_2+\sqrt{\frac{\log n}{n}}\right). \tag{6.14}$$

Consider then a subset Z of $\mathscr{L}_0$ that is maximal for the property

$$\forall f,g\in Z,\qquad \|f-g\|_2\ge\sqrt{\frac{\log n}{n}}. \tag{6.15}$$

Then, by maximality, given f in $\mathscr{L}_0$, there exists g in Z with $\|f-g\|_2\le\sqrt{\log n/n}$. Thus, by (6.14), we get $d_U(f,g)\le 2R\sqrt{\log n/n}$ and by Cauchy–Schwarz

$$\begin{aligned}\frac{1}{n}\left|\sum_{i\le n}\varepsilon_i(f(X_i)-g(X_i))\right|&\le\frac{1}{n}\sum_{i\le n}|f(X_i)-g(X_i)|\\&\le\left(\frac{1}{n}\sum_{i\le n}|f(X_i)-g(X_i)|^2\right)^{1/2}\\&\le d_U(f,g)\le 2R\sqrt{\frac{\log n}{n}}.\end{aligned}$$

Thus we obtain

$$\frac{1}{n}\sup_{f\in\mathscr{L}_0}\left|\sum_{i\le n}\varepsilon_i f(X_i)\right| \le 2R\sqrt{\frac{\log n}{n}} + \frac{1}{n}\sup_{f\in Z}\left|\sum_{i\le n}\varepsilon_i f(X_i)\right|,$$

and it suffices to prove that

$$\frac{1}{n}E_\varepsilon \sup_{f\in Z}\left|\sum_{i\le n}\varepsilon_i f(X_i)\right| \le K\sqrt{\frac{\log n}{n}}.$$

We observe that, by (6.14) and (6.15), we have

$$\forall\ f,g\in Z,\qquad d_U(f,g)\le 2R\|f-g\|_2. \tag{6.16}$$

Consider a probability measure μ on Z. Then, in order to use Proposition 2.4, we compute

$$I(f)=\int_0^\infty \sqrt{\log\frac{1}{\mu(B_U(f,\varepsilon))}}\,d\varepsilon,$$

where B_U denotes the ball for d_U. By (6.16) we have $B_U(f,\varepsilon)\cap Z \supset B(f,\varepsilon/2R)\cap Z$, where the latter ball is for the L^2-distance. Thus

$$I(f)\le \int_0^\infty \sqrt{\log\frac{1}{\mu(B(f,\varepsilon/2R))}}\,d\varepsilon = 2R\int_0^K \sqrt{\log\frac{1}{\mu(B(f,\varepsilon))}}\,d\varepsilon$$

by a change of variable and since the diameter of $\mathscr{L}_0$ is less than or equal to K.

Setting $a_n=\sqrt{\log n/n}$, for $\varepsilon<a_n$, we have $B(f,\varepsilon)=\{f\}$ by (6.15), so that

$$\frac{I(f)}{2R}\le a_n\sqrt{\log\frac{1}{\mu(\{f\})}}+\int_{a_n}^K\sqrt{\log\frac{1}{\mu(B(f,\varepsilon))}}\,d\varepsilon.$$

To evaluate the latter integral, we write

$$\begin{aligned}\int_{a_n}^K\sqrt{\log\frac{1}{\mu(B(f,\varepsilon))}}\,d\varepsilon &\le \left(\int_{a_n}^K\frac{d\varepsilon}{\varepsilon}\right)^{1/2}\left(\int_{a_n}^K\varepsilon\log\frac{1}{\mu(B(f,\varepsilon))}\,d\varepsilon\right)^{1/2}\\ &\le \left(\log\frac{K}{a_n}\right)^{1/2}\left(\int_0^\infty\varepsilon\log\frac{1}{\mu(B(f,\varepsilon))}\,d\varepsilon\right)^{1/2}\\ &\le K\sqrt{\log n}\left(\int_0^\infty\varepsilon\log\frac{1}{\mu(B(f,\varepsilon))}\,d\varepsilon\right)^{1/2}.\end{aligned}$$

Now,

$$\left(\int_0^{a_n}\varepsilon\log\frac{1}{\mu(B(f,\varepsilon))}\,d\varepsilon\right)^{1/2}\ge\sqrt{\log\frac{1}{\mu(\{f\})}}\left(\int_0^{a_n}\varepsilon\,d\varepsilon\right)^{1/2}\ge\frac{1}{K}a_n\sqrt{\log\frac{1}{\mu(\{f\})}},$$

so that

$$I(f)\le KR\sqrt{\log n}\,J(f),$$

where

$$J(f) = \left(\int_0^{\infty} \varepsilon \log \frac{1}{\mu(B(f, \varepsilon))} \, d\varepsilon \right)^{1/2}.$$

What the previous discussion shows when combined with Proposition 2.4 is that

$$\frac{1}{n} E \sup_{f \in Z} \left| \sum_{i \le n} \varepsilon_n f(X_n) \right| \le K \sqrt{\log n} \, \gamma_{2,2}(Z).$$

There is a simple argument showing that $\gamma_{2,2}(U) \le K \gamma_{2,2}(T)$ if $U \subset T$, so the only concern now is to prove that $\gamma_{2,2}(\mathscr{L}_0) < \infty$.

This will be done by using the Fourier transform, proving that $\mathscr{L}_0$ is a subset of an ellipsoid to which Theorem 6.3 applies. For a function f in $L_2([0, 1]^2)$, we define

$$a_{n,m}(f) = \iint_{[0,1]^2} f(x, y) \exp 2i\pi(nx + my) \, dx dy,$$

so that, by the Planchel theorem,

$$\|f\|_2 = \left(\sum_{n,m \in \mathbb{Z}} |a_{n,m}(f)|^2 \right)^{1/2}.$$

Thus it should be obvious that it suffices to prove that $\gamma_{2,2}(T') < 0$, where T' is the set of sequences $\mathscr{F}(f) = (a_{n,m}(f))_{n,m \in \mathbb{Z}}$ for $f \in \mathscr{L}_0$. To avoid messy details, but keep the central point, we will prove the weaker result that $\gamma_{2,2}(T) < \infty$, where T is the set of sequences $\mathscr{F}(f)$ for $f \in \mathscr{L}_1$, where $\mathscr{L}_1$ is the set of functions of $\mathscr{L}_0$ that are 0 on the boundary of $[0, 1]^2$. Then integration by parts yields

$$a_{n,m}(f) = -\frac{1}{2\pi i n} a_{n,m}\left(\frac{\partial f}{\partial x} \right).$$

Since f is Lipschitz, $|\partial f / \partial x| \le 1$, so that, in particular,

$$\sum_{n,m \in \mathbb{Z}} \left| a_{n,m}\left(\frac{\partial f}{\partial x} \right) \right|^2 = \left\| \frac{\partial f}{\partial x} \right\|_2^2 \le 1.$$

Thus $\sum_{n,m \in \mathbb{Z}} n^2 |a_{n,m}(f)|^2 \le 1$. A similar argument yields $\sum_{n,m \in \mathbb{Z}} m^2 |a_{n,m}(f)|^2 \le 1$. Thus

$$a \in T \Rightarrow \sum_{n,m \in \mathbb{Z}} (n^2 + m^2) |a_{n,m}|^2 \le 1.$$

That is, T is contained in a certain ellipsoid.

For $k \ge 0$, the number of values of n, m such that $m^2 + n^2 \le 2^{2k}$ is crudely at most $(2^{k+1} + 1)^2$, since $|n|, |m| \le 2^k$, so it is at most 2^{2k+4}. Relabeling the coordinates and splitting into real and imaginary parts, we view T as a subset of the set of (real) sequences $(x_n)_{n \ge 1}$ such that $\sum n x_n^2 \le K$, and this concludes the proof by Theorem 6.3.

It can be argued that the previous approach is not superior to the original "transportation method" of [2]. It is, however, more "generic" and thus can be used in many situations, some of which are described in [19].

7. Restriction of operators. One towering success of the "probabilistic method" is Bourgain's [3] construction of $\Lambda(p)$-sets of large density. A new approach, based on the ideas of the present paper, has made this result more accessible (see [23]). Several issues related to the $\Lambda(p)$-set problem are not connected with the present circle of ideas, so, rather than discussing this problem, we will concentrate on another result that is actually the essential step of the proof given in [23]. The natural setting of this result is the class of 2-smooth Banach spaces. The reader who wishes to avoid the minimal effort of understanding what this means can assume instead that all Banach spaces are Hilbert spaces; the proofs are identical.

We say that a Banach space W is 2-smooth if, given two vectors x, y in W, with $\|x\| = 1$, we have

$$\frac{1}{2}(\|x+y\| + \|x-y\|) \leq 1 + C\|y\|^2, \tag{7.1}$$

where C is independent of x, y. We will never use (7.1) directly. Rather we will use the following classical facts (see Lindenstrauss and Tzafriri [11]).

FACT 1. If W is 2-smooth, it is of type 2; that is, for each n, each sequence $(Y_i)_{i\leq n}$ of centered, independent r.v.'s valued in W, we have

$$E\left\|\sum_{i\leq n} Y_i\right\|^2 \leq K(C)\sum_{i\leq n} E\|Y_i\|^2. \tag{7.2}$$

There, as well as in the remainder of this section, we denote by $K(C)$ a number that depends only on the number C of (7.1).

FACT 2. If W is 2-smooth, its dual W^* is 2-convex; that is, if $x^*, y^* \in W^*$, $\|x^*\|, \|y^*\| \leq 1$, then

$$\left\|\frac{x^* + y^*}{2}\right\| \leq 1 - \frac{1}{K(C)}\|x^* - y^*\|^2. \tag{7.3}$$

Consider now an operator U from ℓ_m^2 to W. We denote by $(e_i)_{i\leq m}$ the canonical basis of ℓ_m^2. Consider a number $0 < \delta < 1$. We select a random subset I of $\{1, \dots, m\}$ as follows. For each $i \leq m$, we include (independently) i in I with probability δ. In other words, if $(\delta_i)_{i\leq m}$ are independent, with $E\delta_i = \delta$, $\delta_i \in \{0, 1\}$, then $I = \{i \leq m;\ \delta_i = 1\}$. We consider the subspace ℓ_I of ℓ_m^2 generated by the vectors $(e_i)_{i\in I}$. In other words, ℓ_I identifies with the sequences $(x_i)_{i\leq m}$ for which $x_i = 0$ whenever $i \notin I$. We consider the restriction U_I of the operator U to ℓ_I.

THEOREM 7.1.

$$E\|U_I\| \le \frac{K(C)}{\sqrt{\log(1/\delta)}}\Big(\|U\| + \sup_{i\le m}\|U(e_i)\|\sqrt{\log i}\Big).$$

Before we discuss the nature of Theorem 7.1, we explain why even the weaker form

$$E\|U_I\| \le \frac{K(C)}{\sqrt{\log(1/\delta)}}\Big(\|U\| + \sqrt{\log m}\sup_{i\le m}\|U(e_i)\|\Big) \tag{7.4}$$

is sharp. Consider an integer k, and $m = k2^k$. We think of $\{1,\ldots,m\}$ as 2^k consecutive blocks of length k. If i belongs to the ℓth such block, we define $U(e_i) = k^{-1/2}f_\ell$, where $(f_\ell)_{\ell\le 2^k}$ is the canonical basis of $\ell^2_{2^k}$. It should be clear that

$$\|U\| \le 1, \qquad \sqrt{\log m}\sup_{i\le m}\|U(e_i)\| \le K.$$

Thus (7.4) implies

$$E\|U_I\| \le \frac{K}{\sqrt{\log(1/\delta)}}.$$

We will show that, for $\delta \ge 2^{-k}$, we have

$$E\|U_I\| \ge \frac{1}{K\sqrt{\log(1/\delta)}}. \tag{7.5}$$

Consider $1 \le r \le k$.

For each block B of length k, there is (crudely) a probability greater than or equal to δ^r that at least r values of δ_i, $i \in B$, are 1. If $\delta^r \ge 2^{-k}$, with probability greater than or equal to $1/K$ this will occur for at least one block, and thus

$$P\bigg(\|U_I\| \ge \frac{\sqrt{r}}{\sqrt{k}}\bigg) \ge \frac{1}{K},$$

so that $E\|U_I\| \ge \sqrt{r}/K\sqrt{k}$. Taking $r = [k\log 2/\log(1/\delta)]$, we get $\delta^r \ge 2^{-k}$ and hence (7.5).

To explain Theorem 7.1, let us set for simplicity $f_i = U(e_i)$. Then

$$\begin{aligned}\|U_I\| &= \sup\{x^*(U(y));\ x^* \in W_1^*,\ y \in \ell_I,\ \|y\| \le 1\}\\ &= \sup\bigg\{x^*\bigg(\sum_{i\in I} a_i f_i\bigg);\ x^* \in W_1^*;\ \sum_{i\in I} a_i^2 \le 1\bigg\}\\ &= \sup\bigg\{\sum_{i\in I} a_i x^*(f_i);\ x^* \in W_1^*;\ \sum_{i\in I} a_i^2 \le 1\bigg\}\\ &= \sup\bigg\{\bigg(\sum_{i\in I} x^*(f_i)^2\bigg)^{1/2};\ x^* \in W_1^*\bigg\},\end{aligned}$$

where $W_1^* = \{x^* \in W^*;\ \|x^*\| \le 1\}$.

Equivalently,

$$\|U_I\|^2 = \sup\left\{\sum_{i\le m} \delta_i x^*(f_i)^2;\ x^* \in W_1^*\right\}. \tag{7.6}$$

Consider the set

$$T = \{(x^*(f_i)^2)_{i\le m};\ x^* \in W_1^*\}.$$

For a sequence $t = (t_i)_{i\le m}$, consider the random variable $X'_t = \sum_{i\le m} \delta_i t_i$.

Thus we have

$$E\|U_I\|^2 = E\sup_{t\in T} X'_t, \tag{7.7}$$

which shows that the nature of the problem is familiar to us.

Setting $X_t = \sum_{i\le m}(\delta_i - \delta)t_i$ and since [as shown by the computation leading to (7.6)] we have $\sum_{i\le m} t_i \le \|U\|^2$ for $t \in T$, we get

$$E\|U_I\|^2 \le \delta\|U\|^2 + E\sup_{t\in T} X_t. \tag{7.8}$$

The first task to bound the last term is to understand the tails of $X_s - X_t = X_{s-t}$. This is done through the following lemma, which is a weak form of Bennett's inequality. For a sequence $(t_i)_{i\le m}$, we set $\|t\|_\infty = \sup_{i\le m}|t_i|$, $\|t\|_2 = (\sum_{i\le m} t_i^2)^{1/2}$.

LEMMA 7.2 (See [23]). *For all $u > 0$, we have*

$$P\left(\left|\sum_{i\le m}(\delta_i - \delta)t_i\right| \ge u\right) \le 2\exp\left(-\frac{u}{4\|t\|_\infty}\log\frac{u\|t\|_\infty}{\delta\|t\|_2^2}\right). \tag{7.9}$$

This bound is useless for $u\|t\|_\infty \le \delta\|t\|_2^2$, but will never be used in that range. What (7.9) brings to light is that the tails of $X_s - X_t$ do not depend only on $\|s-t\|_2$, but also on $\|s-t\|_\infty$. To find a useful bound, we need a somewhat nontrivial adaptation of Theorem 2.4. This adaptation pertains to the material of Section 9, and will be better discussed there.

LEMMA 7.3. *We have, for each number $A \ge 2$,*

$$E\sup_{t\in T} X_t \le K\left[\frac{\gamma_1(X, \|\cdot\|_\infty)}{\log A} + \sqrt{A\delta}\,\gamma_2(X)\right].$$

Taking $A = 1/\sqrt{\delta}$, we get

$$E\sup_{t\in T} X_t \le K\left[\frac{1}{\log(1/\delta)}\gamma_1(X, \|\cdot\|_\infty) + \delta^{1/4}\gamma_2(X)\right].$$

Combining with (7.8), to prove Theorem 7.1, it suffices to prove that

$$\gamma_1(T, \|\cdot\|_\infty) \le K(C)\Big[\|U\|^2 + \sup_{i\le m}\|U(e_i)\|^2\log i\Big], \tag{7.10}$$

$$\gamma_2(T) \le K(C)\Big[\|U\|^2 + \sup_{i\le m}\|U(e_i)\|^2\log i\Big]. \tag{7.11}$$

It turns out that (7.11) is a consequence of (7.10) and of the fact that $\sum_{i\le m} t_i \le \|U\|^2$ for t in T. This follows, in particular, from [21]. (A more direct proof is possible but apparently not trivial.) Thus we will concentrate on the main point, that is, the proof of (7.10). The main idea of this proof is closely related to the idea of Theorem 6.3. It is to combine a geometric argument and the control of the numbers $N(W_1^*, \|\cdot\|_\infty, \varepsilon)$, where $\|\cdot\|_\infty$ is the norm on W^* given by

$$\|x^*\|_\infty = \sup_{i\le m} \|x^*(f_i)\|,$$

where we recall that $f_i = U(e_i)$.

Thus the first step is to control the numbers $N(W_1^*, \|\cdot\|_\infty, \varepsilon)$.

LEMMA 7.4. *We have*

$$\varepsilon^2 \log N(W_1^*, \|\cdot\|_\infty, \varepsilon) \le K(C) \sup_{i\le m} \|f_i\|^2 \log(i+1). \tag{7.12}$$

It turns out (by geometric arguments that use the 2-convexity of W^*—see [4]) that, to prove (7.12), it suffices to prove the following "dual" statement.

LEMMA 7.5. *Consider a 2-smooth Banach space W and vectors $(f_i)_{i\le m}$ in W. Consider the balanced convex hull T of the vectors $(f_i)_{i\le m}$. Then we have, for all $\varepsilon > 0$,*

$$\varepsilon^2 \log N(T, \|\cdot\|, \varepsilon) \le K(C) \sup_{i\le m} \|f_i\|^2 \log(i+1). \tag{7.13}$$

This lemma is an exception to the policy that we have used in the present paper, not to give any complicated proof. The reason for this exception is that Lemma 7.5 supports some conjectures that are of potential importance and that will be discussed in Section 8. For the reader who contents himself with (7.4) rather than with the full strength of Theorem 7.1, it suffices to prove the following, which is simpler, and serves as an introduction to Lemma 7.5.

LEMMA 7.6. *Under the hypothesis of Lemma 7.5, we have*

$$\varepsilon^2 \log N(T, \|\cdot\|, \varepsilon) \le K(C) \log(m+1) \sup_{i\le m} \|f_i\|^2.$$

PROOF. This is an argument of Maurey. By homogeneity, we can assume $\|f_i\| \le 1$. Consider $x \in T$. Then $x = \sum_{i\le m} \alpha_i f_i$, where $\sum_{i\le m} |\alpha_i| \le 1$. Consider a sequence $(Y_\ell)_{\ell\le k}$ of independent W-valued r.v.'s, with $P(Y_\ell = (\operatorname{sign}\alpha_i) f_i) = |\alpha_i|$ and $P(Y_\ell = 0) = 1 - \sum_{i\le m} |\alpha_i|$. Thus $EY_\ell = x$. Consider $Y'_\ell = Y_\ell - x$, and observe that $\|Y'_\ell\| \le \|Y_\ell\| + \|x\| \le 2$. Thus, by (7.2),

$$E\left\|\sum_{\ell\le k} Y'_\ell\right\|^2 \le kK(C).$$

In particular, there exists a realization of $\sum_{\ell\le k} Y'_\ell$ that is of norm less than or equal to $K(C)\sqrt{k}$. This means that, for this realization,

$$\left\| \frac{1}{k}\sum_{\ell\le k} Y_\ell - x \right\| \le \frac{K(C)}{\sqrt{k}} \le \varepsilon$$

if $k \ge K(C)/\varepsilon^2$. Now, since each Y_ℓ can take at most $m+1$ values, there are at most $(m+1)^k$ points of the type $k^{-1}\sum_{\ell\le k} Y_\ell$. Since each point x of T is within distance ε of such a point, this finishes the proof. □

Proof of Lemma 7.5.

Step 1. For $\ell \ge 0$, we set $I_\ell = \{i \le m,\ 2^{\ell-1} \le \log(i+1) < 2^\ell\}$. By homogeneity, we can assume $\|f_i\|^2 \log(i+1) \le 1$, and we then have to prove that $\varepsilon^2 \log N(T, \|\cdot\|, \varepsilon) \le K(C)$. Consider $\varepsilon \le 1/2$ and the largest integer ℓ_0 with $2^{-\ell_0} \ge \varepsilon$. Consider a sequence $(n_\ell)_{\ell\le\ell_0}$ of integers $0 \le n_\ell \le \ell$ with $\sum_{\ell\ge 0} 2^{-n_\ell} \le 4$. Depending only on that sequence, we will construct a subset Z of T, with card $Z \le \exp(K/\varepsilon^2)$ such that, whenever we consider $x = \sum_{i\le m}\alpha_i f_i$ with $\sum_{i\in I_\ell}|\alpha_i| \le 2^{-n_\ell}$, we can find y in Z with $\|x-y\| \le K(C)\varepsilon$. We explain why this completes the proof. Consider the set Z' that is the reunion of the sets Z for all possible choices of the sequence $(n_\ell)_{\ell\le\ell_0}$. Since $\ell_0 \le K\log(1/\varepsilon)$, very crudely, there are at most

$$(\ell_0+1)^{\ell_0} \le \exp(K/\varepsilon^2)$$

choices of this sequence; thus card $Z' \le \exp(K/\varepsilon^2)$. Now, given x in A, we can write $x = \sum_{i\le m}\alpha_i f_i$ with $\sum_{i\le m}|\alpha_i| \le 1$. Thus, if n_ℓ is the largest integer less than or equal to ℓ with $\sum_{i\in I_\ell}|\alpha_i| \le 2^{-n_\ell}$, we easily see that $\sum_{\ell\le\ell_0} 2^{-n_\ell} \le 4$. Thus we can find y in Z' with $\|x-y\| \le K(C)\varepsilon$.

This shows that

$$\log N(T, \|\cdot\|, K(C)\varepsilon) \le \exp(K/\varepsilon^2)$$

for $\varepsilon \le 1/2$. Since $N(T, \|\cdot\|, \varepsilon) = 1$ for $\varepsilon \ge 2$, this finishes the proof.

Step 2. We observe that, given $\ell \ge 0$, $n_\ell \ge 0$, $k \ge 1$ and $x_\ell = \sum_{i\in I_\ell}\alpha_i f_i$, $\sum_{i\in I_\ell}|\alpha_i| \le 2^{-n_\ell}$, there exists an r.v. $Y_{\ell,k}$ with the following properties:

$$E\|Y_{\ell,k} - x_\ell\|^2 \le \frac{K(C)}{k}\frac{2^{-2n_\ell}}{2^\ell}. \tag{7.14}$$

(7.15) The random variable $Y_{\ell,k}$ takes at most $(1+\operatorname{card} I_\ell)^k$ possible values; these possible values depend only on n_ℓ.

This was shown in the proof of Lemma 7.5; observe simply that x_ℓ belongs to the balanced convex hull of the set $\{2^{-n_\ell}f_i;\ i\in I_\ell\}$ and that $\|2^{-n_\ell}f_i\|^2 \le 2^{-2n_\ell}/2^\ell$.

Step 3. For $\ell \le \ell_0$, set

$$k(\ell) = 1 + \left[\frac{2^{-n_\ell}}{2^\ell\varepsilon^2}\right].$$

Consider the r.v.

$$Y = \sum_{\ell \le \ell_0} Y_{\ell, k(\ell)}$$

and set $y = \sum_{\ell \le \ell_0} x_\ell$. Combining (7.2) and (7.14), we see that

$$\begin{aligned} E\|Y - y\|^2 &\le K(C) \sum_{\ell \le \ell_0} \frac{2^{-2n_\ell}}{k(\ell) 2^\ell} \\ &\le K(C) \sum_{\ell \le \ell_0} 2^{-n_\ell} \varepsilon^2 \le K(C)\varepsilon^2, \end{aligned} \tag{7.16}$$

since $k(\ell) \ge 2^{-n_\ell}/2^\ell \varepsilon^2$. On the other hand, by (7.15), the number of possible values that Y takes is at most

$$\begin{aligned} \prod_{\ell \le \ell_0} (1 + \operatorname{card} I_\ell)^{k(\ell)} &\le \exp\biggl(K \sum_{\ell \le \ell_0} 2^\ell \biggl(1 + \frac{2^{-n_\ell}}{2^\ell \varepsilon^2}\biggr)\biggr) \\ &\le \exp \frac{K}{\varepsilon^2}. \end{aligned}$$

Moreover,

$$\|x - y\| \le \biggl\| \sum_{\ell > \ell_0} x_\ell \biggr\| \le \sum_{\ell > \ell_0} \|x_\ell\| \le \sum_{\ell > \ell_0} \frac{2^{-n_\ell}}{2^\ell} \le 2^{-\ell_0} \le K\varepsilon.$$

The proof is finished. □

We now turn to the proof of (7.10). Consider the function h from $\mathbb{R}$ to $\mathbb{R}$ given by $h(x) = (\operatorname{sign} x)x^2 = x|x|$. For x^* in W_1^*, we set

$$h(x^*) = (h(x^*(f_i)))_{i \le m}$$

and we consider the set

$$T_1 = \{h(x^*);\ x^* \in W_1^*\}.$$

For technical reasons, rather than (7.10) we will prove

$$\gamma_1(T_1, \|\cdot\|_\infty) \le K(C)\Bigl[\|U\|^2 + \sup_{i \le m} \|f_i\|^2 \log i\Bigr]. \tag{7.17}$$

Since T_1 is the image of T under the map

$$(x_i)_{i \le m} \to (|x_i|)_{i \le m},$$

which is a contraction, it is obvious by definition that $\gamma_1(T, \|\cdot\|_\infty) \le \gamma_1(T_1, \|\cdot\|_\infty)$.

The proof of (7.17) will rely on Proposition 4.4 with $\beta = 1$, $\theta(n) = \log n / K(C)$, $r = 8$.

For $t \in T_1$, $j \in \mathbb{Z}$, we set

$$C_j(t) = \{x^* \in W_1^*;\ \|h(x^*) - t\|_\infty \le 2r^{-j}\}.$$

We observe that

$$C_j(t) \text{ is convex.} \tag{7.18}$$

(This is why we replaced T by T_1.)

Also, given s in T_1, such that $\|t-s\|_\infty \le r^{-j}$, we have for all x^* that

$$\|h(x^*)-t\|_\infty \le \|h(x^*)-s\|_\infty + r^{-j},$$

so that

$$\|t-s\|_\infty \le r^{-j} \Rightarrow C_{j+2}(s) \subset C_j(t). \tag{7.19}$$

This is the reason for the factor 2 in the definition of C_j.

We now set, for $t \in T_1$,

$$\psi'_j(t) = \sum_{i\le m}(|t_i| - \min(|t_i|, 2r^{-j})),$$

$$\psi''_j(t) = \inf\{\|x^*\|;\ x^* \in C_j(t)\},$$

$$\psi_j(t) = \psi'_j(t) + S\psi''_j(t),$$

where $S = \sup_{i\ge 1}\|U(e_i)\|^2 \log(i+1)$.

Since for t in T we have $\sum_{i\le m}|t_i| \le \|U\|^2$, it follows that $\psi'_j(t) \le \|U\|^2$, so that $\psi_j(t) \le \|U\|^2 + S$. Thus it suffices to prove (4.4bis). Before we do this, we collect a few facts. The following is well known.

LEMMA 7.7. *Consider a Banach space Y of dimension k and its unit ball B. Then*

$$N(B,\varepsilon) \le \left(1+\frac{2}{\varepsilon}\right)^k.$$

PROOF. Consider points $x_1,\dots,x_n$ in B such that

$$\ell,\ell' \le n \Rightarrow \|x_\ell - x_{\ell'}\| > \varepsilon/2.$$

The balls of radius $\varepsilon/2$, centered at the points x_ℓ, are disjoint and contained in the ball of radius $1+\varepsilon/2$. Thus

$$n\,\mathrm{Vol}\left(\frac{\varepsilon}{2}B\right) \le \mathrm{Vol}\left(\left(1+\frac{\varepsilon}{2}\right)B\right), \tag{7.20}$$

that is,

$$n\left(\frac{\varepsilon}{2}\right)^k \le \left(1+\frac{\varepsilon}{2}\right)^k,$$

so that $n \le (1+2/\varepsilon)^k$. If the family $\{x_1,\dots,x_n\}$ is chosen maximal, the balls of radius $\varepsilon/2$ centered at these points cover B. This completes the proof. □

LEMMA 7.8. *If $s \geq 2r^{-j}$, we have*

$$|s-t| \leq r^{-j} \Rightarrow s - \min(s, 2r^{-j}) + \frac{r^{-j}}{2} \leq t - \min(t, 2r^{-j-2}). \tag{7.21}$$

PROOF. Since $t \geq r^{-j}$, this reduces to

$$s - t \leq \frac{3}{2} r^{-j} - 2r^{-j-2},$$

but the right-hand side is greater than r^{-j}. □

We consider $s \in T_1$, points $t_1, \dots, t_n$ in T_1 such that $\|s - t_\ell\| \leq r^{-j}$ for $\ell \leq n$ and

$$\ell, \ell' \leq n, \qquad \ell \neq \ell' \Rightarrow \|t_\ell - t_{\ell'}\|_\infty \geq r^{-j-1}$$

and we start the proof of (4.4bis). Since, for $i \leq m$, we have $|s_i - t_{\ell,i}| \leq r^{-j}$, we have $||s_i| - |t_{\ell,i}|| \leq r^{-j}$, so that Lemma 7.8 shows that

$$\psi'_{j+2}(t_\ell) \geq \psi'_j(s) + \frac{r^{-j}}{2} \operatorname{card}\{i \leq m;\ |s_i| \geq 2r^{-j}\}. \tag{7.22}$$

We now consider a parameter K_1, to be determined later.

CASE 1. We have

$$\operatorname{card}\{i \leq m;\ |s_i| \geq 2r^{-j}\} \geq \frac{\log n}{K_1}.$$

In that case, we have, by (7.22),

$$\psi'_{j+2}(t_\ell) \geq \psi'_j(s) + \frac{r^{-j}}{2K_1} \log n.$$

Moreover, by (7.19), we have $C_{j+2}(t_\ell) \subset C_j(s)$, so that $\psi''_{j+2}(t_\ell) \geq \psi''_j(s)$ by the definition of ψ''. Thus

$$\forall\, \ell \leq n, \qquad \psi_{j+2}(t_\ell) \geq \psi_j(s) + \frac{r^{-j}}{2K_1} \log n$$

and (4.4bis) holds as soon as $\theta(n) \leq \log n / 2K_1$.

CASE 2. Case 1 does not occur, so that

$$\operatorname{card} I \leq \frac{\log n}{K_1}, \tag{7.23}$$

where $I = \{i \leq m;\ |s_i| \geq 2r^{-j}\}$.

The purpose of the functional ψ' was to create this condition. The main argument starts now.

Step 1. Consider $\delta = \max_{\ell \le N} \psi''_{j+2}(t_\ell)$, and recall that $\delta \le 1$. By compactness, for each $\ell \le n$ we can find x^*_ℓ in $C_{j+2}(t_\ell)$ with $\|x^*_\ell\| \le \delta$. We set $s_\ell = h(x^*_\ell) \in T_1$. Since $r = 8$, we note that

$$\ell \neq \ell' \Rightarrow \|s_\ell - s_{\ell'}\|_\infty \ge r^{-j-1} - 4r^{-j-2} = \frac{r^{-j-1}}{2}. \tag{7.24}$$

Step 2. Provided K_1 has been chosen appropriately, we show that we can find a subset L of $\{1, \dots, n\}$ such that $\operatorname{card} L \ge \sqrt{n}$, and that the following property holds:

$$\forall\, \ell, \ell' \in L,\ \ell \neq \ell' \Rightarrow \exists\, i \le m,\ i \notin I, \qquad |s_{\ell,i} - s_{\ell',i}| \ge \frac{r^{-j-1}}{2}. \tag{7.25}$$

To see this, consider the set

$$B = \{(t_i)_{i \le m};\ \forall\, i \in I,\ |t_i| \le 1\},$$

which depends on $\operatorname{card} I$ coordinates. By Lemma 7.8, the set $s + 2r^{-j}B$ can be covered by sets $W_p = u_p + \frac{1}{5}r^{-j-1}B,\ p \le n_1$, for

$$n_1 \le (1 + 20r)^{\operatorname{card} I}.$$

Thus, by (7.23), we see that $n_1 \le \sqrt{n}$ provided K_1 is large enough. By the "pigeon hole" principle, we can find a given set W_p such that $\operatorname{card} L \ge \sqrt{n}$, where $L = \{\ell \le n;\ s_\ell \in W_p\}$. Given $\ell, \ell' \in L$, we have $|s_{\ell,i} - s_{\ell',i}| \le \frac{2}{5}r^{-j-1}$ whenever $i \in I$. Combining with (7.24), this proves (7.25).

Step 3. We show that

$$\ell, \ell' \in L, \qquad \ell \neq \ell' \Rightarrow \|x^*_\ell - x^*_{\ell'}\|_\infty \ge \frac{r^{-j/2}}{64}. \tag{7.26}$$

Given ℓ, ℓ' as above, we know that, for some $i \notin I$,

$$|s_{\ell,i} - s_{\ell',i}| = |h(x^*_\ell(f_i)) - h(x^*_{\ell'}(f_i))| \ge \frac{r^{-j-1}}{2}. \tag{7.27}$$

However, since $i \notin I$, we have $|s_i| \le 2r^{-j}$, so that

$$|s_{\ell,i}|, |s_{\ell',i}| \le 2r^{-j} + r^{-j-1} \le 4r^{-j}$$

and $|x^*_{\ell'}(f_i)|, |x^*_\ell(f_i)| \le 2r^{-j/2}$.

It remains now to note that

$$\begin{aligned} |h(a) - h(b)| &\le |b - a| \sup\{h'(x);\ |x| \le \max(|a|, |b|)\} \\ &\le 2\max(|a|, |b|)|b - a|, \end{aligned}$$

so that (7.26) implies

$$|x^*_\ell(f_i) - x^*_{\ell'}(f_i)| \ge \frac{r^{-j/2}}{8r} = \frac{r^{-j/2}}{64}.$$

Step 4. We fix ℓ_0 in L, and we consider

$$R = \max_{\ell \in L} \|x^*_\ell - x^*_{\ell_0}\|.$$

The points $y_\ell^* = x_\ell - x_{\ell_0}$ for ℓ in L belong to the ball centered at 0 of radius R. Their mutual distances are at least $r^{-j/2}/64$, as shown in step 3. Thus no ball of radius $r^{-j/2}/129$ can cover any two of them. It then follows from Lemma 7.5 that

$$\left(\frac{r^{-j/2}}{R}\right)^2 \log L \leq K(C)S,$$

so that

$$R^2 \geq r^{-j} \frac{\log n}{SK(C)}.$$

Thus there exists ℓ_1 in L such that

$$\|x_{\ell_1}^* - x_{\ell_0}^*\|^2 = R^2 \geq \frac{r^{-j}}{SK(C)} \log n.$$

We now appeal to (7.3) with $x^* = x_{\ell_1}^*/\delta$, $y^* = x_{\ell_0}^*/\delta$, to get

$$\left\|\frac{x_{\ell_0}^* + x_{\ell_1}^*}{2}\right\| \leq \delta - \frac{r^{-j}\log m}{\delta K(C)S} \leq \delta - \frac{r^{-j}\log m}{K(C)S}$$

since $\delta \leq 1$. Now, we have $x_\ell^* \in C_{j+2}(t_\ell) \subset C_j(s)$, so that, since $C_j(s)$ is convex, $(x_{\ell_0}^* + x_{\ell_1}^*)/2 \in C_j(s)$, and the definition of ψ_j'' shows that

$$\psi_j''(s) \leq \delta - \frac{1}{K(C)} \frac{r^{-j}\log m}{S},$$

that is,

$$\delta = \max_{\ell \leq n} \psi_{j+2}''(t_\ell) \geq \psi_j''(s) + \frac{r^{-j}\log m}{K(C)S}.$$

We now observe that

$$|s - t| \leq r^{-j} \Rightarrow s - \min(s, 2r^{-j}) \leq t - \min(t, 2r^{-j-2}).$$

Indeed, if $s \leq 2r^{-j}$ this is obvious; but if $s \geq 2r^{-j}$ more is true by Lemma 7.9. Thus

$$\psi_{j+2}'(t_\ell) \geq \psi_j'(s),$$

so that

$$\max_{\ell \leq n} \psi_{j+2}(t_\ell) \geq \psi_j(s) + \frac{r^{-j}\log m}{K(C)}.$$

This completes the proof. □

8. Convex hull and open problems. Consider a sequence (t_k) of points of ℓ^2, with

$$\|t_k\| \le \frac{1}{\sqrt{\log(k+1)}}.$$

We claim that

$$E \sup_{k \ge 1} X_{t_k} \le K, \tag{8.1}$$

where, for t in ℓ^2, X_t is given by (3.1). To prove (8.1), it suffices to observe that, for $u \ge 0$,

$$P(X_t \ge u) \le \exp\left(-\frac{u^2}{2\|t\|^2}\right)$$

and hence

$$P(X_{t_k} \ge u) \le \exp\left(-\frac{u^2}{2}\log(k+1)\right),$$

so that

$$\begin{aligned} P\left(\sup_{k\ge 1} X_{t_k} \ge u\right) &\le \sum_{k \ge 1} \exp\left(-\frac{u^2}{2}\log(k+1)\right) \\ &\le \exp\left(-\frac{u^2}{4}\log 2\right) \sum_{k\ge 1} \exp\left(-\frac{u^2}{4}\log(k+1)\right). \end{aligned} \tag{8.2}$$

For $u \ge 3$, the latter series has a sum less than or equal to K, so that (8.2) and integration imply (8.1).

If we set

$$T = \operatorname{conv}(\{t_k;\ k \ge 1\}), \tag{8.3}$$

we have

$$\sup_{t \in T} X_t = \sup_{k \ge 1} X_{t_k},$$

so that

$$E \sup_{t \in T} X_t \le K. \tag{8.4}$$

The importance of that result is as follows. There exists a universal constant K such that, given $U \subset \ell^2$, containing 0, and such that

$$E \sup_{t \in U} X_t \le \frac{1}{K},$$

we can find a set T of the type (8.3) such that $U \subset T$. This is a simple consequence of Theorem 5.1 (see [14]).

If we combine (8.4) and Theorem 5.1 we see that $\gamma_2(T) \le K$ for all sets T of the type (8.3). It seems to me that one would gain deeper understanding by finding a "geometric" proof of this statement. To find such a proof, one should (for example) study the following question.

PROBLEM 8.1. Consider a 2-smooth Banach space W. Consider a sequence $(f_k)_{k\geq 1}$ in W, with $\|f_k\| \leq 1/\sqrt{\log(k+1)}$, and consider the convex hull T of this sequence. Is it true that $\gamma_2(T, \|\cdot\|) \leq K(C)$, where $K(C)$ depends only on the constant C in (7.1)?

The hypothesis that the Banach space is 2-smooth is a natural geometric weakening of the hypothesis that it is a Hilbert space. What is the "correct" hypothesis is certainly not known. Possibly type 2 suffices.

We observe that we have proved in Section 7 that $\varepsilon^2 \log N(T, \|\cdot\|, \varepsilon) \leq K(C)$. It is easily seen that this fact is weaker than the inequality $\gamma_2(T, \|\cdot\|) \leq K(C)$. This provides some (very weak) support for a positive answer to Problem 8.1.

We now will give a "geometric" proof that, when $f_k = e_k/\sqrt{\log(k+1)}$, where (e_k) is the canonical basis of ℓ^2, the convex hull T of the set of vectors f_k satisfies $\gamma_2(T) \leq K$. This proof unfortunately relies heavily on the special position of the vectors e_i. It is interesting to point out that, although there is no convexity involved in the present case, the proof has a lot in common with the proof of (7.17). This reflects the fact that, rather sadly, we apparently have essentially only one useful technique.

The proof combines Propositions 4.4 and 4.5. We will take $r = 12$, $\beta = 1$, $i = 0$, $\theta(n) = K^{-1}\sqrt{\log n}$ for a suitable value of K. We define the functionals

$$\psi_j(t) = \inf\left\{\sum_{k\geq 1} \alpha_k a_k;\ a_k \geq 0,\ \left\|t - \sum_{k\geq 1} a_k e_k\right\| \leq 3r^{-j}\right\},$$

where, for simplicity, we set $\alpha_k = \sqrt{\log(k+1)}$.

Since $\sum a_k e_k = \sum \alpha_k a_k f_k$, we see that $\psi_j(t) \leq 1$, and we now turn to the proof of (4.4bis). Consider s in T and points $t_1, \dots, t_n$ in $B(s, r^{-j})$ such that

$$\forall\ p, q \leq n, \qquad p \neq q \Rightarrow d(t_p, t_q) \geq r^{-j-1}.$$

We consider the number

$$\psi = \inf\left\{\sum_{k\geq 1} \alpha_k a_k;\ a_k \geq 0;\ \left\|s - \sum_{k\geq 1} a_k e_k\right\| \leq 2r^{-j}\right\}.$$

Since

$$p \leq n \Rightarrow B(t_p, 3r^{-j-2}) \subset B(s, 2r^{-j}),$$

we have $\psi_{j+2}(t_p) \geq \psi$.

Consider a parameter A, which will be determined later, and

$$\varepsilon = \frac{r^{-j}}{A}\sqrt{\log n}.$$

Consider $(a_k)_{k\geq 1}$ such that

$$\sum_{k\geq 1} \alpha_k a_k < \psi + \varepsilon, \qquad \left\| s - \sum_{k\geq 1} a_k e_k \right\| \leq 2r^{-j}.$$

Consider the largest number $u \leq 2$ such that

$$\sum_{k\geq 1} \min(a_k, u\alpha_k)^2 \leq r^{-2j}.$$

Setting $b_k = \min(a_k, u\alpha_k)$, we see that $\| \sum_{k\geq 1} b_k e_k \| \leq r^{-j}$. Thus, if $c_k = a_k - b_k$, we have $\| s - \sum_{k\geq 1} c_k e_k \| \leq 3r^{-j}$, so that

$$\psi_j(s) \leq \sum_{k\geq 1} \alpha_k c_k = \sum_{k\geq 1} \alpha_k a_k - \sum_{k\geq 1} \alpha_k b_k \leq \psi + \varepsilon - \sum_{k\geq 1} \alpha_k b_k. \tag{8.5}$$

We first explore the main case, that is,

$$\sum_{k\geq 1} b_k^2 = r^{-2j}. \tag{8.6}$$

Since $b_k \leq u\alpha_k$, we have

$$u \sum_{k\geq 1} \alpha_k b_k \geq r^{-2j}. \tag{8.7}$$

Case 1. We have

$$\frac{r^{-2j}}{u} \geq \frac{2r^{-j}}{A} \sqrt{\log n}.$$

In this case, we have, by (8.5) and (8.7),

$$\psi_j(s) \leq \psi + \varepsilon - \frac{2r^{-j}}{A} \sqrt{\log n} \leq \psi - \frac{r^{-j}}{A} \sqrt{\log n}.$$

Thus, for all $p \leq n$, we have

$$\psi_{j+2}(t_p) \geq \psi \geq \psi_j(s) + \frac{r^{-j}}{A} \sqrt{\log n}$$

and thus (4.4bis) is proved provided $\theta(n) \leq \sqrt{\log n}/A$.

Case 2. We have

$$\frac{r^{-2j}}{u} \leq \frac{2r^{-j}}{A} \sqrt{\log n},$$

so that $r^{-j}/u \leq 2\sqrt{\log n}/A$.

Consider the set $L = \{k \geq 1;\ b_k = u\alpha_k\}$. Thus, by (8.6),

$$u^2 \sum_{k\in L} \alpha_k^2 \leq r^{-2j}. \tag{8.8}$$

From (8.8) we see that

$$\sum_{k\in L}\alpha_k^2 \le \frac{r^{-2j}}{u^2} \le \frac{4}{A^2}\log n. \tag{8.9}$$

For each $\ell \le n$, consider a point $s_\ell \in B(t_\ell, 3r^{-j-2})$ and set $s_\ell = \sum_{k\ge 1} a_{k,\ell}e_k$. We observe that

$$\sum_{k\in L}(a_{k,\ell}-a_k)^2 \le \left\| s_\ell - \sum_{k\ge 1} a_k e_k\right\|^2 \le Kr^{-2j}.$$

Thus, by Cauchy–Schwarz and (8.9),

$$\begin{aligned}\sum_{k\in L}\alpha_k|a_{k,\ell}-a_k| &\le \left(\sum_{k\in L}\alpha_k^2\right)^{1/2}\left(\sum_{k\in L}(a_{k,\ell}-a_k)^2\right)^{1/2}\\ &\le \frac{Kr^{-j}}{A}\sqrt{\log n}.\end{aligned}$$

We observe that

$$\sum_{k\ge 1}\alpha_k c_k \le \sum_{k\in L}\alpha_k a_k$$

since $c_k \le a_k$, $c_k = 0$ if $k \in L$. Thus, using (8.5), we get

$$\sum_{k\in L}\alpha_k a_{k,\ell} \ge \sum_{k\in L}\alpha_k a_k - \frac{Kr^{-j}}{A}\sqrt{\log n} \ge \psi_j(s) - \frac{Kr^{-j}}{A}\sqrt{\log n}. \tag{8.10}$$

We set

$$B = \max_{\ell\le n}\sum_{k\notin L}\alpha_k a_{k,\ell}. \tag{8.11}$$

Thus, combining with (8.10), we see that

$$\max_{\ell\le n}\sum_{k\ge 1}\alpha_k a_{k,\ell} \ge \psi_j(s) + B - \frac{Kr^{-j}}{A}\sqrt{\log n}. \tag{8.12}$$

Thus, if we can prove that $B \ge r^{-j}\sqrt{\log n}/K'$, the proof will be finished by taking $A = 2KK'$, since the point s_ℓ is arbitrary in $B(t_\ell, 3r^{-j-2})$. To prove this, we recall by Lemma 7.5 that

$$\log N(T, \|\cdot\|, \varepsilon) \le \frac{K}{\varepsilon^2},$$

so that, by homogeneity,

$$\log N(BT, \|\cdot\|, \varepsilon) \le \frac{KB^2}{\varepsilon^2}. \tag{8.13}$$

Consider, for $\ell \le n$, the point $v_\ell = \sum_{k\notin L} a_{k,\ell}e_k$, so that $v_\ell \in BT$.
We are going to show that (provided A is large enough) we have

$$\frac{KB^2}{(r^{-j-1}/16)^2} \ge \log\sqrt{n}, \tag{8.14}$$

where K is the constant of (8.13). As explained above, this concludes the proof. Assume, for contradiction, that (8.14) fails. Then, by (8.13), we can cover BT by at most $\sqrt{n}$ balls of radius less than or equal to $r^{-j-1}/16$. The "pigeon hole" principle then implies that we can find a subset M of $\{1, \dots, n\}$, with $\operatorname{card} M \geq \sqrt{n}$ and

$$(8.15) \qquad \forall\, \ell, \ell' \in M, \qquad \|v_\ell - v_{\ell'}\| \leq r^{-j-1}/8.$$

Now, we observe that

$$(8.16) \qquad \ell, \ell' \leq n, \qquad \ell \neq \ell' \Rightarrow \|s_\ell - s_{\ell'}\| \geq r^{-j-1} - 6r^{-j-2} \geq r^{-j-1}/2.$$

Thus, if we set $w_\ell = s_\ell - v_\ell$, combining (8.15) and (8.16), we see that

$$\ell, \ell' \in M, \qquad \ell \neq \ell' \Rightarrow \|w_\ell - w_{\ell'}\| \geq r^{-j-1}/4.$$

The points w_ℓ belong to the space H generated by the vectors $\{e_k;\ k \in L\}$. They belong to the ball $B(w, r^{-j})$, where $w = \sum_{k \in L} a_k e_k$. Thus, by Lemma 7.7, we have

$$(8.17) \qquad \operatorname{card} M \leq \left(1 + \frac{2}{r/4}\right)^{\operatorname{card} L} \leq K^{\operatorname{card} L}.$$

Now, by (8.9), we have $\operatorname{card} L \leq K \log n / A^2$, so that, if A is large enough, (8.17) contradicts the fact that $\operatorname{card} M \geq \sqrt{n}$.

Thus the proof is finished in the case that (8.6) holds. If (8.6) fails, the definition of u shows that this is because $\sum_{k\geq 1} a_k^2 \leq r^{-2j}$; thus $\psi_j(s) = 0$. It then suffices to show that, given an arbitrary point $s_\ell = \sum_{k \geq 1} a_{k,\ell} e_k$, we have $B \geq r^{-j}\sqrt{\log n}/K$, where B is given by (8.11). This follows from Lemma 7.7 as above. □

To conclude this section, we would like to mention an open question related to matching problems and to the transportation cost from the empirical measure to the uniform measure on $[0, 1]^2$ (for notions of transportation cost more elaborate than the one studied in Section 7). The study of transportation costs involves classes of functions; these classes are not so easy to describe, so we will rather consider simpler related classes for which the nature of the difficulty seems the same. Denoting by ψ_a the function $\psi_a(x) = |x| \log(1 + |x|)^a$ for $a \geq 0$, consider the class

$$C(a, b) = \left\{ f\colon\ [0,1]^2 \to \mathbb{R},\ |f| \leq 1, \iint \psi_a\left(\frac{\partial f}{\partial x}\right) d\lambda \leq 1, \iint \psi_b\left(\frac{\partial f}{\partial y}\right) d\lambda \leq 1 \right\}.$$

PROBLEM 8.2. Is it true that, for $a + b \geq 1/2$, we have $\gamma_{2,2}(C(a, b)) < \infty$?

It is proved in [18], Theorem 6.5, that the condition $a + b > 4$ suffices; but the method there cannot yield optimal results. The case $a = 0,\ b = 1/2$ would yield new results on the transportation cost (improving upon Theorem 1.8 of [18]).

There is a kind of similarity between Problem 8.2 and the study of the convex hull of a sequence, in the sense that the weak convexity properties of the functions ψ_a apparently make convexity useless (in contrast with the case of Lipshitz functions of Section 7). The main challenge of Problem 8.2 is, however, that one must apparently take strongly into account not only the integrability properties of $\partial f/\partial x$ and $\partial f/\partial y$, but the fact (which is not used in [18]) that these are the derivatives of the same function, a strong constraint that is hard to use.

9. Families of distances. It is an unfortunate fact that, as exemplified by Lemma 7.2, not all processes of interest satisfy tail conditions as simple as (1.1). Yet the program of extending Theorem 5.1 has been successful in two important situations (see [18] and [20]). Rather than reproducing here the results of these papers, we have chosen to attempt to explain some of the key ideas needed for this purpose. We will consider only the situation of [20]. Throughout this section we consider a number $1 \le \alpha < \infty$ and we denote by $(h_n)_{n\ge1}$ a sequence of independent symmetric random variables such that $P(|h_n| \ge u) = a_\alpha \exp(-u^\alpha)$, where a_α is a normalizing constant. Thus, when $\alpha = 2$, h_n is Gaussian. The second most interesting case is $\alpha = 1$. For $t = (t_n)$ in ℓ^2, we set

$$X_t = \sum_{n\ge1} t_n h_n. \tag{9.1}$$

The following lemma is a standard exercise ([20], Corollary 2.9). In this lemma, for $t = (t_n)_{n\ge1}$, we set $\|t\|_\beta = \left(\sum |t_n|^\beta\right)^{1/\beta}$, where β is the conjugate exponent of α. When $\alpha = 1$, $\beta = \infty$, and then $\|t\|_\infty = \sup_{n\ge1} |t_n|$.

LEMMA 9.1. *If $t \in \ell^2$, we have the following*:

(a) *If $\alpha \ge 2$, for all $u > 0$, we have*

$$P(|X_t| \ge u) \le 2\exp\left(-\frac{1}{K(\alpha)}\max\left(\frac{u^2}{\|t\|_2^2}, \frac{u^\alpha}{\|t\|_\beta^\alpha}\right)\right).$$

(b) *If $\alpha \le 2$, for all $u > 0$, we have*

$$P(|X_t| \ge u) \le 2\exp\left(-\frac{1}{K(\alpha)}\min\left(\frac{u^2}{\|t\|_2^2}, \frac{u^\alpha}{\|t\|_\beta^\alpha}\right)\right). \tag{9.2}$$

Here $K(\alpha)$ is a constant depending on α only.

Consider now a subset T of ℓ^2, and the problem of finding useful bounds for $\sup_{t\in T} X_t$. We discuss only the case $\alpha \le 2$. Certainly we wish to try first to mimic the arguments of Section 2. It is not clear now how to control the size of the sets $A_j(t)$ considered there, so let us simply denote by $d_\beta(A)$ the diameter of $A \subset \ell^2$ for the distance induced by the norm $\|\cdot\|_\beta$.

In the chaining argument, in order to get a usable bound for

$$\sum_{j>i,\ s\in\Pi_j} P(|X_s - X_{\pi_{j-1}(s)}| \geq u a_j(s)) \tag{9.3}$$

in view of (9.2) we have to take

$$\begin{aligned} a_j(s) = K(\alpha)\Bigg[& d_2(A_{j-1}(s))\sqrt{\log\frac{1}{w_j(s)}} \\ & + d_\beta(A_{j-1}(s))\left(\log\frac{1}{w_j(s)}\right)^{1/\alpha}\Bigg], \end{aligned} \tag{9.4}$$

where $w_j(s)$ is the term of (9.3) for $u = 1$, and to require $\sum w_j(s) \leq 1$. Thus, working now with measures μ on T, what we need is to control both

$$S_2 = \sup_{t\in T}\sum_{j>i} d_2(A_{j-1}(s))\sqrt{\log\frac{1}{\mu(A_j(s))}}$$

and

$$S_\alpha = \sup_{t\in T}\sum_{j>i} d_\beta(A_{j-1}(s))\left(\log\frac{1}{\mu(A_j(s))}\right)^{1/\alpha}.$$

An immediate problem is that, even if we have a sequence of partitions $(\mathscr{A}_j)$ for which we control S_2, and one $(\mathscr{A}'_j)$ for which we control S_α, it is unclear how to construct one that controls both terms simultaneously. The partition generated by $\mathscr{A}_j$ and $\mathscr{A}'_j$ does not work because of the different exponents of $\log 1/\mu(A_j(s))$. One way around the problem is to observe that, for conjugate exponents p, q, we have $ab \leq a^p + b^q$, so that

$$\begin{aligned} & d_q(A_{j-1}(s))\left(\log\frac{1}{\mu(A_j(s))}\right)^{1/p} \\ & \quad \leq r^{jq/p}\, d_q(A_{j-1}(s))^q + r^{-j}\log\frac{1}{\mu(A_j(s))}. \end{aligned} \tag{9.5}$$

Thus we can try to control the quantity

$$\sup_{t\in T}\sum_{j>i}(r^{j/p}\, d_q(A_{j-1}(t)))^q + r^{-j}\log\frac{1}{\mu(A_j(t))},$$

when either $p = q = 2$ or $p = \alpha,\ q = \beta,\ \alpha > 1$. When $p = \alpha = 1,\ q = \infty$, we rather require that $d_\infty(A_{j-1}(s)) \leq 2r^{-j+1}$. [It could be more appropriate in the present setting to require $r^j d_\infty(A_{j-1}(s)) \leq 1$; but the choice above is consistent with Section 2.]

The first positive result is that such a sequence of partitions can be constructed using a majorizing measure condition on T.

PROPOSITION 9.2. *Consider a metric space (T, d), and $\alpha > 1$. Consider an integer $i \in \mathbb{Z}$. Then we can find an increasing sequence of partitions $(\mathscr{A}_j)_{j\geq i}$ and a probability measure μ on T such that*

$$\sup_{t\in T} \sum_{j\geq i} r^{j\beta/\alpha} d(A_j(t))^\beta + r^{-j} \log \frac{1}{\mu(A_j(t))} \leq K(\alpha)\gamma_\alpha(T, d), \tag{9.6}$$

where $d(A)$ denotes the diameter of A.

The proof of Proposition 9.2 requires a simple, yet nontrivial construction (in the spirit of Proposition 4.3). See [18].

We pursue the discussion of chaining under condition (9.1). Consider a subset T of ℓ^2 and a number i to be determined later on.

First, we use Proposition 9.2 for the distance induced by the norm $\|\cdot\|_2$; we obtain a sequence of partitions $(\mathscr{B}_j)_{j\geq i}$ and a probability measure μ_1 on T such that

$$\sup_{t\in T} \sum_{j\geq i} r^j d_2(B_j(t))^2 + r^{-j} \log \frac{1}{\mu_1(B_j(t))} \leq K\gamma_2(T). \tag{9.7}$$

Using again Proposition 9.2 for the distance induced by the norm $\|\cdot\|_\beta$, we obtain a sequence of partitions $(\mathscr{C}_j)_{j\geq i}$ and a probability measure μ_2 on T such that

$$\sup_{t\in T} \sum_{j\geq i} r^{j\beta/\alpha} d_\beta(C_j(t))^\beta + r^{-j} \log \frac{1}{\mu_2(C_j(t))} \leq K\gamma_\alpha(T, \|\cdot\|_\beta). \tag{9.8}$$

Consider the sequence $(\mathscr{A}_j)_{j\geq i}$ of partitions of T such that $A_j(t) = B_j(t) \cap C_j(t)$. Consider a probability measure μ on T such that

$$\forall\ j \geq i,\ \forall\ A \in \mathscr{A}_j,\ \forall\ B \in \mathscr{B}_j, \qquad \mu(A \cap B) \geq 2^{i-j-1}\mu_1(A)\mu_2(B).$$

A simple computation using (9.6) and (9.7) shows that

$$\begin{aligned}\sup_{t\in T} \sum_{j\geq i} r^j d_2^2(A_j(t)) + r^{j\beta/\alpha} d_\beta(A_j(t))^\beta + r^{-j} \log \frac{1}{\mu(A_j(t))}&\\ \leq K(\gamma_\alpha(T, \|\cdot\|_\beta) + \gamma_2(T) + r^{-i}).&\end{aligned} \tag{9.9}$$

To use the chaining argument, we have to use a sequence of partitions for which the first partition is T; thus we set $\mathscr{A}_{i-1} = T$. Using chaining and (9.2) and (9.9), we then see by an easy adaptation of the arguments of Section 2 that

$$E \sup_{t\in T} X_t \leq K(\gamma_\alpha(T, \|\cdot\|_\beta) + \gamma_2(T) + r^{-i} + r^i d_2^2(T) + r^{i\beta/\alpha} d_\beta(T)^\beta). \tag{9.10}$$

Since $d_2(T) \leq K\gamma_2(T)$, $d_\beta(T) \leq \gamma_\alpha(T, \|\cdot\|_\beta)$, we then see that (since $\beta/\alpha = \beta - 1$) if we take i such that r^{-i} is of order $\gamma_2(T) + \gamma_\alpha(T, \|\cdot\|_\beta)$, then we get

$$E \sup_{t\in T} X_t \leq K(\gamma_2(T) + \gamma_\alpha(T, \|\cdot\|_\beta)). \tag{9.11}$$

The proof of Lemma 7.2 is similar to the proof of (9.11) (but somewhat harder).

The proof of (9.11) does not really require Proposition 9.2. It is possible to proceed directly (with essentially the same proof) as is done in [10], Chapter 11. The idea latent in the right-hand side of (9.6), however, becomes very precious when trying to prove a converse for (9.11), as we will soon explain.

Before that, it is useful to simplify the notation and to set

$$\varphi_j(s,t) = r^{2j}\|s-t\|_2^2 + r^{j\beta}\|s-t\|_\beta^\beta, \tag{9.12}$$

$$D_j(A) = \sup\{\varphi_j(s,t);\ s,t \in A\}, \tag{9.13}$$

so that controlling the left-hand side of (9.9) requires the control of

$$\sup_{t\in T}\sum_{j\ge i} r^{-j}\left(D_j(A_j(t)) + \log\frac{1}{\mu(A_j(t))}\right). \tag{9.14}$$

A converse to (9.11) is a generalization of Theorem 5.1. To see how to proceed, let us first investigate how we could expect to extend the crucial inequality (5.2) which is a key step of the proof of Theorem 5.1. In order to find a lower bound for $E\sup_{\ell\le n} X_{t_\ell}$, where X_t is given by (9.1), under the condition that the points $(t_\ell)_{\ell\le n}$ are well separated, it seems reasonable to require that for two numbers a_2, a_β we have

$$\ell \ne \ell' \Rightarrow \text{either } \|t_\ell - t_{\ell'}\|_2 \ge a_2 \text{ or } \|t_\ell - t_{\ell'}\|_\beta \ge a_\beta.$$

We can then hope that

$$E\sup_{\ell\le n} X_{t_\ell} \ge \frac{1}{K(\alpha)}\min\left(a_2\sqrt{\log n}, a_\beta(\log n)^{1/\alpha}\right) \tag{9.15}$$

Thus, in order to obtain $E\sup_{\ell\le n} X_{t_\ell} \ge A/K(\alpha)$, it seems natural to require $a_2 \ge A\sqrt{\log n}$, $a_\beta \ge A/(\log n)^{1/\alpha}$, so that

$$\begin{aligned}\ell \ne \ell' \Rightarrow \varphi_j(t_\ell, t_{\ell'}) &\ge r^{2j}\frac{A^2}{\log n} + \left(\frac{r^jA}{(\log n)^{1/\alpha}}\right)^\beta \\ &= \log n\left(\left(\frac{r^jA}{\log n}\right)^2 + \left(\frac{r^jA}{\log n}\right)^\beta\right).\end{aligned}$$

The appearance of the quantity $r^jA/\log n$ at two different powers apparently means that it would be wise to consider only the case where this quantity is of order 1. Thus we see that we are led to conjecture that

$$\ell \ne \ell' \Rightarrow \varphi_j(t_\ell, t_{\ell'}) \ge \log n \Rightarrow E\sup_{\ell\le n} X_{t_\ell} \ge \frac{1}{K(\alpha)} r^{-j}\log n. \tag{9.16}$$

This is indeed true (see [20]). The most remarkable feature of this statement is that, in contrast with Sudakov minoration (5.2), there is a precise relationship between the number of points to consider and how well they are separated.

This success in extending (5.2) provides motivation to prove a suitable extension of Theorem 4.2. The setting is as follows. On a space T, we assume that, for $j \in \mathbb{Z}$, we are given a function $\varphi_j(s,t)$ on $T \times T$. The function φ_j also depends on a number r. This number is not indicated in the notation, because it will remain fixed (after having been appropriately chosen). We assume that, for some number κ independent of r, we have

$$\varphi_j(s,t) \leq \kappa(\varphi_j(s,v) + \varphi_j(v,t)) \tag{9.17}$$

for all s, t, v in T. This is a substitute for the triangle inequality which takes into account the fact that φ_j resembles a power of a distance rather than a distance. We assume that, for a certain number $\delta > 0$, we have, for all r and all s, t in T,

$$\varphi_{j+1}(s,t) \geq r^{1+\delta}\varphi_j(s,t). \tag{9.18}$$

This condition is obvious in the case (9.12); unfortunately, it does not hold in many cases of interest, and in those cases many open problems remain, the most important of which is the Bernoulli conjecture of [21]. We assume, for simplicity, that $r = 2^\tau$, $\tau \in \mathbb{N}$. We assume that, for each subset S of T, there is associated a number $F(S)$. [Our typical choice for $F(S)$ will be $E\sup_{t\in S} X_t$.] We assume $F(S) \geq 0$, and that, if $S \subset S'$, $F(S) \leq F(S')$. We assume that the following holds, for certain numbers $\eta,\ \xi > 0$ and all $j \geq i,\ p \geq \tau - 1$:

(9.19) Assume that for some $p \geq \tau - 1$ there exists points $t_1, \dots, t_n$ in T, with $n = 2^{2^p}$, and assume that

$$\ell \neq \ell' \Rightarrow \varphi_j(t_\ell, t_{\ell'}) \geq 2^p, \tag{9.19a}$$

$$\forall\, \ell, \ell' \leq n, \qquad \varphi_{j-1}(t_\ell, t_{\ell'}) \leq \kappa 2^{p-\tau+2}. \tag{9.19b}$$

Consider then, for each $\ell \leq n$, a subset A_ℓ of T such that

$$t \in A_\ell \Rightarrow \varphi_j(t, t_\ell) \leq \eta 2^p. \tag{9.19c}$$

Then

$$F\left(\bigcup_{\ell\leq n} A_\ell\right) \geq \xi r^{-j}2^p + \min_{\ell\leq n} F(A_\ell). \tag{9.19d}$$

We hope that (9.15) motivates (9.19a) and (9.19d). To understand these conditions, it might also help to consider the case where T is a metric space and where $\varphi_j(s,t) = r^{2j}d(s,t)$. In that case $\kappa = 2$ and (9.19a)–(9.19c) become, respectively (since $r = 2^\tau$),

$$\ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq r^{-j}2^{p/2},$$

$$d(t_\ell, t_{\ell'}) \leq r^{-j}2^{p/2}2^{5/2},$$

$$A_\ell \subset B(t_\ell, r^{-j}\sqrt{\eta}2^{p/2}).$$

When $T \subset \ell^2$, $F(S) = E \sup_{t \in S} X_t$, it follows from Lemma 5.4 that (9.19d) holds when η is small enough. We observe that in this case condition (9.19b) is actually not needed. In [21] the following extension of Theorem 4.2 is proved. It is a key ingredient of the extensions of Theorem 5.1 considered in [18] and [20] (although the original arguments of [18] are somewhat more complicated).

THEOREM 9.3. *If, together with the previous conditions, we assume $\eta r^{\delta} \geq 4$, we can find an increasing sequence of finite partitions $(\mathscr{A}_j)_{j \geq i}$ of T and a probability measure μ such that*

$$\forall\, t \in T, \qquad \sum_{j \geq i} r^{-j}\left(D_j(A_j(t)) + \log \frac{1}{\mu(A_j(t))}\right)$$
$$\leq K(\kappa, \xi)(F(T) + r^{-i}(1 + D_{i-1}(T))).$$

The way to use this result is first to determine η small enough such that (9.19d) holds; then to take r large enough such that $\eta r^{\delta} \geq 4$.

Acknowledgments. The way to introduce majorizing measures used in Section 2 was shown to me by J. Zinn. I am grateful to T. Gowers and I. Pinelis for numerous comments.

REFERENCES

[1] ANDERSEN, N. T., GINÉ, E., OSSIANDER, M. and ZINN J. (1988). The central limit theorem and the law of the iterated logarithm for empirical processes under local conditions. *Probab. Theory Related Fields* **77** 105–119.

[2] AJTAI, H., KOMLÒS, J. and Tusnàdy, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264.

[3] BOURGAIN, J. (1989). Bounded orthogonal sets and the $\Lambda(p)$-set problem. *Acta Math.* **162** 227–246.

[4] BOURGAIN, J., PAJOR, A., SZAREK, S. J. and TOMCZAK-JAEGERMAN, N. (1989). On the duality problem for entropy numbers of operators. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **1376** 50–63. Springer, New York.

[5] FERNIQUE, X. (1971). Regularité de processus gaussiens. *Invent. Math.* **12** 304–321.

[6] FERNIQUE, X. (1975). Regularité des trajectoires des fonctions aléatoires gaussiennes. *Ecole d'Eté de Probabilités de St. Flour, 1974. Lecture Notes in Math.* **480** 1–95. Springer, New York.

[7] FERNIQUE, X. (1995). Fonctions aléatoires gaussiennes; vecteurs aléatoires gaussiens. Unpublished manuscript.

[8] IBRAGIMOV, I. A., SUDAKOV, V. N. and TSIRELSON, B. S. (1976). Norms of Gaussian sample functions. *Proceedings of the Third Japan–USSR Symposium on Probability Theory. Lecture Notes in Math.* **550** 2–41. Springer, New York.

[9] KOLMOGOROV, A. and TICHOMIROV, V. (1959). *Uspekhi Mat. Nauk* **14** 3–86.

[10] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Space*. Springer, New York.

[11] LINDENSTRAUSS, J. and TZAFRIRI, L. (1979). *Classical Banach Spaces* **2**. Springer, New York.

[12] MITYAGIN, B. (1961). *Uspekhi Mat. Nauk* **16** 63–132 [English translation in *Russian Math. Surveys* **16** 59–128].

[13] PISIER, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Harmonic Analysis and Probability, Univ. Connecticut 1980–81. Lecture Notes in Math.* **995** 123–154. Springer, New York.

[14] TALAGRAND, M. (1987). Regularity of Gaussian processes. *Acta Math.* **159** 99–149.

[15] TALAGRAND, M. (1988). The structure of sign-invariant GB-sets, and of certain Gaussian measures. *Ann. Probab.* **16** 172–179.

[16] TALAGRAND, M. (1990). Sample boundedness of stochastic processes under increment conditions. *Ann Probab.* **18** 1–49.

[17] TALAGRAND, M. (1992). A simple proof of the majorizing measure theorem. *Geometric and Functional Analysis* **2** 119–125.

[18] TALAGRAND, M. (1993). Regularity of infinitely divisible processes. *Ann. Probab.* **21** 362–432.

[19] TALAGRAND, M. (1994). Matching theorems and discrepancy computations using majorizing measures. *J. Amer. Math. Soc.* **94** 455–537.

[20] TALAGRAND, M. (1994). The supremum of some canonical processes. *Amer. J. Math.* **116** 284–325.

[21] TALAGRAND, M. (1994). Construction of majorizing measures, Bernoulli processes and cotype. *Geometric and Functional Analysis* **4** 660–717.

[22] TALAGRAND, M. (1994). The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.* **22** 919–959.

[23] TALAGRAND, M. (1995). Sections of smooth convex bodies via majorizing measures. *Acta Math.* **175** 273–300.

[24] TALAGRAND, M. (1996). A new look at independence. *Ann. Probab.* **24** 1–34.

EQUIPE D'ANALYSE-TOUR 46
U. A. AU C.N.R.S NO. 294
4 PLACE JUSSIEU
75230 PARIS CEDEX 05
FRANCE
E-MAIL: talagran@math.ohio-state.edu