2. Gaussian Processes and the Generic Chaining

2.1 Overview

The overview of this chapter is given in Chapter 1, Section 1.4. More generally, Section $1.n$ is the overview of Chapter $n-2$.

2.2 The Generic Chaining

In this section we consider a metric space (T,d) and a process $(X_t)_{t\in T}$ that satisfies the increment condition:

$$\forall u > 0 \ , \ \mathsf{P}(|X_s - X_t| \geq u) \leq 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right) \ . \tag{1.4}$$

In particular this is the case when $(X_t)_{t\in T}$ is a Gaussian process and $d(s,t)^2 = \mathsf{E}(X_s - X_t)^2$. Unless explicitly specified otherwise (and even when we forget to repeat it) we will *always* assume that the process is centered, i.e.

$$\forall t \in T \ , \ \ \mathsf{E}X_t = 0 \ . \tag{2.1}$$

We will measure the "size of the process $(X_t)_{t\in T}$" by the quantity $\mathsf{E}\sup_{t\in T} X_t$. (The reader who is impatient to understand why this quantity is a good measure of the "size of the process" can peek ahead to Lemma 2.2.1 below.)

A side issue (in particular when T is uncountable) is that what is meant by the quantity $\mathsf{E}\sup_{t\in T} X_t$ is not obvious. An efficient method is to *define* this quantity by the following formula:

$$\mathsf{E}\sup_{t\in T} X_t = \sup\Big\{\mathsf{E}\sup_{t\in F} X_t \ ; \ F \subset T \ , \ F \text{ finite}\Big\} \ , \tag{2.2}$$

where the right-hand side makes sense as soon as each r.v. X_t is integrable. This will be the case in almost all the situations considered in this book. For the next few dozens of pages, we make the effort to explain in every case how to reduce the study of the supremum of the r.v.s under consideration to the supremum of a finite family, until the energy available for this sterile exercise runs out, see Section 1.2.

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_2,

Let us say that a process $(X_t)_{t\in T}$ is *symmetric* if it has the same law as the process $(-X_t)_{t\in T}$. Almost all the processes we shall consider are symmetric (although for some of our results this hypothesis is not necessary). The following justifies using the quantity $\mathsf{E}\sup_t X_t$ to measure "the size of a symmetric process".

Lemma 2.2.1. *If the process* $(X_t)_{t\in T}$ *is symmetric then*

$$\mathsf{E}\sup_{s,t\in T}|X_s - X_t| = 2\mathsf{E}\sup_{t\in T} X_t\ .$$

Proof. We note that

$$\sup_{s,t\in T}|X_s - X_t| = \sup_{s,t\in T}(X_s - X_t) = \sup_{s\in T} X_s + \sup_{t\in T}(-X_t)\,,$$

and we take expectations. $\square$

Exercise 2.2.2. Consider a symmetric process $(X_t)_{t\in T}$. Given any t_0 in T prove that

$$\mathsf{E}\sup_{t\in T}|X_t| \le 2\mathsf{E}\sup_{t\in T} X_t + \mathsf{E}|X_{t_0}| \le 3\mathsf{E}\sup_{t\in T}|X_t|\ . \tag{2.3}$$

Generally speaking, and unless mentioned otherwise, the exercises have been designed to be easy. The author however never taught this material in a classroom, so it might happen that some exercises are not that easy after all for the beginner. Please do not be discouraged if this should be the case. (In fact, as it would have taken supra-human dedication for the author to write in detail all the solutions, there is no real warranty that each of the exercise is really feasible or even correct.) The exercises have been designed to shed some light on the material at hand, and to shake the reader out of her natural laziness by inviting her to manipulate some simple objects. (Please note that it is probably futile to sue me over the previous statement, since the reader is referred as "she" through the entire book and not only in connection with the word "laziness".)

In this book, we often state inequalities about the supremum of a symmetric process using the quantity $\mathsf{E}\sup_{t\in T} X_t$ simply because this quantity looks typographically more elegant than the equivalent quantity $\mathsf{E}\sup_{s,t\in T}|X_s - X_t|$. Of course, it is not always enough to control the first moment of $\sup_{s,t\in T}|X_s - X_t|$. We also need to control the tails of this r.v. Emphasis is given to the first moment simply because, as the reader will eventually realize, this is the difficult part, and once this is achieved, control of higher moments is often provided by the same arguments.

Our goal is to find bounds for $\mathsf{E}\sup_{t\in T} X_t$ depending on the structure of the metric space (T,d). We will assume that T is finite, which, as shown by (2.2), does not decrease generality.

Given any t_0 in T, the centering hypothesis (2.1) implies

$$\mathsf{E}\sup_{t\in T} X_t = \mathsf{E}\sup_{t\in T}(X_t - X_{t_0}) \; . \tag{2.4}$$

The latter form has the advantage that we now seek estimates for the expectation of the non-negative r.v. $Y = \sup_{t\in T}(X_t - X_{t_0})$. Then,

$$\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y > u)\,\mathrm{d}u \; . \tag{2.5}$$

Thus it is natural to look for bounds of

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \ge u\Big) \; . \tag{2.6}$$

The first bound that comes to mind is the "union bound"

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \ge u\Big) \le \sum_{t\in T} \mathsf{P}(X_t - X_{t_0} \ge u) \; . \tag{2.7}$$

It seems worthwhile to draw right away some consequences from this bound, and to discuss at leisure a number of other simple, yet fundamental facts. This will take a bit over three pages, after which we will come back to the main story of bounding Y. Throughout this work, $\Delta(T)$ denotes the diameter of T,

$$\Delta(T) = \sup_{t_1,t_2\in T} d(t_1,t_2) \; . \tag{2.8}$$

When we need to make clear which distance we use in the definition of the diameter, we will write $\Delta(T,d)$ rather than $\Delta(T)$. Consequently (1.4) and (2.7) imply

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \ge u\Big) \le 2\,\mathrm{card}\,T \exp\Big(-\frac{u^2}{2\Delta(T)^2}\Big) \; . \tag{2.9}$$

Let us now record a simple yet important computation, that will allow us to use the information (2.9).

Lemma 2.2.3. *Consider a r.v. $Y \ge 0$ which satisfies*

$$\forall\, u > 0\;,\; \mathsf{P}(Y \ge u) \le A\exp\Big(-\frac{u^2}{B^2}\Big) \tag{2.10}$$

for certain numbers $A \ge 2$ and $B > 0$. Then

$$\mathsf{E}Y \le LB\sqrt{\log A} \; . \tag{2.11}$$

Here, as in the entire book, L denotes a universal constant. We make the convention that this constant **is not necessarily** the same on each occurrence. This convention is very convenient, but it certainly needs to get used to, as e.g. in the formula $\sup_x xy - Lx^2 = y^2/L$. This convention should be remembered at all times.

When meeting an unknown notation such as this previous L, the reader might try to look at the **index**, where some of the most common notation is recorded.

Proof of Lemma 2.2.3. We use (2.5) and we observe that since $\mathsf{P}(Y \ge u) \le 1$, then for any number u_0 we have

$$\begin{aligned}\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y \ge u)\mathrm{d}u &= \int_0^{u_0} \mathsf{P}(Y \ge u)\mathrm{d}u + \int_{u_0}^\infty \mathsf{P}(Y \ge u)\mathrm{d}u \\ &\le u_0 + \int_{u_0}^\infty A\exp\Big(-\frac{u^2}{B^2}\Big)\mathrm{d}u \\ &\le u_0 + \frac{1}{u_0}\int_{u_0}^\infty uA\exp\Big(-\frac{u^2}{B^2}\Big)\mathrm{d}u \\ &= u_0 + \frac{AB^2}{2u_0}\exp\Big(-\frac{u_0^2}{B^2}\Big)\,, \qquad (2.12)\end{aligned}$$

and the choice of $u_0 = B\sqrt{\log A}$ completes the proof. □

Combining (2.11) and (2.9) we obtain that (considering separately the case where $\operatorname{card} T = 1$)

$$\mathsf{E}\sup_{t\in T} X_t \le L\Delta(T)\sqrt{\log\operatorname{card} T}\,. \qquad (2.13)$$

The following special case is fundamental.

Lemma 2.2.4. *If $(g_k)_{k\ge 1}$ are standard Gaussian r.v.s then*

$$\mathsf{E}\sup_{k\le N} g_k \le L\sqrt{\log N}\,. \qquad (2.14)$$

Exercise 2.2.5. (a) Prove that (2.14) holds as soon as the r.v.s g_k are centered and satisfy

$$\mathsf{P}(g_k \ge t) \le 2\exp\Big(-\frac{t^2}{2}\Big) \qquad (2.15)$$

for $t > 0$.
(b) For $N \ge 2$ construct N centered r.v.s $(g_k)_{k\le N}$ satisfying (2.15), and taking only the values $0, \pm\sqrt{\log N}$ and for which $\mathsf{E}\sup_{k\le N} g_k \ge \sqrt{\log N}/L$. (You are not yet asked to make these r.v.s independent.)
(c) After learning (2.16) below, solve (b) with the further requirement that the r.v.s g_k are independent. If this is too hard, look at Exercise 2.2.7, (b) below.

This is taking us a bit ahead, but an equally fundamental fact is that when the r.v.s (g_k) are jointly Gaussian, and "significantly different from each other" i.e. $\mathsf{E}(g_k - g_\ell)^2 \ge a^2 > 0$ for $k \ne \ell$, the bound (2.14) can be reversed, i.e. $\mathsf{E}\sup_{k\le N} g_k \ge a\sqrt{\log N}/L$, a fact known as Sudakov's minoration. Sudakov's minoration is a non-trivial fact, but it should be really helpful to solve Exercise 2.2.7 below. Before that let us point out a simple fact, that will be used many times.

Exercise 2.2.6. Consider independent events $(A_k)_{k\geq 1}$. Prove that

$$\mathsf{P}\Big(\bigcup_{k\leq N} A_k\Big) \geq 1 - \exp\Big(-\sum_{k\leq N}\mathsf{P}(A_k)\Big)\,. \tag{2.16}$$

(Hint: $\mathsf{P}(\cup_{k\leq N} A_k) = 1 - \prod_{k\leq N}(1-\mathsf{P}(A_k))$.)

In words: independent events such that the sum of their probabilities is small are basically disjoint.

Exercise 2.2.7. (a) Consider independent r.v.s $Y_k \geq 0$ and $u > 0$ with

$$\sum_{k\leq N}\mathsf{P}(Y_k \geq u) \geq 1\,. \tag{2.17}$$

Prove that

$$\mathsf{E}\sup_{k\leq N} Y_k \geq \frac{u}{L}\,.$$

(Hint: use (2.16) to prove that $\mathsf{P}(\sup_{k\leq N} Y_k \geq u) \geq 1/L$.)
(b) We assume (2.17), but now Y_k need not be ≥ 0. Prove that

$$\mathsf{E}\sup_{k\leq N} Y_k \geq \frac{u}{L} - \mathsf{E}|Y_1|\,.$$

(Hint: observe that for each event Ω we have $\mathsf{E}\mathbf{1}_\Omega \sup_k Y_k \geq -\mathsf{E}|Y_1|$.)
(c) Prove that if $(g_k)_{k\geq 1}$ are independent standard Gaussian r.v.s then $\mathsf{E}\sup_{k\leq N} g_k \geq \sqrt{\log N}/L$.

Before we go back to our main story, it might be worth for the reader to consider in detail consequences of an “exponential decay of tails” such as in (2.10). This is the point of the next exercise.

Exercise 2.2.8. (a) Assume that for a certain $B > 0$ the r.v. $Y \geq 0$ satisfies

$$\forall\, u > 0\,,\ \mathsf{P}(Y \geq u) \leq 2\exp\Big(-\frac{u}{B}\Big)\,. \tag{2.18}$$

Prove that

$$\mathsf{E}\exp\Big(\frac{Y}{2B}\Big) \leq L\,. \tag{2.19}$$

Prove that for $a > 0$ one has $(x/a)^a \leq \exp x$. Use this for $a = p$ and $x = Y/2B$ to deduce from (2.19) that for $p \geq 1$ one has

$$(\mathsf{E}Y^p)^{1/p} \leq LpB\,. \tag{2.20}$$

(b) Assuming now that for a certain $B > 0$ one has

$$\forall\, u > 0\,,\ \mathsf{P}(Y \geq u) \leq 2\exp\Big(-\frac{u^2}{B^2}\Big)\,, \tag{2.21}$$

prove similarly (or deduce from (a)) that $\mathsf{E}\exp(Y^2/2B^2) \leq L$ and that for $p \geq 1$ one has

$$(\mathsf{E}Y^p)^{1/p} \leq LB\sqrt{p}\,. \tag{2.22}$$

In words, (2.20) states that "as p increases, the L^p norm of an exponentially integrable r.v. does not grow faster than p," and (2.22) asserts that if the square of the r.v. is exponentially integrable, then its L^p norm does not grow faster than $\sqrt{p}$. (These two statements are closely related.) More generally it is very classical to relate the size of the tails of a r.v. with the rate of growth of its L^p norm. This is not explicitly used in the sequel, but is good to know as background information. As the following shows, (2.22) provides the correct rate of growth in the case of Gaussian r.v.s.

Exercise 2.2.9. If g is a standard Gaussian r.v. it follows from (2.22) that for $p \geq 1$ one has $(\mathsf{E}|g|^p)^{1/p} \leq L\sqrt{p}$. Prove one has also

$$(\mathsf{E}|g|^p)^{1/p} \geq \frac{\sqrt{p}}{L} \,. \tag{2.23}$$

One knows how to compute exactly $\mathsf{E}|g|^p$, from which one can deduce (2.23). You are however asked to provide a proof in the spirit of this work by deducing (2.23) solely from the information that, say, for $u > 0$ we have (choosing on purpose crude constants) $\mathsf{P}(|g| \geq u) \geq \exp(-u^2/3)/L$.

You will find basically no exact computations in this book. The aim is different. We study quantities which are far too complicated to be computed exactly, and we try to bound them from above, and sometimes from below by simpler quantities with as little a gap as possible between the upper and the lower bounds, the gap being ideally only a multiplicative constant.

We go back to our main story. The bound (2.13) will be effective if the variables $X_t - X_{t_0}$ are rather uncorrelated (and if there are not too many of them). But it will be a disaster if many of the variables $(X_t)_{t \in T}$ are nearly identical. Thus it seems a good idea to gather those variables X_t which are nearly identical. To do this, we consider a subset T_1 of T, and for t in T we consider a point $\pi_1(t)$ in T_1, which we think of as a (first) approximation of t. The elements of T to which corresponds the same point $\pi_1(t)$ are, at this level of approximation, considered as identical. We then write

$$X_t - X_{t_0} = X_t - X_{\pi_1(t)} + X_{\pi_1(t)} - X_{t_0} \,. \tag{2.24}$$

The idea is that it will be effective to use (2.7) for the variables $X_{\pi_1(t)} - X_{t_0}$, because there are not too many of them, and they are rather different (at least in some global sense and if we have done a good job at finding $\pi_1(t)$). On the other hand, since $\pi_1(t)$ is an approximation of t, the variables $X_t - X_{\pi_1(t)}$ are "smaller" than the original variables $X_t - X_{t_0}$, so that their supremum should be easier to handle. The procedure will then be iterated.

Let us set up the general procedure. For $n \geq 0$, we consider a subset T_n of T, and for $t \in T$ we consider $\pi_n(t)$ in T_n. (The idea is of course that the points $\pi_n(t)$ are successive approximations of t.) We assume that T_0 consists of a single element t_0, so that $\pi_0(t) = t_0$ for each t in T. The fundamental relation is

$$X_t - X_{t_0} = \sum_{n \geq 1} \big(X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \big) \,, \tag{2.25}$$

which holds provided we arrange that $\pi_n(t) = t$ for n large enough, in which case the series is actually a finite sum. Relation (2.25) decomposes the increments of the process $X_t - X_{t_0}$ along the "chain" $(\pi_n(t))_{n \geq 0}$ (and this is why this method is called "chaining").

It will be convenient to control the set T_n through its cardinality, with the condition

$$\operatorname{card} T_n \leq N_n \tag{2.26}$$

where

$$N_0 = 1 \,;\; N_n = 2^{2^n} \quad \text{if} \quad n \geq 1 \,. \tag{2.27}$$

The notation (2.27) will be used throughout the book. The reader who has studied Appendix A will observe that the procedure to control T_n is rather different here. This is a crucial point of the generic chaining method.

It is good to notice right away that $\sqrt{\log N_n}$ is about $2^{n/2}$, which explains the ubiquity of this latter quantity. The occurrence of the function $\sqrt{\log x}$ itself is related to the fact that in some sense this is the inverse of the function $\exp(-x^2)$ that governs the size of the tails of a Gaussian r.v. Let us also observe the fundamental inequality

$$N_n^2 \leq N_{n+1} \,,$$

which makes it very convenient to work with this sequence.

Since $\pi_n(t)$ approximates t, it is natural to assume that

$$d(t, \pi_n(t)) = d(t, T_n) = \inf_{s \in T_n} d(t, s) \,. \tag{2.28}$$

For $u > 0$, (1.4) implies

$$\mathsf{P}\big(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \geq u 2^{n/2} d(\pi_n(t), \pi_{n-1}(t)) \big) \leq 2 \exp(-u^2 2^{n-1}) \,.$$

The number of possible pairs $(\pi_n(t), \pi_{n-1}(t))$ is bounded by

$$\operatorname{card} T_n \cdot \operatorname{card} T_{n-1} \leq N_n N_{n-1} \leq N_{n+1} = 2^{2^{n+1}} \,.$$

Thus, if we denote by Ω_u the event defined by

$$\forall n \geq 1 \,,\, \forall t \,,\, |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \leq u 2^{n/2} d(\pi_n(t), \pi_{n-1}(t)) \,, \tag{2.29}$$

we obtain

$$\mathsf{P}(\Omega_u^c) \leq p(u) := \sum_{n \geq 1} 2 \cdot 2^{2^{n+1}} \exp(-u^2 2^{n-1}) \,. \tag{2.30}$$

Here again, at the crucial step, we have used the "union bound": indeed we bound the probability that one of the events (2.29) fails by the sum of the probabilities that the individual events fail. When Ω_u occurs, (2.25) yields

$$|X_t - X_{t_0}| \le u \sum_{n\ge 1} 2^{n/2} d(\pi_n(t), \pi_{n-1}(t))\,,$$

so that

$$\sup_{t\in T} |X_t - X_{t_0}| \le uS$$

where

$$S := \sup_{t\in T} \sum_{n\ge 1} 2^{n/2} d(\pi_n(t), \pi_{n-1}(t))\,.$$

Thus

$$\mathsf{P}\Big(\sup_{t\in T} |X_t - X_{t_0}| > uS\Big) \le p(u)\,.$$

For $n \ge 1$ and $u \ge 3$ we have

$$u^2 2^{n-1} \ge \frac{u^2}{2} + u^2 2^{n-2} \ge \frac{u^2}{2} + 2^{n+1}\,,$$

from which it follows that

$$p(u) \le L \exp\Big(-\frac{u^2}{2}\Big)\,.$$

We observe here that since $p(u) \le 1$ the previous inequality holds not only for $u \ge 3$ but also for $u > 0$. (This type or argument will be used repeatedly.) Therefore

$$\mathsf{P}\Big(\sup_{t\in T} |X_t - X_{t_0}| > uS\Big) \le L \exp\Big(-\frac{u^2}{2}\Big)\,. \tag{2.31}$$

In particular (2.31) implies

$$\mathsf{E} \sup_{t\in T} X_t \le LS\,.$$

The triangle inequality and (2.5) yield

$$\begin{aligned} d(\pi_n(t), \pi_{n-1}(t)) &\le d(t, \pi_n(t)) + d(t, \pi_{n-1}(t)) \\ &\le d(t, T_n) + d(t, T_{n-1})\,, \end{aligned}$$

so that $S \le L \sup_{t\in T} \sum_{n\ge 0} 2^{n/2} d(t, T_n)$, and we have proved the fundamental bound

$$\mathsf{E} \sup_{t\in T} X_t \le L \sup_{t\in T} \sum_{n\ge 0} 2^{n/2} d(t, T_n)\,. \tag{2.32}$$

Now, how do we construct the sets T_n? It is obvious that we should try to make the right-hand side of (2.32) small, but this is obvious only because we have used an approach which naturally leads to this bound. The "traditional chaining method" (as used e.g. in Appendix A) chooses them so that

$$\sup_{t\in T} d(t, T_n)$$

is as small as possible for $\operatorname{card} T_n \le N_n$, where of course

$$d(t, T_n) = \inf_{s \in T_n} d(t, s) \; . \tag{2.33}$$

Thus we define

$$e_n(T) = e_n(T, d) = \inf \sup_t d(t, T_n) \, , \tag{2.34}$$

where the infimum is taken over all subsets T_n of T with $\operatorname{card} T_n \le N_n$. (Since here T is finite, the infimum is actually a minimum.) We call the numbers $e_n(T)$ the **entropy numbers**. This definition is convenient for our purposes. It is unfortunate that it is not consistent with the conventions of Operator Theory, which uses e_{2^n} to denote what we call e_n, but we can't help it if Operator Theory gets it wrong. When T is infinite, the numbers $e_n(T)$ are also defined by (2.34) but are not always finite (e.g. when $T = \mathbb{R}$).

It is good to observe that (since $N_0 = 1$) ,

$$\frac{\Delta(T)}{2} \le e_0(T) \le \Delta(T) \; . \tag{2.35}$$

Recalling that T is finite, let us then choose for each n a subset T_n of T with $\operatorname{card} T_n \le N_n$ and $e_n(T) = \sup_{t \in T} d(t, T_n)$. Since $d(t, T_n) \le e_n(T)$ for each t, (2.32) implies the following.

Proposition 2.2.10 (Dudley's entropy bound [2]). *Under the increment condition (1.4), we have*

$$\mathsf{E} \sup_{t \in T} X_t \le L \sum_{n \ge 0} 2^{n/2} e_n(T) \; . \tag{2.36}$$

We proved this bound only when T is finite, but using (2.2) it also extends to the case where T is infinite, as is shown by the following easy fact.

Lemma 2.2.11. *If U is a subset of T, we have $e_n(U) \le 2e_n(T)$.*

Proof. Indeed, if $a > e_n(T)$, by definition one can cover T by N_n balls (for the distance d) with radius a, and the intersections of these balls with U are of diameter $\le 2a$, so U can be covered by N_n balls in U with radius $2a$. □

The reader already familiar with Dudley's entropy bound might not recognize it. Usually this bound is formulated as in (1.5) using covering numbers. The covering number $N(T, d, \epsilon)$ is defined to be the smallest integer N such that there is a subset F of T, with $\operatorname{card} F \le N$ and

$$\forall t \in T \; , \; d(t, F) \le \epsilon \; .$$

The covering numbers relate to the entropy numbers by the formula

$$e_n(T) = \inf\{\epsilon \; ; \; N(T, d, \epsilon) \le N_n\} \, .$$

Indeed, it is obvious by definition of $e_n(T)$ that for $\epsilon > e_n(T)$, we have $N(T,d,\epsilon) \le N_n$, and that if $N(T,d,\epsilon) \le N_n$ we have $e_n(T) \le \epsilon$. Consequently,

$$\begin{aligned}\epsilon < e_n(T) &\Rightarrow N(T,d,\epsilon) > N_n \\ &\Rightarrow N(T,d,\epsilon) \ge 1 + N_n \ .\end{aligned}$$

Therefore

$$\sqrt{\log(1+N_n)}(e_n(T) - e_{n+1}(T)) \le \int_{e_{n+1}(T)}^{e_n(T)} \sqrt{\log N(T,d,\epsilon)}\, \mathrm{d}\epsilon \ .$$

Since $\log(1+N_n) \ge 2^n \log 2$ for $n \ge 0$, summation over $n \ge 0$ yields

$$\sqrt{\log 2} \sum_{n\ge 0} 2^{n/2}(e_n(T) - e_{n+1}(T)) \le \int_0^{e_0(T)} \sqrt{\log N(T,d,\epsilon)}\, \mathrm{d}\epsilon \ . \tag{2.37}$$

Now,

$$\begin{aligned}\sum_{n\ge 0} 2^{n/2}(e_n(T) - e_{n+1}(T)) &= \sum_{n\ge 0} 2^{n/2} e_n(T) - \sum_{n\ge 1} 2^{(n-1)/2} e_n(T) \\ &\ge \left(1 - \frac{1}{\sqrt{2}}\right) \sum_{n\ge 0} 2^{n/2} e_n(T)\,,\end{aligned}$$

so (2.37) yields

$$\sum_{n\ge 0} 2^{n/2} e_n(T) \le L \int_0^\infty \sqrt{\log N(T,d,\epsilon)}\, \mathrm{d}\epsilon\,.$$

Hence Dudley's bound now appears in the familiar form

$$\mathsf{E} \sup_{t\in T} X_t \le L \int_0^\infty \sqrt{\log N(T,d,\epsilon)}\, \mathrm{d}\epsilon \ . \tag{2.38}$$

Of course, since $\log 1 = 0$, the integral takes place in fact over $0 \le \epsilon \le \Delta(T)$. The right-hand side is often called Dudley's entropy integral.

Exercise 2.2.12. Prove that

$$\int_0^\infty \sqrt{\log N(T,d,\epsilon)}\, \mathrm{d}\epsilon \le L \sum_{n\ge 0} 2^{n/2} e_n(T)\,,$$

showing that (2.36) is not an improvement over (2.38).

Exercise 2.2.13. Assume that for each $\epsilon > 0$ we have $\log N(t,d,\epsilon) \le (A/\epsilon)^\alpha$. Prove that $e_n(T) \le K(\alpha) A 2^{-n/\alpha}$.

Here of course $K(\alpha)$ is a number depending only on α. This, and similar notation are used throughout the book. It is understood that such numbers *need not be the same on every occurrence* and it would help to **remember this at all times**. The difference between the notations K and L is that L is a universal constant, i.e. a number that does not depend on anything, while K might depend on some parameters, such as α here.

How does one estimate covering numbers (or, equivalently, entropy numbers)? The next exercise introduces the reader to "volume estimates", a simple yet fundamental method for this purpose. It deserves to be fully understood. If this exercise is too hard, you can find all the details below in the proof of Lemma 2.5.5.

Exercise 2.2.14. (a) If (T, d) is a metric space, define the packing number $N^*(T, d, \epsilon)$ as the largest integer N such that T contains N points with mutual distances $\geq \epsilon$. Prove that $N(T, d, \epsilon) \leq N^*(T, d, \epsilon)$. Prove that if $\epsilon' > 2\epsilon$ then $N^*(T, d, \epsilon') \leq N(T, d, \epsilon)$.
(b) Let us denote by d the Euclidean distance in $\mathbb{R}^m$, and by B the unit Euclidean ball of center 0. Let us denote by $\mathrm{Vol}(A)$ the m-dimensional volume of a subset A of $\mathbb{R}^m$. By comparing volumes, prove that for any subset A of $\mathbb{R}^m$,

$$N(A, d, \epsilon) \geq \frac{\mathrm{Vol}(A)}{\mathrm{Vol}(\epsilon B)} \tag{2.39}$$

and

$$N(A, d, 2\epsilon) \leq N^*(A, d, 2\epsilon) \leq \frac{\mathrm{Vol}(A + \epsilon B)}{\mathrm{Vol}(\epsilon B)} . \tag{2.40}$$

(c)Conclude that

$$\left(\frac{1}{\epsilon}\right)^m \leq N(B, d, \epsilon) \leq \left(\frac{2+\epsilon}{\epsilon}\right)^m . \tag{2.41}$$

(d) Use (c) to find estimates of $e_n(B)$ for the correct order for each value of n. (Hint: $e_n(B)$ is about $\min(1, 2^{-2^n/m})$. This decreases very fast as n increases.) Estimate Dudley's bound for B provided with the Euclidean distance.
(e) Use (c) to prove that if T is a subset of $\mathbb{R}^m$ and if n_0 is any integer such that $m2^{-n_0} \leq 1$ then for $n > n_0$ one has $e_n(T) \leq L2^{-2^n/2m} e_{n_0}(T)$. (Hint: cover T by N_{n_0} balls of radius $2e_{n_0}(T)$ and cover each of these by balls of smaller radius using (c).)
(f) This part provides a generalization of (2.39) and (2.40) to a more abstract setting, but with the same proofs. Consider a metric space (T, d) and a positive measure μ on T such all balls of a given radius have the same measure, $\mu(B(t, \epsilon)) = \varphi(\epsilon)$ for each $\epsilon > 0$ and $t \in T$. For a subset A of T and $\epsilon > 0$ let $A_\epsilon = \{t \in T; d(t, A) \leq \epsilon\}$, where $d(t, A) = \inf_{s \in A} d(t, s)$. Prove that

$$\frac{\mu(A)}{\varphi(2\epsilon)} \leq N(A, d, 2\epsilon) \leq \frac{\mu(A_\epsilon)}{\varphi(\epsilon)} .$$

There are many simple situations where Dudley's bound is not of the correct order. Although this takes us a bit ahead, we give such an example in the next exercise. There the set T is particularly appealing: it is a simplex in $\mathbb{R}^m$. Another classical example which is in a sense canonical occurs on page 44. Yet other examples based on fundamental geometry (ellipsoids in $\mathbb{R}^m$) are explained in Section 2.5.

Exercise 2.2.15. Consider an integer m and an i.i.d. standard Gaussian sequence $(g_i)_{i \le m}$. For $t = (t_i)_{i \le m}$, let $X_t = \sum_{i \le m} t_i g_i$. This is called the canonical Gaussian process on $\mathbb{R}^m$. Its associated distance is the Euclidean distance on $\mathbb{R}^m$. It will be much used later. Consider the set

$$T = \Big\{ (t_i)_{i \le m} \; ; \; t_i \ge 0 \, , \, \sum_{i \le m} t_i = 1 \Big\} \, , \tag{2.42}$$

the convex hull of the canonical basis. By (2.14) we have $\mathsf{E} \sup_{t \in T} X_t = \mathsf{E} \sup_{i \le m} g_i \le L \sqrt{\log m}$. Prove that however the right-hand side of (2.36) is $\ge (\log m)^{3/2}/L$. (Hint: For an integer $k \le m$ consider the subset T_k of T consisting of sequences $t = (t_i)_{i \le m} \in T$ for which $t_i \in \{0, 1/k\}$. Using part (f) of Exercise 2.2.14 with $T = A = T_k$ and μ the counting measure prove that $\log N(T_k, d, 1/(L\sqrt{k})) \ge k \log(em/k)/L$ and conclude. You need to be fluent with Stirling's formula to succeed.) Thus in this case Dudley's bound is off by a factor about $\log m$. Exercise 2.3.4 below will show that in $\mathbb{R}^m$ the situation cannot be worse than this.

The bound (2.32) seems to be genuinely better than the bound (2.36) because when going from (2.32) to (2.36) we have used the somewhat brutal inequality

$$\sup_{t \in T} \sum_{n \ge 0} 2^{n/2} d(t, T_n) \le \sum_{n \ge 0} 2^{n/2} \sup_{t \in T} d(t, T_n) \, .$$

The method leading to the bound (2.32) is probably the most important idea of this work. Of course the fact that it appears now so naturally does not reflect the history of the subject, but rather that the proper approach is being used. When using this bound, we will choose the sets T_n in order to minimize the right-hand side of (2.32) instead of choosing them as in (2.34). The true importance of this procedure is that as will be demonstrated later, this provides essentially the best possible bound for $\mathsf{E} \sup_{t \in T} X_t$. To understand that matters are not trivial, the reader should try, in the situation of Exercise 2.2.15, to find sets T_n such that the right-hand side of (2.32) is of the correct order $\sqrt{\log m}$. It would probably be quite an athletic feat to succeed at this stage, but the reader is encouraged to keep this question in mind as her understanding deepens.

The next exercise provides a simple (and somewhat "extremal") situation showing that (2.32) is an actual improvement over (2.36).

Exercise 2.2.16. (a) Consider a finite metric space (T,d). Assume that it contains a point t_0 with the property that for $n \geq 0$ we have $\operatorname{card}(T \setminus B(t_0, 2^{-n/2})) \leq N_n - 1$. Prove that T contains sets T_n with $\operatorname{card} T_n \leq N_n$ and $\sup_{t\in T} \sum_{n\geq 0} 2^{n/2} d(t, T_n) \leq L$. (Hint: $T_n = \{t_0\} \cup \{t \in T; d(t,t_0) > 2^{-n/2}\}$.)
(b) Given an integer $s \geq 10$, construct a finite metric space (T,d) with the above property, such that $\operatorname{card} T \leq N_s$ and that $e_n(T) \geq 2^{-n/2}/L$ for $1 \leq n \leq s-1$, so that Dudley's integral is of order s. (Hint: this might be hard if you really never though about metric spaces. Try then a set of the type $T = \{a_\ell f_\ell; \ell \leq M\}$ where $a_\ell > 0$ is a number and $(f_\ell)_{\ell \leq M}$ is the canonical basis of $\mathbb{R}^M$.)

It turns out that the idea behind the bound (2.32) admits a technically more convenient formulation.

Definition 2.2.17. *Given a set T an* admissible sequence *is an increasing sequence $(\mathcal{A}_n)$ of partitions of T such that* $\operatorname{card} \mathcal{A}_n \leq N_n$, *i.e.* $\operatorname{card} \mathcal{A}_0 = 1$ *and* $\operatorname{card} \mathcal{A}_n \leq 2^{2^n}$ *for $n \geq 1$.*

By an increasing sequence of partitions we mean that every set of $\mathcal{A}_{n+1}$ is contained in a set of $\mathcal{A}_n$. Throughout the book we denote by $A_n(t)$ the unique element of $\mathcal{A}_n$ which contains t. The double exponential in the definition of (2.27) of N_n occurs simply since for our purposes the proper measure of the "size" of a partition $\mathcal{A}$ is $\log \operatorname{card} \mathcal{A}$. This double exponential ensures that "the size of the partition $\mathcal{A}_n$ doubles at every step". This offers a number of technical advantages which will become clear gradually.

Theorem 2.2.18 (The generic chaining bound). *Under the increment condition (1.4) (and if $\mathsf{E} X_t = 0$ for each t) then for each admissible sequence $(\mathcal{A}_n)$ we have*

$$\mathsf{E} \sup_{t\in T} X_t \leq L \sup_{t\in T} \sum_{n\geq 0} 2^{n/2} \Delta(A_n(t)) \,. \tag{2.43}$$

Here of course, as always, $\Delta(A_n(t))$ denotes the diameter of $A_n(t)$ for d. One could think that (2.43) could be much worse than (2.32), but is will turn out that this is not the case when the sequence $(\mathcal{A}_n)$ is appropriately chosen.

Proof. We may assume T to be finite. We construct a subset T_n of T by taking exactly one point in each set A of $\mathcal{A}_n$. Then for $t \in T$ and $n \geq 0$, we have $d(t, T_n) \leq \Delta(A_n(t))$ and the result follows from (2.32). □

Definition 2.2.19. *Given $\alpha > 0$, and a metric space (T,d) (that need not be finite) we define*

$$\gamma_\alpha(T,d) = \inf \sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} \Delta(A_n(t)) \,,$$

where the infimum is taken over all admissible sequences.

It is useful to observe that since $A_0(t) = T$ we have $\gamma_\alpha(T,d) \geq \Delta(T)$.

Exercise 2.2.20. (a) Prove that if $d \le d'$ then $\gamma_2(T,d) \le \gamma_2(T,d')$.
(b) More generally prove that if $d \le Bd'$ then $\gamma_2(T,d) \le B\gamma_2(T,d')$.

Exercise 2.2.21. (a) If T is finite, prove that $\gamma_2(T,d) \le L\Delta(T)\sqrt{\log \operatorname{card} T}$.
(Hint: Ensure that $\Delta(A_n(t)) = 0$ if $N_n \ge \operatorname{card} T$.)
(b) Prove that for $n \ge 0$ we have

$$2^{n/2} e_n(T) \le L\gamma_2(T,d) \; . \tag{2.44}$$

(Hint: observe that $2^{n/2}\max\{\Delta(A); A \in \mathcal{A}_n\} \le \sup_{t\in T}\sum_{n\ge 0} 2^{n/2}\Delta(A_n(t))$.)
(c) Prove that, equivalently, for $\epsilon > 0$ we have

$$\epsilon\sqrt{\log N(T,d,\epsilon)} \le L\gamma_2(T,d) \; .$$

The reader should compare (2.44) with Corollary 2.3.2 below.

Combining Theorem 2.2.18 with Definition 2.2.19 yields

Theorem 2.2.22. *Under (1.4) and (2.1) we have*

$$\mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T,d) \; . \tag{2.45}$$

Of course to make (2.45) of interest we must be able to control $\gamma_2(T,d)$, i.e. we must learn how to construct admissible sequences, a topic we shall first address in Section 2.3.

Let us also point out, recalling (2.31), and observing that

$$|X_s - X_t| \le |X_s - X_{t_0}| + |X_t - X_{t_0}| \; , \tag{2.46}$$

we have actually proved

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge Lu\gamma_2(T,d)\Big) \le 2\exp(-u^2) \; . \tag{2.47}$$

There is no reason other than the author's fancy to feature the phantom coefficient 1 in the exponent of the right-hand side, but it might be good at this stage for the reader to write every detail on how this is deduced from (2.31). The different exponents in (2.31) and (2.47) are of course made possible by the fact that the constant L is not the same in these inequalities.

We note that (2.47) implies a lot more than (2.45). Indeed, for each $p \ge 1$, using (2.22)

$$\mathsf{E}\big(\sup_{s,t} |X_s - X_t|\big)^p \le K(p)\gamma_2(T,d)^p \; , \tag{2.48}$$

and in particular

$$\mathsf{E}\big(\sup_{s,t} |X_s - X_t|\big)^2 \le L\gamma_2(T,d)^2 \; . \tag{2.49}$$

One of the ideas underlying Definition 2.2.19 is that partitions of T are really handy. For example, given a partition $\mathcal{B}$ of T whose elements are "small" for a certain distance d_1 and a partition $\mathcal{C}$ whose elements are "small" for another distance d_2, then the elements of the partition generated by $\mathcal{B}$ and $\mathcal{C}$, i.e. the partition which consists of the sets $B \cap C$ for $B \in \mathcal{B}$ and $C \in \mathcal{C}$, are "small" for both d_1 and d_2. This is illustrated in the proof of the following theorem, which applies to processes with a weaker tail condition than (1.4). This theorem will be used many times (the reason being that a classical inequality of Bernstein naturally produces tail conditions such as in (2.50)).

Theorem 2.2.23. *Consider a set T provided with two distances d_1 and d_2. Consider a centered process $(X_t)_{t\in T}$ which satisfies*

$$\forall s,t \in T\ ,\ \forall u > 0\,,$$

$$\mathsf{P}(|X_s - X_t| \geq u) \leq 2\exp\left(-\min\left(\frac{u^2}{d_2(s,t)^2}, \frac{u}{d_1(s,t)}\right)\right)\ . \tag{2.50}$$

Then

$$\mathsf{E}\sup_{s,t\in T} |X_s - X_t| \leq L(\gamma_1(T,d_1) + \gamma_2(T,d_2))\ . \tag{2.51}$$

This theorem will be applied when d_2 is the ℓ_2 distance and d_1 is the ℓ_∞ distance (but it sounds funny, when considering two distances, to call them d_2 and d_∞).

Proof. We denote by $\Delta_j(A)$ the diameter of the set A for d_j. We consider an admissible sequence $(\mathcal{B}_n)_{n\geq 0}$ such that

$$\forall t \in T\,,\ \sum_{n\geq 0} 2^n \Delta_1(B_n(t)) \leq 2\gamma_1(T,d_1) \tag{2.52}$$

and an admissible sequence $(\mathcal{C}_n)_{n\geq 0}$ such that

$$\forall t \in T\,,\ \sum_{n\geq 0} 2^{n/2} \Delta_2(C_n(t)) \leq 2\gamma_2(T,d_2)\ . \tag{2.53}$$

Of course here $B_n(t)$ is the unique element of $\mathcal{B}_n$ that contains t (etc.). We define partitions $\mathcal{A}_n$ of T as follows. We set $\mathcal{A}_0 = \{T\}$, and, for $n \geq 1$, we define $\mathcal{A}_n$ as the partition generated by $\mathcal{B}_{n-1}$ and $\mathcal{C}_{n-1}$, i.e. the partition that consists of the sets $B \cap C$ for $B \in \mathcal{B}_{n-1}$ and $C \in \mathcal{C}_{n-1}$. Thus

$$\operatorname{card} \mathcal{A}_n \leq N_{n-1}^2 \leq N_n\,,$$

and the sequence $(\mathcal{A}_n)$ is admissible. (Let us repeat here that the fundamental inequality $N_n^2 \leq N_{n+1}$ is the reason why it is so convenient to work with the sequence N_n.) For each $n \geq 0$ let us consider a set T_n that intersects each element of $\mathcal{A}_n$ in exactly one point, and for $t \in T$ let us denote by $\pi_n(t)$ the element of T_n that belongs to $A_n(t)$. To use (2.50) we observe that for $v > 0$ it implies

$$\mathsf{P}(|X_s - X_t| \ge v d_1(s,t) + \sqrt{v} d_2(s,t)) \le 2\exp(-v) ,$$

and thus, given $u \ge 1$, we have, since $u \ge \sqrt{u}$,

$$\mathsf{P}\Big(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \ge u\big(2^n d_1(\pi_n(t), \pi_{n-1}(t)) + 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t))\big)\Big) \\ \le 2\exp(-u2^n) , \tag{2.54}$$

so that, proceeding as in (2.30), with probability $\ge 1 - L\exp(-u)$ we have

$$\forall n \;,\; \forall t \;,\; |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \le u\big(2^n d_1(\pi_n(t), \pi_{n-1}(t)) \\ + 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t))\big) . \tag{2.55}$$

Now, under (2.55) we get

$$\sup_{t \in T} |X_t - X_{t_0}| \le u \sup_{t \in T} \sum_{n \ge 1} \big(2^n d_1(\pi_n(t), \pi_{n-1}(t)) + 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t))\big) .$$

When $n \ge 2$ we have $\pi_n(t), \pi_{n-1}(t) \in A_{n-1}(t) \subset B_{n-2}(t)$, so that

$$d_1(\pi_n(t), \pi_{n-1}(t)) \le \Delta_1(B_{n-2}(t)) .$$

Hence, since $d_1(\pi_1(t), \pi_0(t)) \le \Delta_1(B_0(t)) = \Delta_1(T)$, using (2.52) in the last inequality, (and remembering that the value of L need not be the same on each occurrence)

$$\sum_{n \ge 1} 2^n d_1(\pi_n(t), \pi_{n-1}(t)) \le L \sum_{n \ge 0} 2^n \Delta_1(B_n(t)) \le 2L\gamma_1(T,d) = L\gamma_1(T,d) .$$

Proceeding similarly for d_2 shows that under (2.55) we obtain

$$\sup_{s,t \in T} |X_t - X_{t_0}| \le Lu(\gamma_1(T,d_1) + \gamma_2(T,d_2)) ,$$

and therefore using (2.46),

$$\mathsf{P}\Big(\sup_{s,t \in T} |X_s - X_t| \ge Lu(\gamma_1(T,d_1) + \gamma_2(T,d_2))\Big) \le L\exp(-u) , \tag{2.56}$$

which using (2.5) implies the result. □

Exercise 2.2.24. Consider a space T equipped with two different distances d_1 and d_2. Prove that

$$\gamma_2(T, d_1 + d_2) \le L(\gamma_2(T,d_1) + \gamma_2(T,d_2)) . \tag{2.57}$$

(Hint: given an admissible sequence of partitions $\mathcal{A}_n$ (resp. $\mathcal{B}_n$) which behaves well for d_1 (resp. d_2) consider as in the beginning of the proof of Theorem 2.2.23 the partitions generated by $\mathcal{A}_n$ and $\mathcal{B}_n$.)

Exercise 2.2.25 (R. Latała, S. Mendelson). Consider a process $(X_t)_{t\in T}$ and for a subset A of T and $n \geq 0$ let

$$\Delta_n(A) = \sup_{s,t\in A} (\mathsf{E}|X_s - X_t|^{2^n})^{2^{-n}} .$$

Consider an admissible sequence of partitions $(\mathcal{A}_n)_{n\geq 0}$.
(a) Prove that

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \leq \sup_{t\in T} \sum_{n\geq 0} \Delta_n(A_n(t)) .$$

(Hint: Use chaining and (A.11) for $\varphi(x) = x^{2^n}$.)
(b) Explain why this result implies Theorem 2.2.23. (Hint: Use Exercise 2.2.8.)

The following exercise assumes that you are familiar with the contents of Appendix B. It develops the theme of "chaining with varying distances" of Exercise 2.2.25 in a different direction. Variations on this idea will turn out later to be fundamental.

Exercise 2.2.26. Assume that for $n \geq 0$ we are given a distance d_n on T and a convex function φ_n with $\varphi_n(0) = 0$, $\varphi_n(x) = \varphi_n(-x) \geq 0$. Assume that

$$\forall\, s,t \in T\ ,\ \mathsf{E}\varphi_n\Big(\frac{X_s - X_t}{d_n(s,t)}\Big) \leq 1 .$$

Consider a sequence $\epsilon_n > 0$ and assume that $N(T, d_0, \epsilon_0) = 1$. Prove that

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \leq \sum_{n\geq 0} \epsilon_n \varphi_n^{-1}(N(T, d_n, \epsilon_n)) .$$

Prove that this implies Theorem B.2.3. (Hint: simple modification of the argument of Theorem B.2.3.)

We now prove some more specialized results, which may be skipped at first reading. This is all the more the case since for many processes of importance the machinery of "concentration of measure" allows one to find very competent bounds for the quantity $\mathsf{P}(|\sup_{t\in T} X_t - \mathsf{E}\sup_{t\in T} X_t| \geq u)$. For example in the case of Gaussian processes, (2.58) below is a consequence of (2.96) and (2.45). The point of (2.58) is that it improves on (2.47) using only the increment condition (1.4).

Theorem 2.2.27. *If the process (X_t) satisfies (1.4) then for $u > 0$ one has*

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \geq L(\gamma_2(T,d) + u\Delta(T))\Big) \leq L\exp(-u^2) . \tag{2.58}$$

Proof. This is one of a very few instances where one must use some care when using the generic chaining. Consider an admissible sequence $(\mathcal{A}_n)$ of partitions with $\sup_{t\in T}\sum_{n\geq 0} 2^{n/2}\Delta(A_n(t)) \leq 2\gamma_2(T,d)$ and for each n consider a set T_n with $\operatorname{card} T_n \leq N_n$ and such that every element A of $\mathcal{A}_n$ meets T_n. Let $U_n = \cup_{q\leq n} T_q$, so that $U_0 = T_0$ and $\operatorname{card} U_n \leq 2N_n$, and the sequence (U_n) increases. For $u > 0$ consider the event $\Omega(u)$ given by

$$\forall\, n \geq 1\ ,\ \forall\, s,t \in U_n\ ,\ |X_s - X_t| \leq 2(2^{n/2}+u)d(s,t)\ , \tag{2.59}$$

so that (somewhat crudely)

$$\mathsf{P}(\Omega^c(u)) \leq 2\sum_{n\geq 1}(\operatorname{card} U_n)^2 \exp(-2(2^n+u^2)) \leq L\exp(-2u^2)\ . \tag{2.60}$$

Consider now $t\in T$. We define by induction over $q\geq 0$ integers $n(t,q)$ as follows. We start with $n(t,0)=0$, and for $q\geq 1$ we define

$$n(t,q) = \inf\Big\{n\ ;\ n\geq n(t,q-1)\ ;\ d(t,U_n) \leq \frac{1}{2}d(t,U_{n(t,q-1)})\Big\}\ . \tag{2.61}$$

We then consider $\pi_q(t) \in U_{n(t,q)}$ with $d(t,\pi_q(t)) = d(t,U_{n(t,q)})$. Thus, by induction, and denoting by t_0 the unique element of $T_0 = U_0$, for $q\geq 0$, it holds

$$d(t,\pi_q(t)) \leq 2^{-q}d(t,t_0) \leq 2^{-q}\Delta(T)\ . \tag{2.62}$$

Also, when $\Omega(u)$ occurs, using (2.59) for $n = n(t,q)$, and since $\pi_q(t)\in U_n$ and $\pi_{q-1}(t) \in U_{n(t,q-1)} \subset U_n$,

$$|X_{\pi_q(t)} - X_{\pi_{q-1}(t)}| \leq 2(2^{n(t,q)/2}+u)d(\pi_q(t),\pi_{q-1}(t))\ .$$

Assuming that $\Omega(u)$ occurs, we thus obtain

$$\begin{aligned}
|X_t - X_{t_0}| &\leq \sum_{q\geq 1}|X_{\pi_q(t)} - X_{\pi_{q-1}(t)}| \\
&\leq \sum_{q\geq 1} 2(2^{n(t,q)/2}+u)d(\pi_q(t),\pi_{q-1}(t)) \\
&\leq \sum_{q\geq 1} 2(2^{n(t,q)/2}+u)d(t,\pi_q(t)) \\
&\quad + \sum_{q\geq 1} 2(2^{n(t,q)/2}+u)d(t,\pi_{q-1}(t))\ .
\end{aligned} \tag{2.63}$$

We now control the four summations on the right-hand side. First,

$$\begin{aligned}
\sum_{q\geq 1} 2^{n(t,q)/2}d(t,\pi_q(t)) &\leq \sum_{q\geq 1} 2^{n(t,q)/2}d(t,T_{n(t,q)}) \\
&\leq \sum_{n\geq 0} 2^{n/2}d(t,T_n) \leq 2\gamma_2(T,d)\ ,
\end{aligned}$$

since by definition of $n(t,q)$ we have $n(t,q) > n(t,q-1)$ unless $d(t,T_{n(t,q-1)}) = 0$. Now, the definition of $n(t,q)$ implies

$$d(t,\pi_{q-1}(t)) = d(t,U_{n(t,q-1)}) \le 2d(t,U_{n(t,q)-1}) ,$$

so that

$$\sum_{q\ge 1} 2^{n(t,q)/2} d(t,\pi_{q-1}(t)) \le 2\sum_{q\ge 1} 2^{n(t,q)/2} d(t,T_{n(t,q)-1}) ,$$

and as above this is $\le L\gamma_2(T,d)$. Next, (2.62) implies $\sum_{q\ge 1} d(\pi_q(t),t) \le 2\Delta(T)$ and $\sum_{q\ge 1} d(\pi_{q-1}(t),t) \le 2\Delta(T)$. In summary, when $\Omega(u)$ occurs, we have $|X_t - X_{t_0}| \le L(\gamma_2(T,d) + u\Delta(T))$. $\square$

One idea underlying the proof (and in particular the definition (2.61) of $n(t,q)$) is that for an efficient chaining the distance $d(\pi_n(t),\pi_{n+1}(t))$ decreases geometrically. In Chapter 15 we shall later see situations where this is not the case.

We will at times need the following more precise version of (2.56), in the spirit of Theorem 2.2.27.

Theorem 2.2.28. *Under the conditions of Theorem 2.2.23, for all values* $u_1, u_2 > 0$ *we have*

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge L\big(\gamma_1(T,d_1) + \gamma_2(T,d_2) + u_1 D_1 + u_2 D_2\big)\Big) \le L\exp(-\min(u_2^2, u_1)) , \tag{2.64}$$

where for $j = 1,2$ *we set* $D_j = \sum_{n\ge 0} e_n(T,d_j)$.

We observe from (2.44) that $e_n(T,d_2) \le L2^{-n/2}\gamma_2(T,d_2)$ so that by summation $D_2 \le L\gamma_2(T,d_2)$ and similarly for D_1. Thus (2.64) recovers (2.56). Moreover, in many practical situations, one has $D_j \le Le_0(T,d_j) \le L\Delta_j(T) = L\Delta(T,d_j)$. Still, the occurrence of the unwieldy quantities D_j makes the statement of Theorem 2.2.28 a bit awkward. It would be pleasing if in the statement of this theorem one could replace D_j by the smaller quantity $\Delta_j(T)$. Unfortunately this does not seem to be true. The reader might like to consider the case where card $T = N_n$ and $d_1(s,t) = 1$ for $s \ne t$ to understand where the difficulty lies.

Proof. There exists a partition $\mathcal{U}_n$ of T into N_n sets, each of which have a diameter $\le 2e_n(T,d_1)$ for d_1. Consider the partition $\mathcal{B}'_n$ generated by $\mathcal{U}_0,\dots,\mathcal{U}_{n-1}$. These partitions form an admissible sequence such that

$$\forall B \in \mathcal{B}'_n ,\ \Delta_1(B) \le 2e_{n-1}(T,d_1) . \tag{2.65}$$

Let us also consider an admissible sequence $(\mathcal{C}'_n)$ which has the same property for d_2,

$$\forall C \in \mathcal{C}'_n\,,\ \Delta_2(C) \le 2e_{n-1}(T, d_2)\,.$$

We define $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$, and for $n \ge 2$ we define $\mathcal{A}_n$ as being the partition generated by $\mathcal{B}_{n-2}\,,\ \mathcal{B}'_{n-2}\,,\ \mathcal{C}_{n-2}$ and $\mathcal{C}'_{n-2}$, where $\mathcal{B}_n$ and $\mathcal{C}_n$ are as in (2.52) and (2.53) respectively. Let us define a chaining $\pi_n(t)$ associated as usual to the sequence $(\mathcal{A}_n)$ of partitions. (That is we select a set T_n which meets every element of $\mathcal{A}_n$ in exactly one point, and $\pi_n(t)$ denote the element of T_n which belongs to $A_n(t)$.)

$$U = (2^n + u_1)d_1(\pi_n(t), \pi_{n-1}(t)) + (2^{n/2} + u_2)d_2(\pi_n(t), \pi_{n-1}(t))\,,$$

so that (2.50) implies somewhat crudely that

$$\mathsf{P}(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \ge U) \le 2\exp(-2^n - \min(u_2^2, u_1))\,.$$

For $n \ge 3$ we have $\pi_n(t), \pi_{n-1}(t) \in B_{n-3}(t)$, so that $d_1(\pi_n(t), \pi_{n-1}(t)) \le \Delta_1(B_{n-3}(t))$, and $\pi_n(t), \pi_{n-1}(t) \in B'_{n-3}(t)$ so that, using (2.65) in the last inequality,

$$d_1(\pi_n(t), \pi_{n-1}(t)) \le \Delta_1(B'_{n-3}(t)) \le 2e_{n-3}(T, d_1)\,.$$

Proceeding in the same fashion for d_2 it follows that with probability at least $1 - L\exp(-\min(u_2^2, u_1))$ we have

$$\begin{aligned}\forall n \ge 3\,,\ \forall t \in T\,,\ |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \le\ & 2^n \Delta_1(B_{n-3}(t)) + 2^{n/2}\Delta_2(C_{n-3}(t)) \\ & + 2u_1 e_{n-3}(T, d_1) + 2u_2 e_{n-3}(T, d_2)\,.\end{aligned}$$

This inequality remains true for $n = 1, 2$ if in the right-hand side one replaces $n - 3$ by 0, and chaining (i.e. use of (2.25)) completes the proof. □

2.3 Functionals

To make Theorem 2.2.18 useful, we must be able to construct good admissible sequences. In this section we explain our basic method. This method, and its variations, are at the core of the book.

Let us recall that we have defined $\gamma_a(T, d)$ as

$$\gamma_\alpha(T, d) = \inf \sup_{t \in T} \sum_{n \ge 0} 2^{n/\alpha} \Delta(A_n(t))$$

where the infimum is taken over all admissible sequences $(\mathcal{A}_n)$ of partitions of T. Let us now define the quantity

$$\gamma^*_\alpha(T, d) = \inf \sup_{t \in T} \sum_{n \ge 0} 2^{n/\alpha} d(t, T_n)\,,$$

where the infimum is over all choices of the sets T_n with $\operatorname{card} T_n \le N_n$.

It is rather obvious that $\gamma^*_\alpha(T,d) \le \gamma_\alpha(T,d)$. To prove this, consider an admissible sequence $(\mathcal{A}_n)$ of partitions of T. Choose T_n such that each set of $\mathcal{A}_n$ contains one element of T_n. Then for each $t \in T$

$$\sum_{n\ge 0} 2^{n/\alpha} d(t,T_n) \le \sum_{n\ge 0} 2^{n/\alpha} \Delta(A_n(t)) ,$$

and this proves the claim. This is simply the argument of Theorem 2.2.18. Now, we would like to go the other way, that is to prove

$$\gamma_\alpha(T,d) \le K(\alpha)\gamma^*_\alpha(T,d) . \tag{2.66}$$

This is achieved by the following result.

Theorem 2.3.1. *Consider a metric space (T,d), an integer $\tau' \ge 0$ and for $n \ge 0$, consider subsets T_n of T with* $\operatorname{card} T_0 = 1$ *and* $\operatorname{card} T_n \le N_{n+\tau'} = 2^{2^{n+\tau'}}$ *for $n \ge 1$. Consider numbers $\alpha > 0$, $S > 0$, and let*

$$U = \Big\{ t \in T\,;\, \sum_{n\ge 0} 2^{n/\alpha} d(t,T_n) \le S \Big\} .$$

Then $\gamma_\alpha(U,d) \le K(\alpha,\tau')S$.

Of course here $K(\alpha,\tau')$ denotes a number depending on α and τ' only. When $U = T$ and $\tau' = 0$, this proves (2.66), and shows that the bound (2.43) is as good as the bound (2.32), if one does not mind the possible loss of a constant factor. The superiority of the bound (2.43) is that it uses admissible sequences, and as explained before Theorem 2.2.23 these are very convenient.

It is also good to observe that Theorem 2.3.1 allows us to control $\gamma_\alpha(U,d)$ using sets T_n that need not be subsets of U.

It seems appropriate to state the following obvious consequence of (2.66).

Corollary 2.3.2. *For any metric space (T,d) we have*

$$\gamma_\alpha(T,d) \le K(\alpha) \sum_{n\ge 0} 2^{n/\alpha} e_n(T) .$$

Exercise 2.3.3. Find a simple direct proof of Corollary 2.3.2. (Hint. You do have to construct the partitions. If this is too difficult, try first to read the proof of Theorem 2.3.1, and simplify it suitably.)

Exercise 2.3.4. Use (2.44) and Exercise 2.2.14 (d) to prove that if $T \subset \mathbb{R}^m$ then

$$\sum_{n\ge 0} 2^{n/2} e_n(T) \le L \log(m+1)\gamma_2(T,d) . \tag{2.67}$$

In words, Dudley's bound is never off by more than a factor about $\log(m+1)$ in $\mathbb{R}^m$.

The following simple observation allows one to construct a sequence which is admissible from one which is slightly too large. It will be used a great many times.

Lemma 2.3.5. *Consider* $\alpha > 0$, *an integer* $\tau \geq 0$ *and an increasing sequence of partitions* $(\mathcal{B}_n)_{n\geq 0}$ *with* $\operatorname{card}\mathcal{B}_n \leq N_{n+\tau}$. *Let*

$$S := \sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} \Delta(B_n(t)) \ .$$

Then we can find an admissible sequence $(\mathcal{A}_n)_{n\geq 0}$ *such that*

$$\sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} \Delta(A_n(t)) \leq 2^{\tau/\alpha}(S + K(\alpha)\Delta(T)) \ . \tag{2.68}$$

Of course (for the last time) here $K(\alpha)$ denotes a number depending on α only (that need not be the same at each occurrence).

Proof. We set $\mathcal{A}_n = \{T\}$ if $n < \tau$ and $\mathcal{A}_n = \mathcal{B}_{n-\tau}$ if $n \geq \tau$ so that $\operatorname{card}\mathcal{A}_n \leq N_n$ and

$$\sum_{n\geq\tau} 2^{n/\alpha} \Delta(A_n(t)) = 2^{\tau/\alpha} \sum_{n\geq 0} 2^{n/\alpha} \Delta(B_n(t)) \ .$$

Using the bound $\Delta(A_n(t)) \leq \Delta(T)$, we obtain

$$\sum_{n\leq\tau} 2^{n/\alpha} \Delta(A_n(t)) \leq K(\alpha) 2^{\tau/\alpha} \Delta(T) \ . \qquad \square$$

Exercise 2.3.6. Prove that (2.68) might fail if one replaces the right-hand side by $K(\alpha, \tau)S$. (Hint: S does not control $\Delta(T)$.)

Proof of Theorem 2.3.1. There is no other way than to roll up our sleeves and actually construct a partition. For $u \in T_n$, let

$$V(u) = \{t \in U \,;\, d(t, T_n) = d(t, u)\} \ .$$

(This is a well known construction, the sets $V(u)$ are simply the closures of the Voronoï cells associated to the points of T_n.) The sets $V(u)$ cover U i.e. $U = \bigcup_{u\in T_n} V(u)$, but they are not disjoint. First find a partition $\mathcal{C}_n$ of U, with $\operatorname{card}\mathcal{C}_n \leq N_{n+\tau'}$, and the property that

$$\forall C \in \mathcal{C}_n \,,\, \exists u \in T_n \,,\, C \subset V(u) \ .$$

This cannot be the partitions we are looking for since the sequence $(\mathcal{C}_n)$ need not be increasing. A more serious problem is that for $t \in V(u)$ it might happen that $d(t, T_n) \ll \Delta(V(u))$, and hence that $\Delta(C_n(t)) \gg d(t, T_n)$, in which case we have no control over $\Delta(C_n(t))$. To alleviate this problem, we will suitably break the sets of $\mathcal{C}_n$ into smaller pieces. Consider C as above, let b be the smallest integer $b > 1/\alpha + 1$, and consider the set

$$C_{bn} = \{t \in C\,;\, d(t,u) \le 2^{-bn}\Delta(U)\}\ ,$$

so that $\Delta(C_{bn}) \le 2^{-bn+1}\Delta(U)$. Similarly, consider, for $0 \le k < bn$, the set

$$C_k = \{t \in C\,;\, 2^{-k-1}\Delta(U) < d(t,u) \le 2^{-k}\Delta(U)\}\ .$$

Thus $\Delta(C_k) \le 2^{-k+1}\Delta(U)$, and, when $k < bn$, and since $C \subset V(u)$,

$$\forall t \in C_k\ ,\ \Delta(C_k) \le 2^{-k+1}\Delta(U) \le 4d(t,u) \le 4d(t,T_n)\ .$$

Therefore,

$$\forall k \le bn\,,\ \forall t \in C_k\ ,\ \Delta(C_k) \le 4d(t,T_n) + 2^{-bn+1}\Delta(U)\ . \tag{2.69}$$

Consider the partition $\mathcal{B}_n$ consisting of the sets C_k for $C \in \mathcal{C}_n\ ,\ 0 \le k \le bn$, so that $\operatorname{card}\mathcal{B}_n \le (bn+1)N_{n+\tau'}$. Consider the partition $\mathcal{A}_n$ generated by $\mathcal{B}_0\ ,\ \dots\ ,\ \mathcal{B}_n$, so that the sequence $(\mathcal{A}_n)$ increases, and $\operatorname{card}\mathcal{A}_n \le N_{n+\tau}$, where τ depends on α and τ' only. (The reader is advised to work out this fact in complete detail.) From (2.69) we get

$$\forall A \in \mathcal{A}_n\ ,\ \forall t \in A\,,\ \Delta(A) \le 4d(t,T_n) + 2^{-bn+1}\Delta(U)\,,$$

and thus

$$\begin{aligned}\sum_{n\ge 0} 2^{n/\alpha}\Delta(A_n(t)) &\le 4\sum_{n\ge 0} 2^{n/\alpha} d(t,T_n) + \Delta(U)\sum_{n\ge 0} 2^{n/\alpha - bn+1}\\ &\le 4(S + \Delta(U))\ .\end{aligned}$$

Since for t in U we have $d(t,T_0) \le S$ where T_0 contains a unique point, we have $\Delta(U) \le 2S$, and the conclusion follows from Lemma 2.3.5. □

Exercise 2.3.7. In Theorem 2.3.1, carry out the correct dependence of $K(\alpha,\tau')$ upon τ'.

Let us now explain the crucial idea of functionals (and the reason behind the name). We will say that a map F is a *functional* on a set T if, to each subset H of T it associates a number $F(H) \ge 0$, and if it is increasing, i.e.

$$H \subset H' \subset T \Rightarrow F(H) \le F(H')\,. \tag{2.70}$$

Intuitively a functional is a measure of “size” for the subsets of T. It allows to identify which subsets of T are “large” for our purposes. Suitable partitions of T will then be constructed through an exhaustion procedure that selects first the large subsets of T.

When reading the words “measure of the size of a subset of T” the reader might form the picture of the functional $F(H) = \mu(H)$ where μ is a measure on T. For our purposes, this picture is incorrect, because our goal is to understand in a sense what are the smallest functionals which satisfy a certain property to be explained below, and these do not look at all like the measure μ example. A first fundamental example of a functional is

$$F(H) = \gamma_2(H,d)\ . \tag{2.71}$$

A second, equally important, is the quantity

$$F(H) = \mathsf{E}\sup_{t\in H} X_t$$

where $(X_t)_{t\in T}$ is a process indexed by T.

Now we wish to explain the basic property needed for a functional. That this property is relevant is by no means intuitively obvious yet (but we shall soon see that the functional (2.71) does enjoy this property). Let us first try it in words: if we consider a set that is the union of many small pieces far enough from each other, then this set is significantly larger (as measured by the functional) than the *smallest* of its pieces. "Significantly larger" depends on the scale of the pieces, and on their number. This is a kind of "growth condition".

First, let us explain what we mean by "small pieces far from each other". There is a scale, say $a > 0$ at which this happens. The pieces are small at that scale: they are contained in balls with radius $a/100$. The balls are far from each other: any two centers of such balls are at mutual distance $\geq a$. Wouldn't you say that such pieces are "well separated"? Of course there is nothing specific about the choice of the radius $a/100$, and sometimes the radius has to be smaller, so we introduce a parameter $r \geq 4$, and we ask that the "small pieces" be contained in balls with radius a/r rather than $a/100$. The reason why we require $r \geq 4$ is that we want the following: two points taken in different balls with radius a/r whose centers are at distance $\geq a$ cannot be too close to each other. This would not be true for $r = 2$, so we give ourselves some room, and take $r \geq 4$. Here is the formal definition.

Definition 2.3.8. *Given $a > 0$ and an integer $r \geq 4$ we say that subsets $H_1, \ldots, H_m$ of T are (a,r)-separated if*

$$\forall \ell \leq m\,,\ H_\ell \subset B(t_\ell, a/r)\ , \tag{2.72}$$

where the points $t_1, t_2, \ldots, t_m$ in T satisfy

$$\forall \ell \leq m\,,\ t_\ell \in B(s, ar)\ ;\ \forall \ell\,,\ \ell' \leq m\,,\ \ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a \tag{2.73}$$

for a certain point $s \in T$.

Of course here $B(s,a)$ denotes the closed ball with center s and radius a in the metric space (T,d). A secondary feature of this definition is that the small pieces H_ℓ are not only well separated (on a scale a), but they are in the "same region of T" (on the larger scale ra). This is the content of the first part of condition (2.73):

$$\forall \ell \leq m\,,\ t_\ell \in B(s, ar)\ .$$

Exercise 2.3.9. Find interesting examples of metric spaces for which there are no points $t_1, \dots, t_m$ as in (2.73), for all values of n (respectively all large enough values of n).

Now, what do we mean by "the union of the pieces is significantly larger than the *smallest* of these pieces"? In this first version of the growth condition, this means that the size of this union is larger than the size of the smallest piece by a quantity $a\sqrt{\log N}$ where N is the number of pieces. (We remind the reader that the function $\sqrt{\log y}$ arises from the fact that this it in a sense the inverse of the function $\exp(-x^2)$.) Well, sometimes it will only be larger by a quantity of say $a\sqrt{\log N}/100$. This is how the parameter c^* below comes into the picture. Of course, one could also multiply the functionals by a suitable constant (i.e. $1/c^*$) to always reduce to the case $c^* = 1$ but this is a matter of taste.

Another feature is that we do not need to consider the case with N pieces for a general value of N, but only for the case where $N = N_n$ for some n. This is simply because we care about the value of $\log N$ only within, say, a factor of 2, and this is precisely what motived the definition of N_n. In order to understand the definition below one should also recall that $\sqrt{\log N_n}$ is about $2^{n/2}$.

It will be rather convenient to consider not only a single functional but a whole sequence (F_n) of functionals, but at first reading one might assume that F_n does not depend on n. So, consider a metric space (T, d) (that need *not* be finite), and a decreasing sequence $(F_n)_{n \ge 0}$ of functionals on T, that is

$$\forall H \subset T \,,\; F_{n+1}(H) \le F_n(H) \,. \tag{2.74}$$

Definition 2.3.10. *We say that the functionals F_n satisfy the* growth condition *with parameters $r \ge 4$ and $c^* > 0$ if for any integer $n \ge 0$ and any $a > 0$ the following holds true, where $m = N_{n+1}$. For each collection of subsets $H_1, \dots, H_m$ of T that are (a, r)-separated we have*

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge c^* a 2^{n/2} + \min_{\ell \le m} F_{n+1}(H_\ell) \,. \tag{2.75}$$

We observe that the functional F_n occurs on the left-hand side of (2.75), while the smaller functional F_{n+1} occurs on the right-hand side (which gives us a little extra room to check this condition).

Exercise 2.3.11. Find example of spaces (T, d) where the growth condition holds while $F_n(H) = 0$ for each n and each $H \subset T$. (Hint: use Exercise 2.3.9.)

We now note the non-obvious fact that condition (2.75) imposes strong restrictions on the metric space (T, d), and we explain this now. We prove that (2.75) implies that if $a > 2^{-n/2} F_0(T)/c^*$, each ball $B(s, ar)$ can be covered by N_{n+1} balls $B(t, a)$. Consider points $t_1, \dots, t_k$ in $B(s, ar)$ such that $d(t_\ell, t_{\ell'}) \ge a$ whenever $\ell \ne \ell'$. Assume that $k \ge m = N_{n+1}$ for a certain $n \ge 0$.

Taking $H_\ell = \{t_\ell\}$, and since $F_{n+1} \geq 0$, the separation condition implies $F_0(T) \geq F_n(T) \geq c^* a 2^{n/2}$. Consequently, if we assume that $c^* a 2^{n/2} > F_0(T)$, we must have $k < N_{n+1}$. If k is as large as possible, the ball $B(s, ar)$ is covered by the balls $B(t_\ell, a)$ for $\ell \leq k$, proving the claim.

Exercise 2.3.12. If $r \geq 4$, prove that the preceding property yields the inequality $2^{n/2} e_n(T) \leq K(r) F_0(T)/c^* + L\Delta(T)$. (Hint: Iterate the process of covering a ball with radius ar by balls with radius a to bound the minimum number $N(T, d, \epsilon)$ of balls with radius ϵ needed to cover T and use Exercise 2.2.13.) Explain why the term $L\Delta(T)$ is necessary. (Hint: use Exercise 2.3.11.)

The following illustrates how we shall use the first part of (2.73).

Exercise 2.3.13. Let (T, d) be isometric to a subset of $\mathbb{R}^k$ provided with the distance induced by a norm. Prove that in order to check that a sequence of functionals satisfies the growth condition of Definition 2.3.10, it suffices to consider the values of n for which $N_{n+1} \leq (1 + 2/r)^k$. (Hint: it follows from (2.41) that for larger values of n there are no points $t_1, \ldots, t_m$ as in (2.73).)

As we shall soon see, the existence of a sequence of functionals satisfying the separation property will give us a lot more information than the crude result of Exercise 2.3.12.

Before we come to this, what is the point of considering such sequences of functionals? As the following result shows, decreasing sequences of functionals satisfying the growth condition of Definition 2.3.10 are "built into" the definition of $\gamma_2(T, d)$.

Theorem 2.3.14. *For any metric space (T, d) there exists a decreasing sequence of functionals $(F_n)_{n \geq 0}$ with $F_0(T) = \gamma_2(T, d)$ which satisfies the growth condition of Definition 2.3.10 for $r = 4$ and $c^* = 1/2$.*

In words, one can find a decreasing sequence of functionals satisfying the growth condition with $F_0(T)$ as large as $\gamma_2(T, d)$.

Proof. This proof provides a good opportunity to understand the typical way a sequence (F_n) of functionals might depend on n. For a subset H of T we define

$$F_n(H) = \inf \sup_{t \in H} \sum_{k \geq n} 2^{k/2} \Delta(A_k(t)) ,$$

where the infimum is taken over all admissible sequences $(\mathcal{A}_n)$ of partitions of H. (The dependence on n is that the summation starts at $k = n$. This feature will often occur.) Thus $F_0(T) = \gamma_2(T, d)$. To prove the growth condition of Definition 2.3.10, consider $m = N_{n+1}$ and consider points $(t_\ell)_{\ell \leq m}$ of T, with $d(t_\ell, t_{\ell'}) \geq a$ if $\ell \neq \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/4)$, and the set $H = \bigcup_{\ell \leq m} H_\ell$. Consider an admissible sequence $(\mathcal{A}_n)$ of H, and

$$I = \{\ell \leq m \, ; \, \exists A \in \mathcal{A}_n \, , \, A \subset H_\ell\} \, .$$

Since the sets $(H_\ell)_{\ell \le m}$ are disjoint, we have $\operatorname{card} I \le N_n$, and thus there exists $\ell \le m$ with $\ell \notin I$. Then for $t \in H_\ell$, we have $A_n(t) \not\subset H_\ell$, so that since $A_n(t) \subset H$, the set $A_n(t)$ must meet a set $H_{\ell'}$ for a certain $\ell' \ne \ell$, and consequently it meets the ball $B(t_{\ell'}, a/4)$. Since $d(t, B(t_{\ell'}, a/4)) \ge a/2$, this implies that $\Delta(A_n(t)) \ge a/2$. Therefore

$$\sum_{k \ge n} 2^{k/2} \Delta(A_k(t)) \ge \frac{1}{2} a 2^{n/2} + \sum_{k \ge n+1} 2^{k/2} \Delta(A_k(t) \cap H_\ell) . \tag{2.76}$$

Since $\mathcal{A}'_n = \{A \cap H_\ell ; A \in \mathcal{A}_n\}$ is an admissible sequence of H_ℓ, we have by definition

$$\sup_{t \in H_\ell} \sum_{k \ge n+1} 2^{k/2} \Delta(A_k(t) \cap H_\ell) \ge F_{n+1}(H_\ell) .$$

Hence, taking the supremum over t in H_ℓ in (2.76) we get

$$\sup_{t \in H_\ell} \sum_{k \ge n} 2^{k/2} \Delta(A_k(t)) \ge \frac{1}{2} a 2^{n/2} + F_{n+1}(H_\ell) .$$

Since the admissible sequence $(\mathcal{A}_n)$ is arbitrary, we have shown that

$$F_n(H) \ge \frac{1}{2} a 2^{n/2} + \min_\ell F_{n+1}(H_\ell) ,$$

which is (2.75) for $c^* = 1/2$. □

The previous proof demonstrates how to use functionals F_n which actually depend on n. This will be a very useful technical device. However it is not really needed here, since we also have the following.

Theorem 2.3.15. *The functionals $F_n(H) = \gamma_2(H, d)$ satisfy the growth condition of Definition 2.3.10 for $r = 8$ and $c^* = 1/4$.*

Proof. The proof is almost the same as that of Theorem 2.3.14. Consider points $(t_\ell)_{\ell \le m}$ of T, with $d(t_\ell, t_{\ell'}) \ge a$ if $\ell \ne \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/8)$, and the set $H = \bigcup_{\ell \le m} H_\ell$. Consider an admissible sequence $(\mathcal{A}_n)$ of H, and

$$I = \{\ell \le m \, ; \, \exists A \in \mathcal{A}_n \, , \, A \subset H_\ell\} .$$

Since the sets $(H_\ell)_{\ell \le m}$ are disjoint, we have $\operatorname{card} I \le N_n$, and thus there exists $\ell \le m$ with $\ell \notin I$. Then for $t \in H_\ell$, we have $A_n(t) \not\subset H_\ell$, so that since $A_n(t) \subset H$, the set $A_n(t)$ must meet a set $H_{\ell'}$ for a certain $\ell' \ne \ell$, and consequently it meets the ball $B(t_{\ell'}, a/8)$. Since $d(t, B(t_{\ell'}, a/8)) \ge a/2$, this implies that $\Delta(A_n(t)) \ge a/2$. Therefore, since $\Delta(A_n(t) \cap H_\ell) \le \Delta(H_\ell) \le a/4$,

$$\sum_{k \ge 0} 2^{k/2} \Delta(A_k(t)) \ge \frac{1}{4} a 2^{n/2} + \sum_{k \ge 0} 2^{k/2} \Delta(A_k(t) \cap H_\ell) . \tag{2.77}$$

From this point on the proof is identical to that of Theorem 2.3.14. □

Our next result is fundamental. It is a kind of converse to Theorem 2.3.14. The size of $\gamma_2(T,d)$ cannot be really larger than $F_0(T)$ when the sequence (F_n) of functionals satisfies the growth condition of Definition 2.3.10. Put it another way, it says that in a sense $F(H) = \gamma_2(H,d)$ is the smallest functional which satisfies the growth condition of Definition 2.3.10. This also explains why we did not give any very simple example of functional satisfying the growth condition. This seems to be the simplest example.

Theorem 2.3.16. *Let (T,d) be a metric space. Assume that there exists a decreasing sequence of functionals $(F_n)_{n\geq 0}$ which satisfies the growth condition of Definition 2.3.10. Then*

$$\gamma_2(T,d) \leq \frac{Lr}{c^*} F_0(T) + Lr\Delta(T)\,. \tag{2.78}$$

This theorem and its generalizations form the backbone of this book. The essence of this theorem is that it produces (by actually constructing them) a sequence of partitions that witnesses the inequality (2.78). For this reason, it could be called "the fundamental partitioning theorem." The proof of Theorem 2.3.16 is not really difficult, but since one has to construct the partitions, it does require again to roll up our sleeves and even get a bit of grease on our hands. Thus this proof will be better presented (in Section 2.6) after the power of this principle has been demonstrated in Section 2.4 and the usefulness of its consequences illustrated again in Section 2.5.

Exercise 2.3.17. Consider a metric space T consisting of exactly two points. Prove that the sequence of functionals given by $F_n(H) = 0$ for each $H \subset T$ satisfies the growth condition of Definition 2.3.10 for $r = 4$ and any $c^* > 0$. Explain why we cannot replace (2.78) by the inequality $\gamma_2(T,d) \leq LrF_0(T)/c^*$.

Given the functionals F_n, Theorem 2.3.16 yields partitions, but it does not say how to find these functionals. One must understand that there is no magic. Admissible sequences are not going to come out of thin air, but rather they will reflect the geometry of the space (T,d). Once this geometry is understood, it is usually possible to guess a good choice for the functionals F_n. Many examples will be given in subsequent chapters. It seems, at least to the author, that it is much easier to guess the functionals F_n rather than the partitions that witness the inequality (2.78). Besides, as Theorem 2.3.14 shows, we really have no choice. Functionals with the growth property are intimately connected with admissible sequences of partitions.

2.4 Gaussian Processes and the Mysteries of Hilbert Space

Consider a Gaussian process $(X_t)_{t\in T}$, that is, a jointly Gaussian family of centered r.v.s indexed by T. We provide T with the canonical distance

$$d(s,t) = \left(\mathsf{E}(X_s - X_t)^2\right)^{1/2} . \tag{2.79}$$

Recall the functional γ_2 of Definition 2.2.19.

Theorem 2.4.1 (The Majorizing Measure Theorem). *For some universal constant L we have*

$$\frac{1}{L}\gamma_2(T,d) \le \mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T,d) . \tag{2.80}$$

The reason for the name is explained in Section 6.2. We can reformulate this theorem by the statement

$$\textbf{Chaining suffices to explain the size of a Gaussian process}. \tag{2.81}$$

By this statement we simply means that (as witnessed by the left-hand side inequality in (2.80)) the "natural" chaining bound for the size of a Gaussian process (as witnessed by the right-hand side inequality in (2.80)) is of correct order, provided of course one uses the best possible chaining. The author believes that this is an occurrence of a much more general phenomenon, several aspects of which will be investigated in later chapters.

The right-hand side inequality in (2.80) follows from Theorem 2.2.22. To prove the lower bound we will use Theorem 2.3.16 and the functionals

$$F_n(H) = F(H) = \sup_{H^*\subset H, H^*\,\text{finite}} \mathsf{E}\sup_{t\in H^*} X_t ,$$

so that F_n does not depend on n. To apply (2.78) we need to prove that the functionals F_n satisfy the growth condition with c^* a universal constant and to bound $\Delta(T)$ (which is easy). We strive to give a proof that relies on general principles, and lends itself to generalizations.

Lemma 2.4.2 (Sudakov minoration). *Assume that*

$$\forall p\,,\, q \le m\,, \quad p \ne q \quad \Rightarrow d(t_p, t_q) \ge a\,.$$

Then we have

$$\mathsf{E}\sup_{p\le m} X_{t_p} \ge \frac{a}{L_1}\sqrt{\log m}\,. \tag{2.82}$$

Here and below $L_1, L_2, \ldots$ are specific universal constants. Their values remain the same (at least within the same section).

Exercise 2.4.3. Prove that Lemma 2.4.2 is equivalent to the following statement. If $(X_t)_{t\in T}$ is a Gaussian process, and d is the canonical distance, then

$$e_n(T,d) \le 2^{-n/2}\mathsf{E}\sup_{t\in T} X_t\,. \tag{2.83}$$

Compare with Exercise 2.3.12.

A proof of Sudakov minoration can be found in [5], p. 83. The same proof is actually given further in the present book, and the ambitious reader may like to try to understand this now, using the following steps.

Exercise 2.4.4. Use Lemma 8.3.6 and Lemma 16.8.10 to prove that for a Gaussian process $(X_t)_{t\in T}$ we have $e_n(T,d) \le 2^{-n/2}\mathsf{E}\sup_{t\in T}|X_t|$. Then use Exercise 2.2.2 to deduce (2.83).

To understand the relevance of Sudakov minoration, let us consider the case where $\mathsf{E}X_{t_p}^2 \le 100a^2$ (say) for each p. Then (2.82) means that the bound (2.13) is of the correct order in this situation.

Exercise 2.4.5. Prove (2.82) when the r.v.s X_{t_p} are independent. (Hint: use Exercise 2.2.7 (b).)

Exercise 2.4.6. A natural approach ("the second moment method") to prove that $\mathsf{P}(\sup_{p\le m} X_{t_p} \ge u)$ is at least $1/L$ for a certain value of u is as follows. Consider the r.v. $Y = \sum_p \mathbf{1}_{\{X_{t_p}\ge u\}}$, prove that $\mathsf{E}Y^2 \le L(\mathsf{E}Y)^2$, and then use the Paley-Zygmund inequality (7.30) below to prove that $\sup_{p\le m} X_{t_p} \ge a\sqrt{\log m}/L_1$ with probability $\ge 1/L$. Prove that this approach works when the r.v.s X_{t_ℓ} are independent, but find examples showing that this naive approach does not work in general to prove (2.82).

The following is a very important property of Gaussian processes, and one of the keys to Theorem 2.4.1. It is a facet of the theory of concentration of measure, a leading idea of modern probability theory. The reader is referred to the (very nice) book of M. Ledoux [4] to learn about this.

Lemma 2.4.7. *Consider a Gaussian process $(X_t)_{t\in U}$, where U is finite and let $\sigma = \sup_{t\in U}(\mathsf{E}X_t^2)^{1/2}$. Then for $u > 0$ we have*

$$\mathsf{P}\Big(\Big|\sup_{t\in U} X_t - \mathsf{E}\sup_{t\in U} X_t\Big| \ge u\Big) \le 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big)\,. \tag{2.84}$$

Let us stress in words what this means. The size of the fluctuations of $\mathsf{E}\sup_{t\in U} X_t$ is governed by the size of the individual r.v.s X_t, rather than by the (typically much larger) quantity $\mathsf{E}\sup_{t\in U} X_t$.

Exercise 2.4.8. Find an example of a Gaussian process for which

$$\mathsf{E}\sup_{t\in T} X_t \gg \sigma = \sup_{t\in T}(\mathsf{E}X_t^2)^{1/2}\,,$$

whereas the fluctuations of $\sup_{t\in T} X_t$ are of order σ, e.g. the variance of $\sup_t X_t$ is about σ^2. (Hint: $T = \{(t_i)_{i\le n}; \sum_{i\le n} t_i^2 \le 1\}$ and $X_t = \sum_{i\le n} t_i g_i$ where g_i are independent standard Gaussian. Observe first that $(\sup_t X_t)^2 = \sum_{i\le n} g_i^2$ is of order n and has fluctuations of order $\sqrt{n}$ by the central limit theorem. Conclude that $\sup_t X_t$ has fluctuations of order 1 whatever the value of n.)

Proposition 2.4.9. *Consider points $(t_\ell)_{\ell\le m}$ of T. Assume that $d(t_\ell, t_{\ell'}) \ge a$ if $\ell \ne \ell'$. Consider $\sigma > 0$, and for $\ell \le m$ a finite set $H_\ell \subset B(t_\ell, \sigma)$. Then if $H = \bigcup_{\ell\le m} H_\ell$ we have*

$$\mathsf{E}\sup_{t\in H} X_t \ge \frac{a}{L_1}\sqrt{\log m} - L_2\sigma\sqrt{\log m} + \min_{\ell\le m} \mathsf{E}\sup_{t\in H_\ell} X_t \ . \tag{2.85}$$

When $\sigma \le a/(2L_1L_2)$, (2.85) implies

$$\mathsf{E}\sup_{t\in H} X_t \ge \frac{a}{2L_1}\sqrt{\log m} + \min_{\ell\le m} \mathsf{E}\sup_{t\in H_\ell} X_t \ , \tag{2.86}$$

which can be seen as a generalization of (2.82).

Proof. We can and do assume $m \ge 2$. For $\ell \le m$, we consider the r.v.

$$Y_\ell = \big(\sup_{t\in H_\ell} X_t\big) - X_{t_\ell} = \sup_{t\in H_\ell}(X_t - X_{t_\ell}) \ .$$

We set $U = H_\ell$ and for $t \in U$ we set $Z_t = X_t - X_{t_\ell}$. Since $H_\ell \subset B(t_\ell, \sigma)$ we have $\mathsf{E}Z_t^2 = d(t, t_\ell)^2 \le \sigma^2$ and, for $u \ge 0$ equation (2.84) used for the process $(Z_t)_{t\in U}$ implies

$$\mathsf{P}(|Y_\ell - \mathsf{E}Y_\ell| \ge u) \le 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big) \ .$$

Thus if $V = \max_{\ell\le m} |Y_\ell - \mathsf{E}Y_\ell|$ then

$$\mathsf{P}(V \ge u) \le 2m\exp\Big(-\frac{u^2}{2\sigma^2}\Big) , \tag{2.87}$$

and (2.11) implies $\mathsf{E}V \le L_2\sigma\sqrt{\log m}$. Now, for each $\ell \le m$,

$$Y_\ell \ge \mathsf{E}Y_\ell - V \ge \min_{\ell\le m} \mathsf{E}Y_\ell - V \ ,$$

and thus

$$\sup_{t\in H_\ell} X_t = Y_\ell + X_{t_\ell} \ge X_{t_\ell} + \min_{\ell\le m} \mathsf{E}Y_\ell - V$$

so that

$$\sup_{t\in H} X_t \ge \max_{\ell\le m} X_{t_\ell} + \min_{\ell\le m} \mathsf{E}Y_\ell - V \ .$$

We then take expectations and use (2.82). □

Exercise 2.4.10. Prove that (2.86) might fail if one allows $\sigma = a$. (Hint: the intersection of the balls $B(t_\ell, a)$ might contain a ball with positive radius.)

Exercise 2.4.11. Prove that

$$\mathsf{E}\sup_{t\in H} X_t \le La\sqrt{\log m} + \max_{\ell\le m} \mathsf{E}\sup_{t\in H_\ell} X_t \ . \tag{2.88}$$

Try to find improvements on this bound. (Hint: peek at (16.81) below.)

Proof of Theorem 2.4.1. We fix $r \geq 2L_1L_2$. To prove the growth condition for the functionals F_n we simply observe that (2.86) implies that (2.75) holds for $c^* = 1/L$. Using Theorem 2.3.16, it remains only to control the term $\Delta(T)$. But

$$\mathsf{E}\max(X_{t_1}, X_{t_2}) = \mathsf{E}\max(X_{t_1} - X_{t_2}, 0) = \frac{1}{\sqrt{2\pi}} d(t_1, t_2)\,,$$

so that $\Delta(T) \leq \sqrt{2\pi}\mathsf{E}\sup_{t\in T} X_t$. □

The proof of Theorem 2.4.1 displays an interesting feature. This theorem aims at understanding $\mathsf{E}\sup_{t\in T} X_t$, and for this we use functionals that are based on precisely this quantity. This is not a circular argument. The content of Theorem 2.4.1 is that there is simply no other way to bound a Gaussian process than to control the quantity $\gamma_2(T,d)$. Of course, to control this quantity in a specific situation, we must in some way gain understanding of the underlying geometry of this situation.

The following is a noteworthy consequence of Theorem 2.4.1.

Theorem 2.4.12. *Consider two processes $(Y_t)_{t\in T}$ and $(X_t)_{t\in T}$ indexed by the same set. Assume that the process $(X_t)_{t\in T}$ is Gaussian and that the process $(Y_t)_{t\in T}$ satisfies the condition*

$$\forall u > 0\ ,\ \forall s\ ,\ t \in T\ ,\ \mathsf{P}(|Y_s - Y_t| \geq u) \leq 2\exp\left(-\frac{u^2}{d(s,t)^2}\right)\ ,$$

where d is the distance (2.79) associated to the process X_t. Then we have

$$\mathsf{E}\sup_{s,t\in T} |Y_s - Y_t| \leq L\mathsf{E}\sup_{t\in T} X_t\ .$$

Proof. We combine (2.49) with the left-hand side of (2.80). □

Let us now turn to a simple (and classical) example that illustrates well the difference between Dudley's bound (2.38) and the bound (2.32). Basically this example reproduces, for a metric space associated to an actual Gaussian process, the metric structure that was described in an abstract setting in Exercise 2.2.16. Consider an independent sequence $(g_i)_{i\geq 1}$ of standard Gaussian r.v.s and for $i \geq 2$ set

$$X_i = \frac{g_i}{\sqrt{\log i}}\ . \tag{2.89}$$

Consider an integer $s \geq 3$ and the process $(X_i)_{2\leq i\leq N_s}$ so the index set is $T = \{2, 3, \ldots, N_s\}$. The distance d associated to the process satisfies for $p \neq q$

$$\frac{1}{\sqrt{\log(\min(p,q))}} \leq d(p,q) \leq \frac{2}{\sqrt{\log(\min(p,q))}}\ . \tag{2.90}$$

Consider $1 \leq n \leq s-2$ and $T_n \subset T$ with $\operatorname{card} T_n = N_n$. There exists $p \leq N_n + 1$ with $p \notin T_n$, so that (2.90) implies $d(p, T_n) \geq 2^{-n/2}/L$ (where the distance from a point to a set is defined in (2.33)). This proves that $e_n(T) \geq 2^{-n/2}/L$. Therefore

$$\sum_n 2^{n/2} e_n(T) \geq \frac{s-2}{L} \, . \tag{2.91}$$

On the other hand, for $n \leq s$ let us now consider $T_n = \{2, 3, \ldots, N_n, N_s\}$, integers $p \in T$ and $m \leq s-1$ such that $N_m < p \leq N_{m+1}$. Then $d(p, T_n) = 0$ if $n \geq m+1$, while, if $n \leq m$,

$$d(p, T_n) \leq d(p, N_s) \leq L 2^{-m/2}$$

by (2.90) and since $p \geq N_m$ and $N_s \geq N_m$. Hence we have

$$\sum_n 2^{n/2} d(p, T_n) \leq \sum_{n \leq m} L 2^{n/2} 2^{-m/2} \leq L \, . \tag{2.92}$$

Comparing (2.91) and (2.92) proves that the bound (2.38) is worse than the bound (2.32) by a factor about s.

Exercise 2.4.13. Prove that when T is finite, the bound (2.38) cannot be worse than (2.32) by a factor greater than about $\log\log \operatorname{card} T$. This shows that the previous example is in a sense extremal. (Hint: use $2^{n/2} e_n(T) \leq L\gamma_2(T, d)$ and $e_n(T) = 0$ if $N_n \geq \operatorname{card} T$.)

Exercise 2.4.14. Prove that the estimate (2.67) is essentially optimal. (Hint: if $m \geq \exp(10s)$, one can produce the situation of Example 2.2.16 (b) inside $\mathbb{R}^m$.)

It follows from (2.92) and (2.32) that $\mathsf{E} \sup_{i \geq 1} X_i < \infty$. A simpler proof of this fact is given in Proposition 2.4.16 below.

Now we generalize the process of Exercise 2.2.15 to Hilbert space. We consider the Hilbert space $\ell^2 = \ell^2(\mathbb{N}^*)$ of sequences $(t_i)_{i \geq 1}$ such that $\sum_{i \geq 1} t_i^2 < \infty$, provided with the norm

$$\|t\| = \|t\|_2 = \Big(\sum_{i \geq 1} t_i^2 \Big)^{1/2} \, . \tag{2.93}$$

To each t in ℓ^2 we associate a Gaussian r.v.

$$X_t = \sum_{i \geq 1} t_i g_i \tag{2.94}$$

(the series converges in $L^2(\Omega)$). In this manner, for each subset T of ℓ^2 we can consider the Gaussian process $(X_t)_{t \in T}$. The distance induced on T by the process coincides with the distance of ℓ^2 since from (2.94) we have $\mathsf{E} X_t^2 = \sum_{i \geq 1} t_i^2$.

The importance of this construction is that it is generic. *All* Gaussian processes can be obtained in this way, at least when there is a countable subset T' of T that is dense in the space (T,d), which is the only case of importance for us. Indeed, it suffices to think of the r.v. Y_t of a Gaussian process as a point in $L^2(\Omega)$, where Ω is the underlying probability space, and to identify $L^2(\Omega)$, which is then separable, and ℓ^2 by choosing an orthonormal basis of $L^2(\Omega)$.

Here is the place to make a general observation. It is not true that all processes of interest can be represented as the sum of a random series as in (2.94). Suppose, however, that one is interested in the boundedness of a random series of functions, $X_u = \sum_{i\geq 1} \xi_i f_i(u)$ for $u \in U$. Then all that matters is the set T of coefficients $T = \{t = (f_i(u))_i; u \in U\}$. For a sequence $t = (t_i)$ we then define

$$X_t = \sum_i t_i \xi_i \tag{2.95}$$

and the fundamental issue becomes to understand the boundedness of the process $(X_t)_{t\in T}$. This is why processes of the type (2.95) play such an important role in this book.

A subset T of ℓ^2 will always be provided with the distance induced by ℓ^2, so we may also write $\gamma_2(T)$ rather than $\gamma_2(T,d)$. We denote by $\operatorname{conv} T$ the convex hull of T, and we write

$$T_1 + T_2 = \big\{t_1 + t_2 \ ; \ t_1 \in T_1 \ , \ t_2 \in T_2\big\} \ .$$

Theorem 2.4.15. *For a subset T of ℓ^2, we have*

$$\gamma_2(\operatorname{conv} T) \leq L\gamma_2(T) \ . \tag{2.96}$$

For two subsets T_1 and T_2 of ℓ^2, we have

$$\gamma_2(T_1 + T_2) \leq L\big(\gamma_2(T_1) + \gamma_2(T_2)\big) \ . \tag{2.97}$$

Proof. To prove (2.96) we observe that since $X_{a_1 t_1 + a_2 t_2} = a_1 X_{t_1} + a_2 X_{t_2}$ we have

$$\sup_{t\in \operatorname{conv} T} X_t = \sup_{t\in T} X_t \ . \tag{2.98}$$

We then use (2.80) to write

$$\frac{1}{L}\gamma_2(\operatorname{conv} T) \leq \mathsf{E} \sup_{\operatorname{conv} T} X_t \leq \mathsf{E} \sup_T X_t \leq L\gamma_2(T) \ .$$

The proof of (2.97) is similar. □

We recall the ℓ^2 norm $\|\cdot\|$ of (2.93). Here is a simple fact.

Proposition 2.4.16. *Consider a set $T = \{t_k \ ; \ k \geq 1\}$ where*

$$\forall\, k \geq 1 \ , \ \|t_k\| \leq 1/\sqrt{\log(k+1)} \ .$$

Then $\mathsf{E} \sup_{t\in T} X_t \leq L$.

Proof. We have

$$\mathsf{P}\Big(\sup_{k\geq 1}|X_{t_k}|\geq u\Big)\leq\sum_{k\geq 1}\mathsf{P}(|X_{t_k}|\geq u)\leq\sum_{k\geq 1}2\exp\Big(-\frac{u^2}{2}\log(k+1)\Big) \quad (2.99)$$

since X_{t_k} is Gaussian with $\mathsf{E}X_{t_k}^2\leq 1/\log(k+1)$. Now for $u\geq 2$, the right-hand side of (2.99) is at most $L\exp(-u^2/L)$. □

Exercise 2.4.17. Deduce Proposition 2.4.16 from (2.32). (Hint: see Exercise 2.2.16 (a).)

Combining with (2.98), Proposition 2.4.16 proves that $\mathsf{E}\sup_{t\in T}X_t\leq L$, where $T=\mathrm{conv}\{t_k;k\geq 1\}$. The following shows that this situation is in a sense generic.

Theorem 2.4.18. *Consider a countable set $T\subset\ell^2$, with $0\in T$. Then we can find a sequence (t_k), such that each element t_k is a multiple of the difference of two elements of T, with*

$$\forall\, k\geq 1\ ,\ \|t_k\|\sqrt{\log(k+1)}\leq L\mathsf{E}\sup_{t\in T}X_t$$

and

$$T\subset\mathrm{conv}(\{t_k\ ;\ k\geq 1\})\ .$$

Proof. By Theorem 2.4.1 we can find an admissible sequence $(\mathcal{A}_n)$ of T with

$$\forall t\in T\,,\ \sum_{n\geq 0}2^{n/2}\Delta(A_n(t))\leq L\mathsf{E}\sup_{t\in T}X_t:=S\ . \quad (2.100)$$

We construct sets $T_n\subset T$, such that each $A\in\mathcal{A}_n$ contains exactly one element of T_n. We ensure in the construction that $T=\bigcup_{n\geq 0}T_n$ and that $T_0=\{0\}$. (To do this, we simply enumerate the elements of T as $(v_n)_{n\geq 1}$ with $v_0=0$ and we ensure that v_n is in T_n.) For $n\geq 1$ consider the set U_n that consists of all the points

$$2^{-n/2}\frac{t-v}{\|t-v\|}$$

where $t\in T_n, v\in T_{n-1}$ and $t\neq v$. Thus each element of U_n has norm $2^{-n/2}$, and U_n has at most $N_nN_{n-1}\leq N_{n+1}$ elements. Let $U=\bigcup_{k\geq 1}U_k$. We observe that U contains at most N_{n+2} elements of norm $\geq 2^{-n/2}$. If we enumerate $U=\{t_k;k=1,\ldots\}$ where the sequence $(\|t_k\|)$ is non-increasing, then if $\|t_k\|\geq 2^{-n/2}$ we have $k\leq N_{n+2}$ and this implies that $\|t_k\|\leq L/\sqrt{\log(k+1)}$.

Consider $t\in T$, so that $t\in T_m$ for some $m\geq 0$. Writing $\pi_n(t)$ for the unique element of $T_n\cap A_n(t)$, since $\pi_0(t)=0$ we have

$$t=\sum_{1\leq n\leq m}\pi_n(t)-\pi_{n-1}(t)=\sum_{1\leq n\leq m}a_n(t)u_n(t)\ , \quad (2.101)$$

where

$$u_n(t) = 2^{-n/2}\frac{\pi_n(t) - \pi_{n-1}(t)}{\|\pi_n(t) - \pi_{n-1}(t)\|} \in U \; ; \; a_n(t) = 2^{n/2}\|\pi_n(t) - \pi_{n-1}(t)\| \; .$$

Since

$$\sum_{1 \le n \le m} a_n(t) \le \sum_{n \ge 1} 2^{n/2}\Delta(A_{n-1}(t)) \le 2S$$

and since $u_n(t) \in U_n \subset U$ we see from (2.101) that

$$t \in 2S \operatorname{conv}(U \cup \{0\}) \; .$$

This concludes the proof. □

It is good to meditate a little about the significance of Theorem 2.4.18. First, we reformulate this theorem in a way which is suitable for generalizations. Consider the class $\mathcal{C}$ of sets of the type $C = \operatorname{conv}\{t_k; k \ge 1\}$ and for $C \in \mathcal{C}$ define the size $s(C)$ as $\inf \sup_k \|t_k\|\sqrt{\log(k+1)}$, where we assume without loss of generality that the sequence $(\|t_k\|)_{k\ge1}$ decreases, and where the infimum is over all possible choices of the sequence (t_k) for which $C = \operatorname{conv}\{t_k; k \ge 1\}$. Proposition 2.4.16 implies $\mathsf{E}\sup_{t\in C} X_t \le Ls(C)$. Theorem 2.4.18 implies that given a countable set T with $0 \in T$ we can find $T \subset C \in \mathcal{C}$ with $s(C) \le L\mathsf{E}\sup_{t\in T} X_t$. In words, the size of T for the Gaussian process is witnessed by the smallest size (as measured by s) of an element of $\mathcal{C}$ containing T.

Also worthy of detailing is a remarkable geometric consequence of Theorem 2.4.18. Consider an integer N and let us provide ℓ^2_N ($= \mathbb{R}^N$ provided with the Euclidean distance) with the canonical Gaussian measure μ, i.e. the law of the i.i.d. Gaussian sequence $(g_i)_{i\le N}$. Let us view an element t of ℓ^2_N as a function on ℓ^2_N by the canonical duality, so t is a r.v. Y_t on the probability space (ℓ^2_N, μ). The processes (X_t) and (Y_t) have the same law, hence they are really the same object viewed in two different ways. Consider a subset T of ℓ^2_N, and assume that $T \subset \operatorname{conv}\{t_k; k \ge 1\}$. Then for any $v > 0$ we have

$$\Big\{\sup_{t\in T} t \ge v\Big\} \subset \bigcup_{k\ge1}\{t_k \ge v\} \; . \tag{2.102}$$

The somewhat complicated set on the left-hand side is covered by a countable union of much simpler sets: the sets $\{t_k \ge v\}$ are *half-spaces*. Assume now that for $k \ge 1$ and a certain S we have $\|t_k\|\sqrt{\log(k+1)} \le S$. Then (2.99) implies that for $u \ge 2$

$$\sum_{k\ge1}\mu(\{t_k \ge Su\}) \le L\exp(-u^2/L) \; .$$

Theorem 2.4.18 implies that may take $S \le L\mathsf{E}\sup_t X_t$. Therefore for $v \ge L\mathsf{E}\sup_t X_t$, the fact that the set in the left-hand side of (2.102) is small (in the sense of probability) can be witnessed by the fact that this set can be covered by a countable union of simple sets (half-spaces) the *sum* of the probabilities of which is small.

Of course, one may hope that the two remarkable phenomena described above occur (at least in some form) in many other settings, a topic to which we shall come back many times.

Exercise 2.4.19. Prove that if $T \subset \ell^2$ and $0 \in T$, then (even when T is not countable) we can find a sequence (t_k) in ℓ^2, with $\|t_k\|\sqrt{\log(k+1)} \le L\mathsf{E}\sup_{t\in T} X_t$ for all k and

$$T \subset \overline{\mathrm{conv}}\{t_k \; ; \; k \ge 1\} ,$$

where $\overline{\mathrm{conv}}$ denotes the closed convex hull. (Hint: do the obvious thing, apply Theorem 2.4.18 to a dense countable subset of T.) Denoting now $\mathrm{conv}^*(A)$ the set of infinite sums $\sum_i \alpha_i a_i$ where $\sum_i |\alpha_i| = 1$ and $a_i \in A$, prove that one can also achieve

$$T \subset \mathrm{conv}^*\{t_k \; ; \; k \ge 1\} .$$

Exercise 2.4.20. Consider a set $T \subset \ell^2$ with $0 \in T \subset B(0,\delta)$. Prove that we can find a sequence (t_k) in ℓ^2, with the following properties:

$$\forall\, k \ge 1 \; , \; \|t_k\|\sqrt{\log(k+1)} \le L\mathsf{E}\sup_{t\in T} X_t \; , \tag{2.103}$$

$$\|t_k\| \le L\delta \; , \tag{2.104}$$

$$T \subset \overline{\mathrm{conv}}\{t_k \; ; \; k \ge 1\} \; , \tag{2.105}$$

where $\overline{\mathrm{conv}}$ denotes the closed convex hull. (Hint: copy the proof of Theorem 2.4.18, observing that since $T \subset B(0,\delta)$ one may chose $\mathcal{A}_n = \{T\}$ and $T_n = \{0\}$ for $n \le n_0$, where n_0 is the smallest integer for which $2^{n_0/2} \ge \delta^{-1}\mathsf{E}\sup_{t\in T} X_t$, and thus $U_n = \emptyset$ for $n \le n_0$.)

The purpose of the next exercise is to derive from Exercise 2.4.20 some results of Banach space theory due to S. Artstein [1]. This exercise is more elaborate, and may be omitted at first reading. A Bernoulli r.v. ε is such that $\mathsf{P}(\varepsilon = \pm 1) = 1/2$. (The reader will not confuse Bernoulli r.v.s ε_i with positive numbers ϵ_k!)

Exercise 2.4.21. In this exercise we consider a subset $T \subset \mathbb{R}^N$, where $\mathbb{R}^N$ is provided with the Euclidean distance. We assume that for some $\delta > 0$, we have

$$0 \in T \subset B(0,\delta) .$$

We consider independent Bernoulli r.v.s $(\varepsilon_{i,p})_{i,p\ge 1}$ and for $q \le N$ we consider the random operator $U_q : \mathbb{R}^N \to \mathbb{R}^q$ given by

$$U_q(x) = \Big(\sum_{i\le N} \varepsilon_{i,p} x_i\Big)_{p\le q} .$$

The purpose of the exercise is to show that there exists a number L such that if

$$q \geq \delta^{-1} \mathsf{E} \sup_{t \in T} \sum_{i \leq N} g_i t_i \ , \tag{2.106}$$

then with high probability

$$U_q(T) \subset B(0, L\delta\sqrt{q}) \ . \tag{2.107}$$

(a) Use the subgaussian inequality (3.2.2) to prove that if $\|x\| = 1$, then

$$\mathsf{E} \exp\Big(\frac{1}{4}\Big(\sum_{i \leq N} \varepsilon_{i,p} x_i\Big)^2\Big) \leq L \ . \tag{2.108}$$

(b) Use (2.108) and independence to prove that for $x \in \mathbb{R}^n$ and $v \geq 1$,

$$\mathsf{P}(\|U_q(x)\| \geq Lv\sqrt{q}\|x\|) \leq \exp(-v^2 q) \ . \tag{2.109}$$

(c) Use (2.109) to prove that with probability close to 1, for each of the vectors t_k of Exercise 2.4.20 one has $\|U_q(t_k)\| \leq L\delta\sqrt{q}$ and conclude.

The simple proof of Theorem 2.4.15 hides the fact that (2.96) is a near miraculous result. It does not provide any real understanding of what is going on. Here is a simple question.

Research problem 2.4.22. Given a subset T of the unit ball of ℓ^2, give a geometrical proof that $\gamma_2(\operatorname{conv} T) \leq L\sqrt{\log \operatorname{card} T}$.

The issue is that, while this result is true whatever the choice of T, the structure of an admissible sequence which witnesses that $\gamma_2(\operatorname{conv} T) \leq L\sqrt{\log \operatorname{card} T}$ must depend on the "geometry" of the set T.

A geometrical proof should of course not use Gaussian processes but only the geometry of Hilbert space. A really satisfactory argument would give a proof that holds in Banach spaces more general than Hilbert space, for example by providing a positive answer to the following, where the concept of q-smooth Banach space is explained in [6].

Research problem 2.4.23. Given a 2-smooth Banach space, is it true that for each subset T of its unit ball $\gamma_2(\operatorname{conv} T) \leq K\sqrt{\log \operatorname{card} T}$? More generally, is it true that for each finite subset T one has $\gamma_2(\operatorname{conv} T) \leq K\gamma_2(T)$? (Here K may depend on the Banach space, but not on T.)

Here of course we use the distance induced by the norm to compute the γ_2 functional.

Research problem 2.4.24. Still more generally, is it true that for a finite subset T of a q-smooth Banach space, one has $\gamma_q(\operatorname{conv} T) \leq K\gamma_\alpha(T)$?

Even when the Banach space is ℓ^p, I do not know the answer to these problems (unless $p = 2$!). (The Banach space ℓ^p is 2-smooth for $p \geq 2$ and q-smooth for $p < 2$, where $1/p + 1/q = 1$.) One concrete case is when the set T consists of the first N vectors of the unit basis of ℓ^p. It is possible to show in this case that $\gamma_q(\operatorname{conv} T) \leq K(p)(\log N)^{1/q}$, where $1/p + 1/q = 1$. We leave this as a challenge to the reader. The proof for the general case is pretty much the same as for the case $p = q = 2$ which was already proposed as a challenge after Exercise 2.2.15.

2.5 A First Look at Ellipsoids

We have illustrated the gap between Dudley's bound (2.38) and the sharper bound (2.32), using the examples (2.42) and (2.89). Perhaps the reader deems these examples artificial, and believes that "in all practical situations" Dudley's bound suffices. Before we prove Theorem 2.3.16 (thus completing the proof of the Majorizing Measure Theorem 2.4.1) in the next section, we feel that it may be useful to provide some more motivation by demonstrating that the gap between Dudley's bound (2.38) and the generic chaining bound (2.32) already exists for *ellipsoids* in Hilbert space. It is hard to argue that ellipsoids are artificial, unnatural or unimportant. Moreover, understanding ellipsoids will be fundamental in several subsequent questions, such as the matching theorems of Chapter 4.

Given a sequence $(a_i)_{i \geq 1}\,, a_i > 0$, we consider the ellipsoid

$$\mathcal{E} = \left\{ t \in \ell^2 \; ; \; \sum_{i \geq 1} \frac{t_i^2}{a_i^2} \leq 1 \right\} . \tag{2.110}$$

Proposition 2.5.1. *We have*

$$\frac{1}{L} \left(\sum_{i \geq 1} a_i^2 \right)^{1/2} \leq \mathsf{E} \sup_{t \in \mathcal{E}} X_t \leq \left(\sum_{i \geq 1} a_i^2 \right)^{1/2} . \tag{2.111}$$

Proof. The Cauchy-Schwarz inequality implies

$$Y := \sup_{t \in \mathcal{E}} X_t = \sup_{t \in \mathcal{E}} \sum_{i \geq 1} t_i g_i \leq \left(\sum_{i \geq 1} a_i^2 g_i^2 \right)^{1/2} . \tag{2.112}$$

Taking $t_i = a_i^2 g_i / (\sum_{j \geq 1} a_j^2 g_j^2)^{1/2}$ yields that actually $Y = (\sum_{i \geq 1} a_i^2 g_i^2)^{1/2}$ and thus $\mathsf{E} Y^2 = \sum_{i \geq 1} a_i^2$. The right-hand side of (2.111) follows from the Cauchy-Schwarz inequality:

$$\mathsf{E} Y \leq (\mathsf{E} Y^2)^{1/2} = \left(\sum_{i \geq 1} a_i^2 \right)^{1/2} . \tag{2.113}$$

For the left-hand side, let $\sigma = \max_{i\geq 1} |a_i|$. Since $Y = \sup_{t\in\mathcal{E}} X_t \geq |a_i||g_i|$ for any i, we have $\sigma \leq L\mathsf{E}Y$. Also,

$$\mathsf{E}X_t^2 = \sum_i t_i^2 \leq \max_i a_i^2 \sum_j \frac{t_j^2}{a_j^2} \leq \sigma^2 \ . \tag{2.114}$$

Then (2.84) implies

$$\mathsf{E}(Y - \mathsf{E}Y)^2 \leq L\sigma^2 \leq L(\mathsf{E}Y)^2 \ ,$$

so that $\sum_{i\geq 1} a_i^2 = \mathsf{E}Y^2 = \mathsf{E}(Y - \mathsf{E}Y)^2 + (\mathsf{E}Y)^2 \leq L(\mathsf{E}Y)^2$. □

As a consequence of Theorem 2.4.1,

$$\gamma_2(\mathcal{E}) \leq L\Bigl(\sum_{i\geq 1} a_i^2\Bigr)^{1/2} \ . \tag{2.115}$$

This statement is purely about the geometry of ellipsoids. The proof we gave was rather indirect, since it involved Gaussian processes. Later on, in Theorem 4.1.11, we shall give a "purely geometric" proof of this result that will have many consequences.

Let us now assume that the sequence $(a_i)_{i\geq 1}$ is non-increasing. Since

$$2^n \leq i \leq 2^{n+1} \Rightarrow a_{2^n} \geq a_i \geq a_{2^{n+1}}$$

we get

$$\sum_{i\geq 1} a_i^2 = \sum_{n\geq 0} \sum_{2^n\leq i<2^{n+1}} a_i^2 \leq \sum_{n\geq 0} 2^n a_{2^n}^2$$

and

$$\sum_{i\geq 1} a_i^2 \geq \sum_{n\geq 0} 2^n a_{2^{n+1}}^2 = \frac{1}{2}\sum_{n\geq 1} 2^n a_{2^n}^2 \ ,$$

and thus $\sum_{n\geq 0} 2^n a_{2^n}^2 \leq 3\sum_{i\geq 1} a_i^2$. So we may rewrite (2.111) as

$$\frac{1}{L}\Bigl(\sum_{n\geq 0} 2^n a_{2^n}^2\Bigr)^{1/2} \leq \mathsf{E}\sup_{t\in\mathcal{E}} X_t \leq \Bigl(\sum_{n\geq 0} 2^n a_{2^n}^2\Bigr)^{1/2} \ . \tag{2.116}$$

Proposition 2.5.1 describes the size of ellipsoids with respect to Gaussian processes. Our next result describes their size with respect to Dudley's entropy bound (2.36).

Proposition 2.5.2. *We have*

$$\frac{1}{L}\sum_{n\geq 0} 2^{n/2} a_{2^n} \leq \sum_{n\geq 0} 2^{n/2} e_n(\mathcal{E}) \leq L\sum_{n\geq 0} 2^{n/2} a_{2^n} \ . \tag{2.117}$$

The right-hand sides in (2.116) and (2.117) are distinctively different. Dudley's bound fails to describe the behavior of Gaussian processes on ellipsoids. This is a simple occurrence of a general phenomenon. In some sense an ellipsoid is smaller than what one would predict just by looking at its entropy numbers $e_n(\mathcal{E})$. This idea will be investigated further in Section 4.1.

Exercise 2.5.3. Prove that for an ellipsoid $\mathcal{E}$ of $\mathbb{R}^m$ one has

$$\sum_{n \geq 0} 2^{n/2} e_n(\mathcal{E}) \leq L\sqrt{\log(m+1)}\gamma_2(T,d) \ ,$$

and that this estimate is essentially optimal. Compare with (2.67).

The proof of (2.117) hinges on ideas which are at least 50 years old, and which relate to the methods of Exercise 2.2.14. The left-hand side is the easier part (it is also the most important for us). It follows from the next lemma, the proof of which is basically a special case of (2.39).

Lemma 2.5.4. *We have* $e_n(\mathcal{E}) \geq \frac{1}{2} a_{2^n}$.

Proof. Consider the following ellipsoid in $\mathbb{R}^{2^n}$:

$$\mathcal{E}_n = \left\{ (t_i)_{i \leq 2^n} \ ; \ \sum_{i \leq 2^n} \frac{t_i^2}{a_i^2} \leq 1 \right\} .$$

Since $\mathcal{E}_n$ is the image of $\mathcal{E}$ by a contraction (namely the "projection on the first 2^n coordinates") it holds that $e_n(\mathcal{E}_n) \leq e_n(\mathcal{E})$.

Let us denote by B the centered unit Euclidean ball of $\mathbb{R}^{2^n}$ and by Vol the volume in this space. Let us consider a subset T of $\mathcal{E}_n$, with $\operatorname{card} T \leq 2^{2^n}$, and $\epsilon > 0$; then

$$\mathrm{Vol}\Bigl(\bigcup_{t \in T} (\epsilon B + t)\Bigr) \leq \sum_{t \in T} \mathrm{Vol}(\epsilon B + t) \leq 2^{2^n} \epsilon^{2^n} \mathrm{Vol} B = (2\epsilon)^{2^n} \mathrm{Vol} B \, .$$

On the other hand, since $a_i \geq a_{2^n}$ for $i \leq 2^n$, we have $a_{2^n} B \subset \mathcal{E}_n$, so that $\mathrm{Vol}\mathcal{E}_n \geq a_{2^n}^{2^n} \mathrm{Vol} B$. Thus when $2\epsilon < a_{2^n}$, we cannot have $\mathcal{E}_n \subset \bigcup_{t \in T} (\epsilon B + t)$. Therefore $e_n(\mathcal{E}_n) \geq \epsilon$. □

We now turn to the upper bound, which relies on a special case of (2.40).

Lemma 2.5.5. *We have*

$$e_{n+3}(\mathcal{E}) \leq 3 \max_{k \leq n} (a_{2^k} 2^{k-n}) \ . \tag{2.118}$$

Proof. We keep the notation of the proof of Lemma 2.5.4. First we show that

$$e_{n+3}(\mathcal{E}) \leq e_{n+3}(\mathcal{E}_n) + a_{2^n} \ . \tag{2.119}$$

To see this, we observe that when $t \in \mathcal{E}$, then

$$1 \geq \sum_{i \geq 1} \frac{t_i^2}{a_i^2} \geq \sum_{i > 2^n} \frac{t_i^2}{a_i^2} \geq \frac{1}{a_{2^n}^2} \sum_{i > 2^n} t_i^2$$

so that $(\sum_{i>2^n} t_i^2)^{1/2} \leq a_{2^n}$ and, viewing $\mathcal{E}_n$ as a subset of $\mathcal{E}$, we have $d(t, \mathcal{E}_n) \leq a_{2^n}$. Thus if we cover $\mathcal{E}_n$ by certain balls with radius ϵ, the balls with the same centers but radius $\epsilon + a_{2^n}$ cover $\mathcal{E}$. This proves (2.119).

Consider now $\epsilon > 0$, and a subset Z of $\mathcal{E}_n$ with the following properties:

$$\text{any two points of } Z \text{ are at mutual distance} \geq 2\epsilon \tag{2.120}$$

$$\operatorname{card} Z \text{ is as large as possible under (2.120).} \tag{2.121}$$

Then by (2.121) the balls centered at points of Z and with radius $\leq 2\epsilon$ cover $\mathcal{E}_n$. Thus

$$\operatorname{card} Z \leq N_{n+3} \Rightarrow e_{n+3}(\mathcal{E}_n) \leq 2\epsilon\ . \tag{2.122}$$

The balls centered at the points of Z, with radius ϵ, have disjoint interiors, so that

$$\operatorname{card} Z \ \mathrm{Vol}(\epsilon B) \leq \mathrm{Vol}(\mathcal{E}_n + \epsilon B)\ . \tag{2.123}$$

Now for $t = (t_i)_{i \leq 2^n} \in \mathcal{E}_n$, we have $\sum_{i \leq 2^n} t_i^2/a_i^2 \leq 1$, and for t' in ϵB, we have $\sum_{i \leq 2^n} t_i'^2/\epsilon^2 \leq 1$. Let $c_i = 2\max(\epsilon, a_i)$. Since

$$\frac{(t_i + t_i')^2}{c_i^2} \leq \frac{2t_i^2 + 2t_i'^2}{c_i^2} \leq \frac{1}{2}\left(\frac{t_i^2}{a_i^2} + \frac{t_i'^2}{\epsilon^2}\right)\ ,$$

we have

$$\mathcal{E}_n + \epsilon B \subset \mathcal{E}^1 := \left\{ t\ ;\ \sum_{i \leq 2^n} \frac{t_i^2}{c_i^2} \leq 1 \right\}\ .$$

Therefore

$$\mathrm{Vol}(\mathcal{E}_n + \epsilon B) \leq \mathrm{Vol}\mathcal{E}^1 = \mathrm{Vol}B \prod_{i \leq 2^n} c_i$$

and comparing with (2.123) yields

$$\operatorname{card} Z \leq \prod_{i \leq 2^n} \frac{c_i}{\epsilon} = 2^{2^n} \prod_{i \leq 2^n} \max\left(1, \frac{a_i}{\epsilon}\right)\ .$$

Assume now that for any $k \leq n$ we have $a_{2^k} 2^{k-n} \leq \epsilon$. Then $a_i \leq a_{2^k} \leq \epsilon 2^{n-k}$ for $2^k < i \leq 2^{k+1}$, so that

$$\begin{aligned}\prod_{i \leq 2^n} \max\left(1, \frac{a_i}{\epsilon}\right) &= \prod_{k \leq n-1} \prod_{2^k < i \leq 2^{k+1}} \max\left(1, \frac{a_i}{\epsilon}\right) \\ &\leq \prod_{k \leq n-1} \left(2^{n-k}\right)^{2^k} = 2^{\sum_{k \leq n} (n-k) 2^k} \\ &\leq 2^{2^{n+2}}\end{aligned}$$

since $\sum_{i \geq 0} i 2^{-i} = 4$.

To sum up, if $\epsilon = \max_{k \le n} a_{2^k} 2^{k-n}$, we have shown that

$$\operatorname{card} Z \le 2^{2^n} \cdot 2^{2^{n+2}} \le N_{n+3} \,,$$

so that $e_{n+3}(\mathcal{E}_n) \le 2\epsilon$. The conclusion follows from (2.119). □

Proof of Proposition 2.5.2. We have, using (2.118)

$$\begin{aligned}\sum_{n\ge 3} 2^{n/2} e_n(\mathcal{E}) &= \sum_{n \ge 0} 2^{(n+3)/2} e_{n+3}(\mathcal{E}) \\ &\le L \sum_{n\ge 0} 2^{n/2} \Bigl(\sum_{k \le n} 2^{k-n} a_{2^k}\Bigr) \\ &\le L \sum_{k \ge 0} 2^k a_{2^k} \sum_{n \ge k} 2^{-n/2} \\ &\le L \sum_{k\ge 0} 2^{k/2} a_{2^k} \;.\end{aligned}$$

Since $\mathcal{E}$ is contained in the ball centered at the origin with radius a_1, we have $e_n(\mathcal{E}) \le a_1$ for each n. The result follows. □

2.6 Proof of the Fundamental Partitioning Theorem

In this section we prove Theorem 2.3.16.

Theorem 2.6.1. *Assume that on the metric space (T, d) there exists a decreasing sequence of functionals $(F_n)_{n \ge 0}$ that satisfies the growth condition of Definition 2.3.10. Then we can find an increasing sequence of partitions $(\mathcal{A}_n)$ with* $\operatorname{card} \mathcal{A}_n \le N_{n+1}$ *and*

$$\sup_{t \in T} \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \le \frac{Lr}{c^*} F_0(T) + Lr\Delta(T) \,. \tag{2.124}$$

This is not exactly Theorem 2.3.16 because here we have $\operatorname{card} \mathcal{A}_n \le N_{n+1}$ rather than $\operatorname{card} \mathcal{A}_n \le N_n$, but Theorem 2.3.16 follows by combining Theorem 2.6.1 with Lemma 2.3.5.

Replacing F_n by F_n/c^* it suffices to consider the case $c^* = 1$, so we assume this condition throughout this section.

Before going into the details let us first explain the principle of the construction. We construct the increasing sequence $(\mathcal{A}_n)$ of partitions by induction, starting of course with $\mathcal{A}_0 = \{T\}$. Together with $C \in \mathcal{A}_n$, we will construct a point $t_{n,C}$ of T, and an integer $j_n(C)$ in $\mathbb{Z}$. We assume

$$C \subset B(t_{n,C}, r^{-j_n(C)}) \,, \tag{2.125}$$

so that in particular

$$\Delta(C) \le 2r^{-j_n(C)} . \tag{2.126}$$

Thus, we may think of $j_n(C)$ as keeping track of the diameter of C. More accurately, $j_n(C)$ keeps track of a convenient upper bound for the diameter of C, as it may well happen that $\Delta(C)$ is much smaller than $2r^{-j_n(C)}$. We do *not* require that $t_{n,C}$ belongs to C.

To start the construction, we set $\mathcal{A}_0 = \{T\}$, and we choose any point $t_{0,T} \in T$. We then take for $j_0(T)$ the largest possible integer such that $T \subset B(t_{0,T}, r^{-j_0(T)})$, so that

$$r^{-j_0(T)} \le r\Delta(T) . \tag{2.127}$$

Let us now assume that for a certain $n \ge 0$ we have already constructed the partition $\mathcal{A}_n$ with $\operatorname{card}\mathcal{A}_n \le N_{n+1}$. To construct $\mathcal{A}_{n+1}$ we will split each set of $\mathcal{A}_n$ in at most N_{n+1} pieces according to Lemma 2.6.2 below. Since $N_{n+1}^2 \le N_{n+2}$ we will have $\operatorname{card}\mathcal{A}_{n+1} \le N_{n+2}$, and in this manner we will construct the corresponding increasing sequence of partitions $\mathcal{A}_n$.

All the magic of course is in the procedure by which we will split a given element of $\mathcal{A}_n$ into pieces and in the information that we gather while doing so. To describe this procedure, let us fix $C \in \mathcal{A}_n$, and let $j = j_n(C)$.

Lemma 2.6.2 (The Decomposition Lemma). *Consider a subset C of T, an integer $n \ge 0$ and $j \in \mathbb{Z}$. Let $m = N_{n+1}$. Assume that for a certain $t_C \in T$ we have $C \subset B(t_C, r^{-j})$. Then we can find $m' \le m$ and a partition $(A_\ell)_{\ell \le m'}$ such that for each $\ell \le m'$ we have* **either**

$$\exists t_\ell \in C \ , \ A_\ell \subset B(t_\ell, r^{-j-1}) \ , \tag{2.128}$$

or else

$$r^{-j-1}2^{n/2-1} + \sup_{t \in A_\ell} F_{n+1}(A_\ell \cap B(t, r^{-j-2})) \le F_n(C) . \tag{2.129}$$

Thus we split C into two kinds of pieces. Those that satisfy (2.128) are of "smaller diameter" than C itself. For those that satisfy (2.129), we gain some (still mysterious) control on the behavior of the functionals F_n. Two noticeable features of this proof are that it is "algorithmic" (the construction is obtained by repeating a basic simple step until the entire set C has been used up) and "greedy" in that the basic simple step maximizes some simple measure of "gain".

Proof. The proof will show in fact that for $\ell < m$ the set A_ℓ satisfies (2.128) and that if $\ell = m = m'$ the set $A_\ell = A_m$ satisfies (2.129). (The present formulation is motivated by pedagogical reason, as it makes the exposition easier in more complicated cases.) To avoid being distracted by secondary issues, let us first assume that T is finite. By induction over $1 \le \ell \le m = N_{n+1}$ we construct points $t_\ell \in C$ and sets $A_\ell \subset C$ as follows.

First, we set $D_0 = C$ and we choose t_1 in C such that

$$F_{n+1}(C \cap B(t_1, r^{-j-2})) = \sup_{t \in C} F_{n+1}(C \cap B(t, r^{-j-2})) . \tag{2.130}$$

We then set $A_1 = C \cap B(t_1, r^{-j-1})$. The idea is simply that "we take the largest possible piece of C" (it is in this sense that the method is "greedy"). The reader notices that the radius of the balls in (2.130) is r^{-j-2} while it is r^{-j-1} in the definition of A_1. This is the main idea of the proof. A "large piece" of C is a piece of the type $A_1 = C \cap B(t_1, r^{-j-1})$ for which $F_{n+1}(C \cap B(t_1, r^{-j-2}))$ (rather than $F_{n+1}(A_1)$) is large. This construction is perfectly appropriate in order to use the growth condition of Definition 2.3.10, as it naturally creates well separated "large" pieces (of which $C \cap B(t_1, r^{-j-2})$ is the first one). The drawback of the construction is that the information we produce "skips a level" since it pertains to smaller balls than those we would like (with radius r^{-j-2} rather than r^{-j-1}), and the key point of the proof will be to show that we can at some stage recover the information about the "skipped level".

To continue the construction, assume now that $t_1, \ldots, t_\ell$ and $A_1, \ldots, A_\ell$ have already been constructed, and set $D_\ell = C \setminus \bigcup_{1 \le p \le \ell} A_p$. If $D_\ell = \emptyset$, we set $m' = \ell$ and the construction stops. Otherwise, we choose $t_{\ell+1}$ in D_ℓ such that

$$F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) = \sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, r^{-j-2})) . \tag{2.131}$$

We set $A_{\ell+1} = D_\ell \cap B(t_{\ell+1}, r^{-j-1})$ and we continue in this manner until either we stop or we construct

$$D_{m-1} = C \setminus \bigcup_{\ell < m} A_\ell .$$

If D_{m-1} is empty, the construction is finished. Otherwise we set $A_m = D_{m-1}$, so that $A_1, \ldots, A_m$ form a partition of C. In this manner we have partitioned C in at most m pieces.

If $\ell < m$ it is obvious by construction that (2.128) holds, so that to finish the proof it suffices to show that (2.129) holds for $\ell = m$. The proof relies on the growth condition. (Let us observe for future use that it actually suffices for the proof that the growth condition holds whenever a is of the type $a = r^{-j'-1}$ for a certain $j' \in \mathbb{Z}$, and that other values of a are not needed.) Then (2.73) rewrites as

$$\forall \ell \le m\,,\, t_\ell \in B(s, r^{-j})\ ;\ \forall \ell\,,\, \ell' \le m\,,\, \ell \ne \ell' \Rightarrow d(t_\ell, t_{\ell'}) \ge r^{-j-1}\,, \tag{2.132}$$

and the content of the growth condition is that this implies (since $c^* = 1$)

$$\begin{aligned} &\forall\ \ell \le m\,,\, H_\ell \subset B(t_\ell, r^{-j-2}) \\ &\Rightarrow F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge r^{-j-1} 2^{n/2} + \min_{\ell \le m} F_{n+1}(H_\ell) . \end{aligned} \tag{2.133}$$

Let us construct a point $t_m \in A_m = D_{m-1}$ as in (2.131) for $\ell = m-1$. All the points $(t_\ell)_{\ell \le m}$ belong to $C \subset B(t_C, r^{-j})$. For $\ell < m$ we have by construction

$$t_{\ell+1} \in D_\ell = C \setminus \bigcup_{1 \le p \le \ell} A_p = C \setminus \bigcup_{1 \le p \le \ell} B(t_p, r^{-j-1}) ,$$

and therefore $d(t_{\ell+1}, t_p) \ge r^{-j-1}$ for $p \le \ell$. Consequently these points satisfy (2.132) for $s = t_C$, and therefore (2.133) holds for $H_\ell = D_{\ell-1} \cap B(t_\ell, r^{-j-2})$, where we recall that $D_0 = C$. Since $H_\ell \subset C$, we obtain

$$F_n(C) \ge F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge r^{-j-1} 2^{n/2} + \min_{\ell \le m} F_{n+1}(H_\ell) . \tag{2.134}$$

Now, it follows from (2.131) that for $1 \le \ell \le m-1$

$$\begin{aligned}\sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, r^{-j-2})) &\le F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) \\ &= F_{n+1}(H_{\ell+1}) ,\end{aligned}$$

and (2.130) implies that this is also true when $\ell = 0$. Since the sequence (D_ℓ) decreases, this implies that for $0 \le \ell < m$ we have

$$\sup_{t \in D_{m-1}} F_{n+1}(D_{m-1} \cap B(t, r^{-j-2})) \le F_{n+1}(H_{\ell+1})$$

and therefore

$$\sup_{t \in D_{m-1}} F_{n+1}(D_{m-1} \cap B(t, r^{-j-2})) \le \min_{1 \le \ell \le m} F_{n+1}(H_\ell) .$$

Combining with (2.134) we finally obtain (since $A_m = D_{m-1}$)

$$r^{-j-1} 2^{n/2} + \sup_{t \in A_m} F_{n+1}(A_m \cap B(t, r^{-j-2})) \le F_n(C) , \tag{2.135}$$

and this finishes the proof when T is finite. When T need not be finite, we set $\epsilon = r^{-j-1} 2^{n/2-1}$ and we replace (2.131) by

$$F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) \ge \sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, r^{-j-2})) - \epsilon , \tag{2.136}$$

and rather than (2.135) we reach

$$r^{-j-1} 2^{n/2} + \sup_{t \in A_m} F_{n+1}(A_m \cap B(t, r^{-j-2})) \le F_n(C) + \epsilon .$$

Recalling the value of ϵ finishes the proof. □

We now continue the construction proving Theorem 2.6.1. We split the set $C \in \mathcal{A}_n$ into at most m pieces using the Decomposition Lemma (Lemma 2.6.2), and we consider one of these pieces A.

If $A = A_\ell$ satisfies (2.128), we define $j_{n+1}(A) = j + 1 = j_n(C) + 1$ and $t_{n+1,A} = t_\ell$, so that

$$A = A_\ell \subset B(t_\ell, r^{-j-1}) = B(t_{n+1,A}, r^{-j_{n+1}(A)}) .$$

Let us stress for further use that in that case $t_{n+1,A} \in C$.

If $A = A_\ell$ satisfies (2.129), we define instead $j_{n+1}(A) = j(= j_n(C))$ and $t_{n+1,A} = t_{n,C}$, so that

$$A \subset C \subset B(t_{n,C}, r^{-j_n(C)}) = B(t_{n+1,A}, r^{-j_{n+1}(A)}) .$$

This completes the basic procedure and the construction, and we turn to the proof of (2.124). First we observe that for any $t \in T$, (2.126) implies

$$\sum_{n\geq 0} 2^{n/2} \Delta(A_n(t)) \leq 2 \sum_{n\geq 0} r^{-j_n(A_n(t))} 2^{n/2} , \tag{2.137}$$

and our objective is to bound the right-hand side. We fix t in T once and for all. It turns out that in the right-hand side of (2.137) only certain terms really contribute. We develop this idea in the next lemma. The basic observation is simply that the sum of a geometric series can be basically bounded by either the first or the last term of the series.

Lemma 2.6.3. *Consider numbers* $(a_n)_{n\geq 0}$*,* $a_n \geq 0$*, and assume* $\sup_n a_n < \infty$*. Consider* $\alpha > 1$ *and define*

$$I = \left\{k \geq 0 \,;\, \forall n \geq 0\,, n \neq k\,,\ a_n < a_k \alpha^{|k-n|}\right\} . \tag{2.138}$$

Then

$$\sum_{n\geq 0} a_n \leq \frac{2\alpha}{\alpha - 1} \sum_{k\in I} a_k . \tag{2.139}$$

Proof. Let us write $n \prec k$ when $a_k \geq a_n \alpha^{|n-k|}$. This relation is a partial order: if $n \prec k$ and $k \prec p$ then $a_p \geq a_n \alpha^{|p-k|+|k-n|} \geq a_n \alpha^{|p-n|}$, so that $n \prec p$. Let us observe that the set I defined above is the set of elements k of $\mathbb{N}$ that are maximal, i.e. $k \prec k' \Rightarrow k = k'$. Since we assume that the sequence (a_n) is bounded, there cannot exist an increasing sequence for the order $\prec$. Consequently, for each n in $\mathbb{N}$ there exists $k \in I$ with $n \prec k$. Then $a_n \leq a_k \alpha^{-|n-k|}$, and therefore

$$\sum_{n\geq 0} a_n \leq \sum_{k\in I} \sum_{n\geq 0} a_k \alpha^{-|k-n|} \leq \frac{2}{1-\alpha^{-1}} \sum_{k\in I} a_k . \qquad \square$$

We go back to the control of the right-hand side of (2.137). We recall that $r \geq 4$. To lighten notation we set $j(n) = j_n(A_n(t))$, and we set $a_n = r^{-j(n)} 2^{n/2}$. This sequence is bounded because either $j(n) > j(n-1)$ and then $a(n) \leq a(n-1)$, or else $a_n \leq F_0(T)$ by (2.129). Consider the set I provided by Lemma 2.6.3 for $\alpha = \sqrt{2}$. We observe the following fundamental relation:

$$k \in I \; , \; k \geq 1 \Rightarrow j(k-1) = j(k) \; , \; j(k+1) = j(k) + 1 \; . \tag{2.140}$$

Indeed, if $j(k+1) = j(k)$, then $a_{k+1} = \sqrt{2} a_k$, so that $k \notin I$ by the definition of I, and if $j(k-1) = j(k) - 1$ then $a_{k-1} = (r/\sqrt{2}) a_k \geq 2 a_k$, and again $k \notin I$ by definition of I.

Lemma 2.6.4. *Consider elements* $1 \leq k < k'$ *of* I. *Then*

$$\frac{1}{4r} a_k \leq F_{k-1}(A_{k-1}(t)) - F_{k'+1}(A_{k'+1}(t)) \; . \tag{2.141}$$

Proof. It follows from (2.125) that if we define $A^* := A_{k'+1}(t)$ and $t^* := t_{k'+1, A^*}$ then

$$A^* \subset B(t^*, r^{-j(k'+1)}) \; .$$

Moreover, since $k' \in I$ we have $j(k'+1) = j(k') + 1$, and as noted we have $t^* \in A_{k'}(t) \subset A_k(t)$. Also $j(k') \geq j(k+1)$, and $j(k+1) = j(k) + 1$ since $k \in I$ and $k \geq 1$. Consequently, $j(k'+1) \geq j(k) + 2$ and therefore

$$A^* \subset A_k(t) \cap B(t^*, r^{-j(k)-2}) \; .$$

Moreover, since $k \in I$ and $k \geq 1$, we have $j(k-1) = j(k)$. By construction, (2.129) used for $n = k-1$ and $C = A_n(t) = A_{k-1}(t)$ implies

$$r^{-j(k)-1} 2^{(k-1)/2-1} + \sup_{u \in A_k(t)} F_k(A_k(t) \cap B(u, r^{-j(k)-2})) \leq F_{k-1}(A_{k-1}(t)) \; , \tag{2.142}$$

so that since $r^{-j(k)-1} 2^{(k-1)/2-1} \geq a_k/4r$, (2.142) implies

$$\frac{1}{4r} a_k + F_k(A^*) \leq F_{k-1}(A_{k-1}(t)) \; . \tag{2.143}$$

Since $k \leq k'$ and since the sequence (F_n) decreases, we have $F_k(A^*) \geq F_{k'+1}(A_{k'+1}(t))$ and (2.143) proves (2.141). $\square$

Proof of Theorem 2.6.1. Let

$$x(n) = F_n(A_n(t)) \; ,$$

so that (2.141) implies

$$\frac{1}{4r} a_k \leq x(k-1) - x(k'+1) \; .$$

Moreover, since the sequence (F_n) of functionals decreases, and since the sequence of sets $(A_n(t))$ decreases, the sequence $(x(n))$ decreases.

Let us assume first that I is infinite and let us enumerate I as an increasing sequence $(k_i)_{i\geq 1}$. For $i \geq 1$ let us define $y(i) = x(k_i)$, so that the sequence $(y(i))$ decreases since the sequence $(x(n))$ decreases. For $i \geq 2$ we have $k_i - 1 \geq k_{i-1}$ so that $x(k_i - 1) \leq y(i-1)$. Similarly, $x(k_{i+1}+1) \geq x(k_{i+2}) = y(i+2)$. Since $k_i \geq 1$ (2.141) implies

$$\frac{1}{4r} a_{k_i} \leq y(i-1) - y(i+2) \,. \tag{2.144}$$

Since $y(i) \leq x(0) = F_0(A_0(t)) = F_0(T)$, summation of the inequalities (2.144) yields

$$\sum_{i \geq 2} a_{k_i} \leq LrF_0(T) \,. \tag{2.145}$$

It only remains to control a_{k_1}. When $k_1 = 0$, then $a_0 = r^{-j_0(T)} \leq r\Delta(T)$. Otherwise $k_1 \geq 1$, and then (2.143) implies $a_{k_1} \leq 4rF_0(T)$. This completes the proof when I is infinite. Only small changes are required when I is finite, and this is left to the reader. □

2.7 A General Partitioning Scheme

Theorem 2.6.1 admits considerable generalizations, which turn out to be very useful. These generalizations admit basically the same proof as Theorem 2.6.1. They require an extension of the "growth condition" of Definition 2.3.10. We consider a function

$$\theta : \mathbb{N} \cup \{0\} \to \mathbb{R}^+ \,.$$

Definition 2.3.10 corresponds to the case $\theta(n) = 2^{(n-1)/2}$.

The condition we are about to state involves two new parameters β and τ. Definition 2.3.10 corresponds to the case $\beta = 1$ and $\tau = 1$. The parameter $\tau \in \mathbb{N}$ is of secondary importance. The larger τ, the more "room there is".

Let us recall that since Definition 2.3.8, we say that sets $(H_\ell)_{\ell \leq m}$ are (a, r) separated if there exist $s, t_1, \ldots, t_m$ for which

$$\forall \ell \leq m \,,\, t_\ell \in B(s, ar) \,;\, \forall \ell \,,\, \ell' \leq m \,,\, \ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a \,, \tag{2.146}$$

and

$$\forall \ell \leq m \,,\, H_\ell \subset B(t_\ell, a/r) \,.$$

Definition 2.7.1. *We say that the functionals F_n satisfy the* growth condition *if for a certain integer $\tau \geq 1$, and for certain numbers $r \geq 4$ and $\beta > 0$, the following holds true. Consider $a > 0$, any integer $n \geq 0$, and set*

$m = N_{n+\tau}$. *Then, whenever the subsets* $(H_\ell)_{\ell \le m}$ *of* T *are* (a, r) *separated in the sense of Definition 2.3.8, then*

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge a^\beta \theta(n+1) + \min_{\ell \le m} F_{n+1}(H_\ell) \; . \tag{2.147}$$

In the right-hand side of (2.147), the term $a^\beta\theta(n+1)$ is the product of a^β, which accounts for the scale at which the sets H_ℓ are separated, and of the term $\theta(n+1)$, which accounts for the number of these sets. The "linear case" $\beta = 1$ is by far the most important. The role of the parameter τ is to give some room. When τ is large, there are more sets and it should be easier to prove (2.147).

The reader noticed that we call "growth condition" both the condition of Definition 2.3.10 and the more general condition of Definition 2.7.1. It is not practical to give different names to these conditions because we shall eventually consider several more conditions in the same spirit. We shall always make precise to which condition we refer.

We will assume the following regularity condition for θ. For some $1 < \xi \le 2$, and all $n \ge 0$, we have

$$\xi\theta(n) \le \theta(n+1) \le \frac{r^\beta}{2}\theta(n) \; . \tag{2.148}$$

When $\theta(n) = 2^{(n-1)/2}$, (2.148) holds for $\xi = \sqrt{2}$. The main result of this section is as follows.

Theorem 2.7.2. *Under the preceding conditions we can find an increasing sequence* $(\mathcal{A}_n)$ *of partitions of* T *with* $\operatorname{card}\mathcal{A}_n \le N_{n+\tau}$ *such that*

$$\sup_{t \in T} \sum_{n \ge 0} \theta(n)\Delta(A_n(t))^\beta \le L(2r)^\beta\Big(\frac{F_0(T)}{\xi - 1} + \theta(0)\Delta(T)^\beta\Big) \; . \tag{2.149}$$

In all the situations we shall consider, it will be true that $F_0(\{t_1, t_2\}) \ge \theta(0)d(t_1, t_2)^\beta$ for any points t_1 and t_2 of T. (Since $F_1(H) \ge 0$ for any set H, this condition is essentially weaker in spirit than (2.147) for $n = 0$.) Then $\theta(0)\Delta(T)^\beta \le F_0(T)$.

The sequence $(\mathcal{A}_n)$ of Theorem 2.7.2 need not be admissible because $\operatorname{card}\mathcal{A}_n$ is too large. To construct good admissible sequences we will combine Theorem 2.7.2 with Lemma 2.3.5.

Not surprisingly, the key to the proof of Theorem 2.7.2 is the following, which is simply an adaptation of Lemma 2.6.2 to the present setting.

Lemma 2.7.3. *If the functionals* F_n *satisfy the growth condition, then, given integers* $n \ge 0$ *and* $j \in \mathbb{Z}$, *for any subset* C *of* T *such that*

$$\exists s \in T \; ; \; C \subset B(s, r^{-j}) \; ,$$

we can find a partition $(A_\ell)_{\ell \le m'}$ of C, where $m' \le m = N_{n+\tau}$, such that for each $\ell \le m'$ we have **either**

$$\exists t_\ell \in C \; ; \; A_\ell \subset B(t_\ell, r^{-j-1}) \; , \tag{2.150}$$

or else

$$\frac{1}{2} r^{-\beta(j+1)} \theta(n+1) + \sup_{t \in A_\ell} F_{n+1}(A_\ell \cap B(t, r^{-j-2})) \le F_n(C) \; . \tag{2.151}$$

The proof is nearly identical to the proof of Lemma 2.6.2 so it is left to the reader.

Proof of Theorem 2.7.2. We construct the sequence of partitions $(\mathcal{A}_n)$ and $t_{n,A}, j_n(A)$ for $A \in \mathcal{A}_n$ as in Theorem 2.6.1, using the Decomposition Lemma at each step. Since, however, there is no point in repeating the same proof, we will organize the argument differently.

The basic idea is that (since we have "skipped levels") we must keep track not only of what we do in the current step of the construction but also of what we do in the previous step. This is implemented by keeping track for each set $C \in \mathcal{A}_n$ of three different "measures of its size", namely

$$a_i(C) = \sup_{t \in C} F_n(C \cap B(t, r^{-j_n(C)-i})) \; ,$$

for $i = 0, 1, 2$. This quantity depends also on n, in the sense that if $C \in \mathcal{A}_n$ and $C \in \mathcal{A}_{n+1}$ than $a_i(C)$ need not be the same whether we see C as an element of $\mathcal{A}_n$ or of $\mathcal{A}_{n+1}$. To lighten notation we shall not indicate this dependence. For technical reasons keeping track of the values $a_j(C)$ is not very convenient, and instead we will keep track of three quantities $b_j(C)$ for $j = 0, 1, 2$, where $b_j(C)(\ge a_j(C))$ is a kind of "regularized version" of $a_j(C)$. (These quantities also depend on n.) We rewrite the conditions $a_j(C) \le b_j(C)$:

$$F_n(C) \le b_0(C) \tag{2.152}$$

$$\forall t \in C \; , \; F_n(C \cap B(t, r^{-j_n(C)-1})) \le b_1(C) \tag{2.153}$$

$$\forall t \in C \; , \; F_n(C \cap B(t, r^{-j_n(C)-2})) \le b_2(C) \; . \tag{2.154}$$

We will also require the following two technical conditions:

$$b_1(C) \le b_0(C) \tag{2.155}$$

and

$$b_0(C) - \frac{1}{2} r^{-\beta(j_n(C)+1)} \theta(n) \le b_2(C) \le b_0(C) \; . \tag{2.156}$$

Moreover, the quantities b_i will satisfy the following fundamental relation: if $n \ge 0$, $A \in \mathcal{A}_{n+1}$, $C \in \mathcal{A}_n$, $A \subset C$, then

$$\begin{aligned}&\sum_{0\le i\le 2} b_i(A)+\frac{1}{2}(1-\frac{1}{\xi})r^{-\beta(j_{n+1}(A)+1)}\theta(n+1)\\&\le \sum_{0\le i\le 2} b_i(C)+\frac{1}{4}(1-\frac{1}{\xi})r^{-\beta(j_n(C)+1)}\theta(n)\;. \qquad (2.157)\end{aligned}$$

As we shall show below in the last step of the proof, summation of these relations over $n \ge 0$ implies (2.149). Let us make a first comment about (2.157). When $j_{n+1}(A) > j_n(C)$, since $r^{-\beta}\theta(n+1) \le \theta(n)/2$ by (2.148) we have

$$r^{-\beta(j_{n+1}(A)+1)}\theta(n+1) \le \frac{1}{2}r^{-\beta(j_n(C)+1)}\theta(n)\;, \qquad (2.158)$$

and in that case (2.157) is satisfied as soon as $\sum_{0\le i\le 2} b_i(A) \le \sum_{0\le i\le 2} b_i(C)$. This is related to the idea, already made explicit in the proof of Theorem 2.6.1, that this case "does not matter". It will be harder to satisfy (2.157) when $j_{n+1}(A) = j_n(C)$.

Before we go into the details of the construction, and of the recursive definition of the numbers $b_j(C)$, we explain how this proof was found. It is difficult here to give a "big picture" why the approach works. We simply gather in each case the available information to make sensible definitions. Analysis of these definitions in the two main cases below will convince the reader that this is exactly how we have proceeded. Of course, when starting such an approach, it is difficult to know whether it will succeed, so we simply crossed our fingers and tried. The overall method seems powerful.

We now define the numbers $b_j(C)$ by induction over n. We start with

$$b_0(T) = b_1(T) = b_2(T) = F_0(T)\,.$$

For the induction step from n to $n+1$, let us first consider the case where, when applying the Decomposition Lemma, the set $A = A_\ell$ satisfies (2.150). We then define

$$b_0(A) = b_2(A) = b_1(C)\,,\; b_1(A) = \min(b_1(C), b_2(C))\;.$$

Relations (2.155) and (2.156) for A are obvious. To prove (2.152) for A, we write

$$\begin{aligned}F_{n+1}(A) &\le F_{n+1}(C\cap B(t_\ell, r^{-j-1}))\\&\le F_n(C\cap B(t_\ell, r^{-j-1})) \le b_1(C) = b_0(A)\,,\end{aligned}$$

using (2.153) for C. In a similar manner, we have, if $t \in A$, and since $j_{n+1}(A) = j+1$,

$$\begin{aligned}F_{n+1}(A\cap B(t, r^{-j_{n+1}(A)-1})) &\le F_{n+1}(C\cap B(t, r^{-j-2}))\\&\le F_n(C\cap B(t, r^{-j-2}))\\&\le \min(b_1(C), b_2(C)) = b_1(A)\;,\end{aligned}$$

and this proves (2.153) for A. Also, (2.154) for A follows from (2.152) for A since $b_2(A) = b_0(A)$.

To prove (2.157), we observe that

$$\sum_{0\le i\le 2} b_i(A) \le 2b_1(C) + b_2(C) \le \sum_{0\le i\le 2} b_i(C)\ , \tag{2.159}$$

since $b_1(C) \le b_0(C)$ by (2.155). We observe that, since $j_{n+1}(A) = j_n(C)+1$, (2.158) holds and combining with (2.159) this proves (2.157).

Next, we consider the case where $A = A_\ell$ satisfies (2.151). We define

$$b_0(A) = b_0(C)\ ;\ b_1(A) = b_1(C)\ ;\ b_2(A) = b_0(C) - \frac{1}{2} r^{-\beta(j+1)}\theta(n+1)\ .$$

It is obvious that A and $n+1$ in place of C and n satisfy the relations (2.125), (2.155) and (2.156). The relations (2.152) and (2.153) for A follow from the fact that similar relations hold for C rather than A, that $F_{n+1} \le F_n$, and that the functional F_{n+1} is increasing. Moreover (2.154) follows from (2.151) and (2.152).

To prove (2.157), we observe that by definition

$$\begin{aligned}
&\sum_{0\le i\le 2} b_i(A) + \frac{1}{2}(1-\frac{1}{\xi})r^{-\beta(j+1)}\theta(n+1) \\
&= 2b_0(C) + b_1(C) - \frac{1}{2\xi} r^{-\beta(j+1)}\theta(n+1) \\
&\le 2b_0(C) + b_1(C) - \frac{1}{2} r^{-\beta(j+1)}\theta(n)\ ,
\end{aligned} \tag{2.160}$$

using the regularity condition (2.148) on $\theta(n)$ in the last inequality. But (2.156) implies

$$b_0(C) \le b_2(C) + \frac{1}{2} r^{-\beta(j+1)}\theta(n)\ ,$$

so that (2.160) implies (2.157).

We have completed the construction, and we turn to the proof of (2.149). By (2.157), for any t in T, any $n \ge 0$, we have, setting $j_n(t) = j_n(A_n(t))$

$$\begin{aligned}
&\sum_{0\le i\le 2} b_i(A_{n+1}(t)) + \frac{1}{2}(1-\frac{1}{\xi})r^{-\beta(j_{n+1}(t)+1)}\theta(n+1) \\
&\le \sum_{0\le i\le 2} b_i(A_n(t)) + \frac{1}{4}(1-\frac{1}{\xi})r^{-\beta(j_n(t)+1)}\theta(n)\ .
\end{aligned}$$

Since $b_i(T) = F_0(T)$ and since $b_i(A) \ge 0$ by (2.152) to (2.154), summation of these relations for $0 \le n \le q$ implies

$$\frac{1}{2}(1-\frac{1}{\xi})\sum_{0\le n\le q} r^{-\beta(j_{n+1}(t)+1)}\theta(n+1) \tag{2.161}$$

$$\le 3F_0(T) + \frac{1}{4}(1-\frac{1}{\xi})\sum_{0\le n\le q} r^{-\beta(j_n(t)+1)}\theta(n) \tag{2.162}$$

and thus

$$\frac{1}{4}(1-\frac{1}{\xi})\sum_{0\le n\le q} r^{-\beta(j_n(t)+1)}\theta(n) \le 3F_0(T) + \frac{1}{4}(1-\frac{1}{\xi})r^{-\beta(j_0(T)+1)}\theta(0)\ .$$

By (2.125), we have $\Delta(A_n(t)) \le 2r^{-j_n(t)}$, and the choice of $j_0(T)$ implies $r^{-j_0(T)-1} \le \Delta(T)$ so that, since $\xi \le 2$,

$$\sum_{n\ge 0}\theta(n)\Delta^\beta(A_n(t)) \le \frac{L(2r)^\beta}{\xi-1}(F_0(T)+\Delta^\beta(T)\theta(0))\ . \qquad \square$$

Exercise 2.7.4. Write in complete detail the proof of Theorem 2.7.2 along the lines of the proof of Theorem 2.6.1.

To illustrate how the parameter τ in Theorem 2.7.2 may be used we give another proof of Theorem 2.3.1. Recall the definition 2.2.19 of $\gamma_\alpha(T,d)$.

Second proof of Theorem 2.3.1. We will use Theorem 2.7.2 with $r = 4$, $\beta = 1$ and $\tau = \tau' + 1$. For $n \ge 0$ and a subset A of U we define

$$F_n(A) = \sup_{t\in A}\sum_{k\ge n} 2^{k/\alpha} d(t, T_k)\ .$$

In order to check (2.147), consider $m = N_{n+\tau'+1}$, and assume that there exist points $t_1, \dots, t_m$ of U such that

$$1 \le \ell < \ell' \le m \Rightarrow d(t_\ell, t_{\ell'}) \ge a\ .$$

Consider then subsets $H_1, \dots, H_m$ of U with $H_\ell \subset B(t_\ell, a/4)$. By definition of F_{n+1}, given any $\epsilon > 0$, we can find $u_\ell \in H_\ell$ such that

$$\sum_{k\ge n+1} 2^{k/\alpha} d(u_\ell, T_k) \ge F_{n+1}(H_\ell) - \epsilon\ .$$

Since $d(t_\ell, t_{\ell'}) \ge a$ for $\ell \ne \ell'$, the open balls $B(t_\ell, a/2)$ are disjoint. Since there are $N_{n+\tau'+1}$ of them, whereas $\operatorname{card} T_n \le N_{n+\tau'}$, one of these balls cannot meet T_n. Thus there is $\ell \le m$ with $d(t_\ell, T_n) \ge a/2$. Since $u_\ell \in H_\ell \subset B(t_\ell, a/4)$, the inequality $d(u_\ell, T_n) \ge a/4$ holds, and

$$\begin{aligned}\sum_{k\ge n} 2^{k/\alpha} d(u_\ell, T_k) &\ge 2^{n/\alpha}\frac{a}{4} + \sum_{k\ge n+1} 2^{k/\alpha} d(u_\ell, T_k)\\ &\ge 2^{n/\alpha-2}a + F_{n+1}(H_\ell) - \epsilon\ .\end{aligned}$$

Since $u_\ell \in H_\ell$ this shows that

$$F_n\Big(\bigcup_{p\le m} H_p\Big) \ge 2^{n/\alpha-2}a + F_{n+1}(H_\ell) - \epsilon\,,$$

and since ϵ is arbitrary, this proves that (2.147) holds with $\theta(n+1) = 2^{n/\alpha-2}$. (Condition (2.148) holds only when $\alpha \geq 1$, which is the most interesting case. We leave to the reader to complete the case $\alpha < 1$ by using a different value of r.) We have $F_0(U) \leq S$, and since $d(t, T_0) \leq S$ for $t \in U$, and $\operatorname{card} T_0 = 1$, we have $\Delta(U) \leq 2S$. To finish the proof one simply applies Theorem 2.7.2 and Lemma 2.3.5. □

We now collect some simple facts, the proof of which will also serve as another (easy) application of Theorem 2.7.2.

Theorem 2.7.5. *(a) If U is a subset of T, then*

$$\gamma_\alpha(U,d) \leq \gamma_\alpha(T,d) \,.$$

(b) If $f : (T,d) \to (U,d')$ is onto and satisfies

$$\forall x, y \in T \,,\; d'(f(x), f(y)) \leq A d(x,y) \,,$$

for some constant A, then

$$\gamma_\alpha(U,d') \leq K(\alpha) A \gamma_\alpha(T,d) \,.$$

(c) We have

$$\gamma_\alpha(T,d) \leq K(\alpha) \sup \gamma_\alpha(F,d) \,, \tag{2.163}$$

where the supremum is taken over $F \subset T$ and F finite.

It seems plausible that with different methods than those used below one should be able to obtain (b) and (c) with $K(\alpha) = 1$, although there is little motivation to do this.

Proof. Part (a) is obvious. To prove (b) we consider an admissible sequence of partitions $\mathcal{A}_n$ with $\sup_t \sum_{n\geq 0} 2^{n/\alpha} \Delta(A_n(t), d) \leq 2\gamma_a(T,d)$. Consider then sets $T_n \subset T$ with $\operatorname{card} T_n \leq N_n$ and $\operatorname{card}(T_n \cap A) = 1$ for each $A \in \mathcal{A}_n$ so that $\sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} d(t, T_n) \leq 2\gamma_\alpha(T,d)$. We observe that $\sup_{s\in U} \sum_{n\geq 0} 2^{n/\alpha} d'(s, f(T_n)) \leq 2A\gamma_\alpha(T,d)$, and we apply Theorem 2.3.1.

To prove (c) we essentially repeat the argument in the proof of Theorem 2.3.14. We define

$$\gamma_{\alpha,n}(T,d) = \inf \sup_{t\in T} \sum_{k\geq n} 2^{k/\alpha} \Delta(A_k(t))$$

where the infimum is over all admissible sequences $(\mathcal{A}_k)$. We consider the functionals

$$F_n(A) = \sup \gamma_{\alpha,n}(G,d)$$

where the supremum is over all finite subsets G of A. We will use Theorem 2.7.2 with $\beta = 1$, $\theta(n+1) = 2^{n/\alpha-1}$, $\tau = 1$, and $r = 4$. (As in the proof of Theorem 2.3.1 this works only for $\alpha \geq 1$, and the case $\alpha < 1$ requires a different choice of r.) To prove (2.147), consider $m = N_{n+1}$ and consider points

$(t_\ell)_{\ell \le m}$ of T, with $d(t_\ell, t_{\ell'}) \ge a$ if $\ell \ne \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/4)$ and $c < \min_{\ell \le m} F_{n+1}(H_\ell)$. For $\ell \le m$, consider finite sets $G_\ell \subset H_\ell$ with $\gamma_{\alpha,n+1}(G_\ell, d) > c$, and $G = \bigcup_{\ell \le m} G_\ell$. Consider an admissible sequence $(\mathcal{A}_n)$ of G, and

$$I = \{\ell \le m \, ; \, \exists A \in \mathcal{A}_n \, , \, A \subset G_\ell\}$$

so that, since the sets G_ℓ for $\ell \le m$ are disjoint, we have $\operatorname{card} I \le N_n$, and thus there exists $\ell \le m$ with $\ell \notin I$. Then for $t \in G_\ell$, we have $A_n(t) \not\subset G_\ell$, so $A_n(t)$ meets a ball $B(t_{\ell'}, a/4)$ for $\ell \ne \ell'$, and hence $\Delta(A_n(t)) \ge a/2$; so that

$$\sum_{k \ge n} 2^{k/\alpha} \Delta(A_k(t)) \ge \frac{a}{2} 2^{n/\alpha} + \sum_{k \ge n+1} 2^{k/\alpha} \Delta(A_k(t) \cap G_\ell)$$

and hence

$$\sup_{t \in G_\ell} \sum_{k \ge n} 2^{k/\alpha} \Delta(A_k(t)) \ge a 2^{n/\alpha - 1} + \gamma_{\alpha,n+1}(G_\ell, d) \, .$$

Since the admissible sequence $(\mathcal{A}_n)$ is arbitrary, we have shown that

$$\gamma_{\alpha,n}(G, d) \ge a 2^{n/\alpha - 1} + c$$

and thus

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge a 2^{n/\alpha - 1} + \min_{\ell \le m} F_{n+1}(H_\ell) \, ,$$

which is (2.147). Finally, we have $F_0(T) = \sup \gamma_\alpha(G, d)$, where the supremum is over all finite subsets G of T, and since $\Delta(G) \le \gamma_\alpha(G, d)$, we have that $\Delta(T) \le F_0(T)$ and we conclude by Lemma 2.3.5 and Theorem 2.7.2. □

There are many possible variations about the scheme of proof of Theorem 2.7.2. We end this section with such a version. This specialized result will be used only in Section 16.8, and its proof could be omitted at first reading.

There are natural situations, where, in order to be able to prove (2.147), we need to know that $H_\ell \subset B(t_\ell, \eta a)$ where η is very small. In order to apply Theorem 2.7.2, we have to take $r \ge 1/\eta$, which (when $\beta = 1$) produces a loss of a factor $1/\eta$. We will give a simple modification of Theorem 2.7.2 that produces only the loss of a factor $\log(1/\eta)$.

For simplicity, we assume $r = 4, \beta = 1, \theta(n) = 2^{n/2}$ and $\tau = 1$. We consider an integer $s \ge 2$.

Theorem 2.7.6. *Assume that the hypotheses of Theorem 2.7.2 are modified as follows. Whenever $t_1, \ldots, t_m$ are as in (2.146), and whenever $H_\ell \subset B(t_\ell, a4^{-s})$, we have*

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge a 2^{(n+1)/2} + \min_{\ell \le m} F_{n+s}(H_\ell) \, . \tag{2.164}$$

Then there exists an increasing sequence of partitions $(\mathcal{A}_n)$ *in* T *such that* $\operatorname{card}\mathcal{A}_n \le N_{n+1}$ *and*

$$\sup_{t\in T}\sum_{n\ge 0} 2^{n/2}\Delta(A_n(t)) \le Ls(F_0(T)+\Delta(T))\,.$$

Now the reader observes that in the last term of (2.164) we have $F_{n+s}(H_\ell)$ rather than the larger quantity $F_{n+1}(H_\ell)$. This will be essential in Section 16.8. It is of course unimportant in the first term of the right-hand side to have the exponent $n+1$ rather than n. We use $n+1$ to mirror (2.73).

Proof. We closely follow the proof of Theorem 2.7.2. For clarity we assume that T is finite. First, we copy the proof of the Decomposition Lemma (Lemma 2.6.2), and rather than obtaining (2.129) (which occurs exactly for $\ell = m$) we now have

$$\frac{1}{2}4^{-j-1}2^{(n+1)/2} + \sup_{t\in A_m} F_{n+s}(A_m\cap B(t,4^{-j-1-s})) \le F_n(C)\,. \tag{2.165}$$

Together with each set C in $\mathcal{A}_n$, we construct numbers $b_i(C)\ge 0$ for $0\le i\le s+1$, such that

$$\forall i\,,\ 1\le i\le s+1\,,\ b_i(C)\le b_0(C)$$

$$b_0(C)\ge b_{s+1}(C)\ge b_0(C)-\frac{1}{2}4^{-j_n(C)-1}2^{n/2}$$

$$F_n(C)\le b_0(C)$$

$$\forall i\,,\ 1\le i\le s+1\,,\ \forall t\in C\,,\ F_{n+i-1}(C\cap B(t,4^{-j_n(C)-i}))\le b_i(C)\,. \tag{2.166}$$

The reader observes that (2.166) is not a straightforward extension of (2.154), since it involves F_{n+i-1} rather than the larger quantity F_n. We set $b_i(T) = F_0(T)$ for $0\le i\le s+1$. For the induction from n to $n+1$ we consider one of the pieces A of the partition of $C\in\mathcal{A}_n$ and $j=j_n(C)$. If $A=A_m$, we set

$$\forall i\,,\ 0\le i\le s\,,\ b_i(A)=b_i(C)$$

$$b_{s+1}(A)=b_0(A)-\frac{1}{2}4^{-j-1}2^{(n+1)/2}\,. \tag{2.167}$$

Since $j_{n+1}(A)=j_n(C)$, for $i=s+1$ condition (2.166) for A follows from (2.165) and (2.167) since $F_n(C)\le b_0(C)=b_0(A)$. For $i\le s$, condition (2.166) for A follows from the same condition for C since $F_{n+i}\le F_{n+i-1}$.

If $A=A_\ell$ with $\ell<m$ we then set

$$b_{s+1}(A)=b_1(C)\ ;\ \forall i\le s\,,\ b_i(A)=\min(b_{i+1}(C),b_1(C))\,.$$

Since $j_{n+1}(A)=j_n(A)+1$, condition (2.166) for $i=s+1$ and A follows from the same condition for C and $i=1$, while condition (2.166) for $i\le s$ and A

follows form the same condition for C and $i+1$. Exactly as previously we show in both cases that

$$\sum_{0\le i\le s+1} b_i(A) + \frac{1}{2}(1-\frac{1}{\sqrt{2}})4^{-j_{n+1}(A)-1}2^{(n+1)/2}$$
$$\le \sum_{0\le i\le s+1} b_i(C) + \frac{1}{4}(1-\frac{1}{\sqrt{2}})4^{-j_n(C)-1}2^{n/2} + \epsilon_{n+1}$$

and we finish the proof in the same manner. □

2.8 Notes and Comments

It seems necessary to say a few words about the history of Gaussian processes. I have heard people saying that the problem of characterizing continuity and boundedness of Gaussian processes goes back (at least implicitly) to Kolmogorov.

The understanding of Gaussian processes was long delayed by the fact that in the most immediate examples the index set is a subset of $\mathbb{R}$ or $\mathbb{R}^n$ and that the temptation to use the special structure of this index set is nearly irresistible. Probably the single most important conceptual progress about Gaussian processes is the realization, in the late sixties, that the boundedness of a (centered) Gaussian process is determined by the structure of the metric space (T,d), where d is the usual distance $d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2}$. It is of course difficult now to realize what a tremendous jump in understanding this was, since this seems so obvious *a posteriori*.

In 1967, R. Dudley obtained the inequality (2.36), which however cannot be reversed in general. (Actually, as R. Dudley pointed out repeatedly, he did not state (2.36). Nonetheless since he performed all the essential steps it seems appropriate to call (2.36) Dudley's bound. It simply does not seem worth the effort to find who deserves the very marginal credit of having stated (2.36) first.) A few years later, X. Fernique proved that in the "stationary case" Dudley's inequality can be reversed [3], i.e. he proved in that case the lower bound of Theorem 2.4.1. This result is historically important, because it was central to the work of Marcus and Pisier [7], [8] who build on it to solve all the classical problems on random Fourier series. A part of their results was presented in Section 3.2. Interestingly, now that the right approach has been found, the proof of Fernique's result is not really easier than that of Theorem 2.4.1.

Another major contribution of Fernique (building on earlier ideas of C. Preston) was an improvement of Dudley's bound based on a new tool called majorizing measures. Fernique conjectured that his inequality was essentially optimal. Gilles Pisier suggested in 1983 that I should work on this conjecture. In my first attempt I proved quite fast that Fernique's conjecture held in the case where the metric space (T, d) is ultrametric. I was quite disappointed to learn that Fernique had already done this, so I was discouraged for a while. In the second attempt, I tried to decide whether a majorizing measure existed on ellipsoids. I had the hope that some simple density with respect to the volume measure would work. It was difficult to form any intuition, and I really struggled in the dark for months. At some point I decided not to use the volume measure, but rather a combination of suitable point masses, and easily found a direct construction of the majorizing measure on ellipsoids. This of course made it quite believable that Fernique's conjecture was true, but I still tried to disprove it. At some point I realized that I did not understand why a direct approach to prove Fernique's conjecture using a partition scheme should fail, while this understanding should be useful to construct a counter example. Once I tried this direct approach, it was only a matter of a few days to prove Fernique's conjecture. Gilles Pisier made two comments about this discovery. The first one was "you are lucky", by which he of course meant that I was lucky that Fernique's conjecture was true, since a counter example would have been of limited interest. I am grateful to this day for his second comment: "I wish I had proved this myself, but I am very glad you did it."

Fernique's concept of majorizing measures is very difficult to grasp at the beginning, and was consequently dismissed by the main body of probabilists as a mere curiosity. (I must admit that I myself did find it very difficult to understand.) However, in 2000, while discussing one of the open problems of this book with K. Ball (be he blessed for his interest in it!) I discovered that one could replace majorizing measures by the totally natural variation on the usual chaining arguments that was presented here. That this was not discovered much earlier is a striking illustration of the inefficiency of the human brain (and of mine in particular).

Some readers wondered why I do not mention Slepian's lemma. Of course this omission is done on purpose and must be explained. Slepian's lemma is very specific to Gaussian processes, and focusing on it seems a good way to guarantee that one will never move beyond these. One notable progress made by the author was to discover the scheme of proof of Proposition 2.4.9 that dispenses with Slepian's lemma, and that we shall use in many situations. Comparison results such as Slepian's lemma are not at the root of results such as the majorizing measure theorem, but rather are (at least qualitatively) a consequence of them. Indeed, if two centered Gaussian processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ satisfy $\mathsf{E}(X_s - X_t)^2 \le \mathsf{E}(Y_s - Y_t)^2$ whenever $s, t \in T$, then (2.80) implies $\mathsf{E}\sup_{t\in T} X_t \le L\mathsf{E}\sup_{t\in T} Y_t$. (Slepian's lemma asserts that this inequality holds with constant $L = 1$.)

It may happen in the construction of Lemma 2.6.2 that $C = A_1$. Thus it may happen in the construction of Theorem 2.6.1 that a same set A belongs both to $\mathcal{A}_n$ and $\mathcal{A}_{n+1}$. When this is the case, the construction shows that one has $j_{n+1}(A) = j_n(A) + 1$. It is therefore incorrect, as was done in the first edition, to use in the construction a number $j(A)$ depending only on A.

I am grateful to J. Lehec for having pointed out this mistake. Fortunately, the only change required in the proofs is to add the proper indexes to the quantities of the type $j(C)$.

In [9] the author presented a particularly simple proof of (an equivalent form of) Theorem 2.4.1. It is also based on a partitioning scheme. For the readers who are familiar with that proof, it might be useful to compare the partitioning scheme of [9] with the partitioning scheme presented here. We shall show that these schemes "produce the same pieces of T", the difference being that these are not gathered to form partitions in the same manner. Consider a metric space (T,d) and a functional $F(H)$ on T. Assume that for a certain number r it satisfies the following growth condition. Given $m \geq 2$, $k \in \mathbb{Z}$ and points $t_1, \ldots, t_m$ of T, with $d(t_\ell, t_{\ell'}) \geq r^{-k}$, and subsets H_ℓ of $B(t_\ell, r^{-k-1})$ then

$$F(\cup_{j \leq m} H_\ell) \geq r^{-k}\sqrt{\log m} + \min_{\ell \leq m} F(H_\ell) \,. \tag{2.168}$$

Let us then perform the construction of Theorem 2.4.1 for the functionals $F_n = F$. Let us define $j_0(T)$ as in (2.127) and we partition T using the Decomposition Lemma 2.6.2. That is, for $j = j_{0,T}$ we inductively construct sets D_ℓ, and we pick t_ℓ in D_ℓ such that

$$F(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) = \sup_{t \in D_\ell} F(D_\ell \cap B(t, r^{-j-2})) \,,$$

we set $A_{\ell+1} = D_\ell \cap B(t_{\ell+1}, r^{-j-1})$ and $D_{\ell+1} = D_\ell \setminus A_{\ell+1}$. Assume that the construction continues until we construct a non-empty last piece $C = A_m = D_{m-1}$, where $m = N_1$. Let us get investigate what happens to this set C at the next stage of the construction. Recall that we have defined $j_1(A) = j$. First, we find u_1 in C with

$$F(C \cap B(u_1, r^{-j-2})) = \sup_{t \in C} F(C \cap B(t, r^{-j-2})) \,,$$

we set $A_1^* = C \cap B(t, r^{-j-1})$ and $D_1^* = C \setminus A_1^*$, and we continue in this manner. The point is that this construction is the exact continuation of the construction by which we obtained A_1, A_2, etc. In consequence, if we consider the sets $A_1, \ldots, A_{m-1}$ together with the sets $A_1^*, \ldots, A_{m^*-1}^*$, where $m^* = N_2$, these pieces are simply obtained by continuing the exhaustion procedure by which we constructed $A_1, \ldots, A_{m-1}$ until $m + m^* - 2$ (etc.). Therefore, as in [9] we construct all the pieces that are obtained by pursuing this exhaustion procedure until the entire space is exhausted, and the same is true at every level of the construction.

The generic chaining as presented here (and the use of a scheme where the functional F might depend on the stage n of the construction) offers at times considerable clarification over the previous approaches. This justifies presenting a proof of Theorem 2.4.1 which is not the simplest we know.

References

1. Artstein, S.: The change in the diameter of a convex body under a random sign-projection. In: Geometric Aspects of Functional Analysis. Springer Lecture Notes in Math., vol. 1850, pp. 31–39 (2004)
2. Dudley, R.M.: The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J. Funct. Anal. **1**, 290–330 (1967)
3. Fernique, X.: Regularité des trajectoires des fonctions aléatoires gaussiennes. In: Ecole d'été de Probabilités de Saint-Flour, IV-1974. Lecture Notes in Math., vol. 480, pp. 1–96. Springer, Berlin (1975)
4. Ledoux, M.: The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001). x+181 pp. ISBN 0-8218-2864-9
5. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
6. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces, II. Function Spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 97. Springer, Berlin (1979). x+243 pp. ISBN: 3-540-08888-1
7. Marcus, M.B., Pisier, G.: Random Fourier Series with Applications to Harmonic Analysis. Annals of Mathematics Studies, vol. 101. Princeton University Press, Princeton; University of Tokyo Press, Tokyo (1981). v+151 pp. ISBN: 0-691-08289-8; 0-691-08292-8
8. Marcus, M.B., Pisier, G.: Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. Acta Math. **152**(3–4), 245–301 (1984)
9. Talagrand, M.: A simple proof of the majorizing measure theorem. Geom. Funct. Anal. **2**, 119–125 (1992)

Springer

http://www.springer.com/978-3-642-54074-5

Upper and Lower Bounds for Stochastic Processes
Modern Methods and Classical Problems
Talagrand, M.
2014, XV, 626 p., Hardcover
ISBN: 978-3-642-54074-5