Michel TALAGRAND

# Upper and Lower Bounds for Stochastic Processes

## Modern Methods and Classical Problems

April 11, 2018

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

# Contents

# 0. Introduction

What is the maximum level a certain river is likely to reach over the next 25 years? What is the likely magnitude of the strongest earthquake to occur during the life of a planned nuclear plant? These fundamental practical questions have motivated (arguably also fundamental) mathematics, some of which are the object of this book. The value $X_t$ of the quantity of interest at time $t$ is modeled by a random variable. What can be said about the maximum value of $X_t$ over a certain range of $t$? How can we guarantee that, with probability close to one, this maximum will not exceed a given threshold?

A collection of random variables $(X_t)_{t\in T}$, where $t$ belongs to a certain index set $T$, is called a stochastic process, and the topic of this book is the study of the supremum of certain stochastic processes, and more precisely the search of upper and lower bounds for these suprema. The key word of the book is

INEQUALITIES.

The "classical theory of processes" deals mostly with the case where $T$ is a subset of the real line or of $\mathbb{R}^n$. We do not focus on that case, and the book does not really expand on the most basic and robust results which are important in this situation. Our most important index sets are "high-dimensional": the large sets of data which are currently the focus of so much attention consist of data which usually depends on many parameters. Our specific goal is to demonstrate the impact and the range of modern abstract methods, in particular through their treatment of several classical questions which are not accessible to "classical methods".

A. Kolmogorov invented the most important idea to bound stochastic processes: chaining. This wonderfully efficient method answers with little effort a number of basic questions, but fails to provide a complete understanding, even in natural situations. This is best discussed in the case of Gaussian processes, where the family $(X_t)_{t\in T}$ consists of centered jointly Gaussian random variables (r.v.s). These are arguably the most important of all. A Gaussian process defines in a canonical manner a distance $d$ on its index set $T$ by the formula

$$d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2} \,. \tag{0.1}$$

Probably the single most important conceptual progress about Gaussian processes was the gradual realization that the metric space $(T,d)$ is the key object to understand them, even if $T$ happens to be an interval of the real line. This led R. Dudley to develop in 1967 an abstract version of Kolmgorov's chaining argument adapted to this situation. The resulting very efficient bound for Gaussian processes is unfortunately not always tight. Roughly speaking, "there sometimes remains a parasitic logarithmic factor in the estimates".

The discovery around 1985 (by X. Fernique and the author) of a precise (and in a sense, *exact*) relationship between the "size" of a Gaussian process and the "size" of this metric space provided the missing understanding in the case of these processes. Attempts to extend this result to other processes spanned a body of work which forms the core of this book.

A significant part of the book is devoted to situations where skill is required to "remove the last parasitic logarithm in the estimates." These situations occur with unexpected frequency in all kinds of problems. A particularly striking example is as follows. Consider $n^2$ independent uniform random points $(X_i)_{i \le n^2}$ which are uniformly distributed in the unit square $[0,1]^2$. How far a typical sample is from being very uniformly spread on the unit square? To measure this we construct a one-to-one map $\pi$ from $\{1, \ldots, n^2\}$ to the vertices $v_1, \ldots, v_{n^2}$ of a uniform $n \times n$ grid in the unit square. If we try to minimize the *average* distance between $X_i$ and $v_{\pi(i)}$ we can do as well as about $\sqrt{\log n}/n$ but no better. If we try to minimize the *maximum* distance between $X_i$ and $v_{\pi(i)}$, we can do as well as about $(\log n)^{3/4}/n$ but no better. The factor $1/n$ is just due to scaling, but the fractional powers of $\log n$ require a surprising amount of work.

The book is largely self-contained, but it mostly deals with rather subtle questions such as the previous one. It also devotes considerable energy to the problem of finding *lower* bounds for certain processes, a topic considerably more difficult and less developed than the search for upper bounds. Even though some of the main ideas of at least Chapter 2 could (and should!) be taught at an elementary level, this is an advanced text.

This book is a sense a continuation of the monograph [61], or at least of part of it. I made no attempt to cover again all the relevant material of [61], but familiarity with [61] is certainly not a prerequisite, and maybe not even helpful. The way certain results are presented there is arguably obsolete, and, more importantly, many of the problems considered in [61] (in particular limit theorems) are no longer the focus of much interest.

One of my main goals is to communicate as much as possible of my experience from working on stochastic processes, and I have covered most of my results in this area. A number of these results were proved many years ago. I still like them, but some seem to be waiting for their first reader. The odds of these results meeting this first reader while staying buried in the original papers seemed nil, but might increase in the present book form. In order to present a somewhat coherent body of work I have also included recent results by others in the same general direction. I find these results deep and very beautiful. They are sometimes difficult to access for the non-specialist. Explaining them here in a unified and often simplified presentation could serve a useful purpose. Still, the choice of topics is highly personal and does not represent a systematic effort to cover all the important directions. I can only hope that the book contains enough state-of-art knowledge about sufficiently many fundamental questions to be useful.

A number of seemingly important questions remain open, and one of my main goals is to popularize these. Opinions differ as to what constitutes a really important problem, but I like those I explain here because they deal with fundamental structures. Several of them were raised a generation ago in [61], but have seen little progress or even attention since. These problems might be challenging. At least, I tried my best to make progress on them.

This book had a previous edition [145]. The change between the two editions are not only cosmetic or pedagogical, and the degree of improvement in the mathematics themselves is almost embarrassing at times, resulting in a large decrease of size. Only a limited quantity of material of secondary importance was removed, and the reduction of well over 100 pages is mostly due to better proofs. Mathematics is indeed a game of iterations. Part of the improvement was permitted by a better understanding of the consequences of the landmark result of Bednorz and Latała [23] (whose proof is one of our main goals). This landmark of modern probability seems bound to have a considerable influence. As I painfully experienced how slowly my understanding develops, I have been very cautious in adding new material.

I would like to express my infinite gratitude to Shahar Mendelson. While he was donating his time to help another of my projects, it became clear through our interactions that, while I had produced great efforts towards the quality of the mathematics contained in my books, I certainly had not put enough efforts in the exposition of this material. I concluded that there should be real room for improvement in the text of [145], and this is why I started to revise it. While tying to explain better the material to others, I ended up understanding it much better myself!

# 1. What is this Book About?

## 1.1 Philosophy and Style

This short chapter describes the philosophy underlying this book, and some of its highlights. This description, using words rather than formulas, is necessarily imprecise, and is only intended to provide some insight into our point of view.

The practitioner of stochastic processes is likely to be struggling at any given time with his favorite model of the moment, a model which typically involves a rich and complicated structure. There is a near infinite supply of such models. The importance with which we view any one of them is likely to vary over time.

The first advice the author received from his advisor Gustave Choquet was as follows: Always consider a problem under the minimum structure in which it makes sense. This advice will probably be as fruitful in the future as it has been in the past, and it has strongly influenced this work. By following it, one is naturally led to study problems with a kind of minimal and intrinsic structure. Not so many structures are really basic, and one may hope that these will remain of interest for a very long time. This book is devoted to the study of such structures.

The feeling, real or imaginary, that one is studying objects of intrinsic importance is enjoyable, but the success of the approach of studying "minimal structures" has ultimately to be judged by its results. As we shall demonstrate, the tools arising from this approach provide the final words to a number of classical problems.

Some readers may be disturbed to see that certain standard considerations are given little or no attention. You will find rather little about "convergence" here, at least explicitly. There are no apparent $\sigma$-algebras, and measurablity is hardly mentioned at all. We prove inequalities, and for this one may basically pretend that every index set is finite. Thus we shall never define "essential supremum" or "separable processes", etc. The missing "details" belong to pre-1950 mathematics, and it would serve no purpose to rewrite them.

## 1.2 Basic Chaining: The Kolmogorov Conditions

Kolmogorov invented chaining, the main tool of this book. He stated the "Kolmogorov conditions", which robustly ensure the good behavior of a stochastic process on $\mathbb{R}^m$. These conditions are studied in any advanced probability course. If you have taken such a course, this section will refresh your mind about these conditions, and the next few sections will present the natural generalization of the chaining method in an abstract metric space, preparing you for the final version of the method which is presented in Chapter 2. Learning in detail about these historical developments now makes sense only if you have already heard of them. For this reason, the material up to Section 1.5 included is directed towards a reader who has already some fluency with probability theory. If, on the other hand, you have never heard of these things, you should start directly with Chapter 2, which *is written at a far greater level of detail and assumes minimal familiarity with even basic probability theory.*

A stochastic process a collection of random variables (r.v.s) $(X_t)_{t\in T}$. The fundamental idea of chaining is to replace the index set $T$ by a sequence of finite approximations $T_n$, and to study the r.v.s $X_t$ through successive approximations $X_{\pi_n(t)}$ where $\pi_n(t) \in T_n$. As a first approximation let us consider a single point $t_0$ so $T_0 = \{t_0\}$. The fundamental relation is then

$$X_t - X_{t_0} = \sum_{n\geq 1} (X_{\pi_n(t)} - X_{\pi_{n-1}(t)}) \;, \tag{1.1}$$

where in all cases we need the sum on the right will be finite. This relation gives its name to the method, the chain of approximations $\pi_n(t)$ links $t_0$ and $t$. To control the differences $X_t - X_{t_0}$ it suffices then to control all the differences $|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}|$.

Let us apply this method to processes that satisfy the so-called Kolmogorov conditions: that is processes $(X_t)_{t\in T}$ where $T = [0,1]^m$, for which

$$\forall\, s,t \in [0,1]^m \;,\; \mathsf{E}|X_s - X_t|^p \leq d(s,t)^\alpha \;. \tag{1.2}$$

where $d(s,t)$ denotes the Euclidean distance and $p > 0, \alpha > m$. Let us study the continuity of such processes. The most obvious candidate for approximating set $T_n$ is the grid $G_n$, the set of points $x$ in $[0,1]^m$ such that the coordinates of $2^n x$ are integers $\neq 0$. Thus $\operatorname{card} G_n = 2^{nm}$. It is completely natural to choose $\pi_n(u) \in G_n$ as close to $u$ as possible, so that $d(u,\pi_n(u)) \leq \sqrt{m}2^{-n}$, and $d(\pi_n(u),\pi_{n-1}(u)) \leq 3\sqrt{m}2^{-n}$.

Forn $n \geq 1$ let us then define

$$U_n = \{(s,t)\;;\; s \in G_n \;,\; t \in G_n \;,\; d(s,t) \leq 3\sqrt{m}2^{-n}\} \;, \tag{1.3}$$

so that we have the crucial property

$$\operatorname{card} U_n \leq K(m)2^{nm} \;, \tag{1.4}$$

where $K(m)$ denotes a number depending only on $m$, which need not be the same on each occurrence. Consider then the r.v.

$$Y_n = \max\{|X_s - X_t| \; ; \; (s,t) \in U_n\} \; , \tag{1.5}$$

so that (and since $G_{n-1} \subset G_n$) for each $u$,

$$|X_{\pi_n(u)} - X_{\pi_{n-1}(u)}| \le Y_n \; .$$

To avoid having to explain what is "a version of the process", and since we care only about inequalities, we will consider only the r.v.s $X_t$ for $t \in G =: \bigcup_{n \ge 0} G_n$. We first claim that

$$\sup_{s,t \in G; d(s,t) \le 2^{-k}} |X_s - X_t| \le 3 \sum_{n \ge k} Y_n \; . \tag{1.6}$$

To prove this consider $m \ge k$ such that $s, t \in G_m$, so that $s = \pi_m(s)$ and $t = \pi_m(t)$. Since $d(s,t) \le 2^{-k}$, we have

$$d(\pi_k(s), \pi_k(t)) \le d(s, \pi_k(s)) + d(s,t) + d(t, \pi_k(t)) \le 3\sqrt{m} 2^{-k} \; ,$$

so that $(\pi_k(s), \pi_k(t)) \in U_k$ and

$$|X_{\pi_k(s)} - X_{\pi_k(t)}| \le Y_k \; .$$

To obtain (1.6) we then use the previous inequalities and the identities

$$X_s - X_t = X_s - X_{\pi_k(s)} + X_{\pi_k(s)} - X_{\pi_k(t)} + X_{\pi_k(t)} - X_t \; ,$$

and, for $u \in \{s,t\}$,

$$X_u - X_{\pi_k(u)} = X_{\pi_m(u)} - X_{\pi_k(u)} = \sum_{k \le n < m} X_{\pi_{n+1}(u)} - X_{\pi_n(u)} \; .$$

Let us now draw some consequences of (1.6). For a finite family of numbers $V_i \ge 0$, we have

$$(\max_i V_i)^p \le \sum_i V_i^p \; , \tag{1.7}$$

and thus

$$\mathsf{E} Y_n^p \le \mathsf{E} \sum_{(s,t) \in U_n} |X_s - X_t|^p \le K(m,\alpha) 2^{n(m-\alpha)} \; ,$$

since $\mathsf{E}|X_s - X_t|^p \le K(m,\alpha) 2^{-n\alpha}$ for $(s,t) \in U_n$ and using (1.4). To proceed one needs to distinguish whether or not $p \ge 1$. For specificity we assume $p \ge 1$. Since, as we just proved, $\|Y_n\|_p := (\mathsf{E}|Y_n|^p)^{1/p} \le K(m,p) 2^{n(m-\alpha)/p}$, the triangle inequality in $L^p$ yields

$$\Big\| \sum_{n \ge k} Y_n \Big\|_p \le K(m,p,\alpha) 2^{k(m-\alpha)/p} \; . \tag{1.8}$$

Combining with (1.6) we then obtain

$$\Big\| \sup_{s,t\in G; d(s,t)\le 2^{-k}} |X_s - X_t| \Big\|_p \le K(m,p,\alpha) 2^{k(m-\alpha)/p} \,, \tag{1.9}$$

a sharp inequality from which it is then simple to prove (with some loss of sharpness) results such as the fact that for $0 < \beta < \alpha - m$ one has

$$\mathsf{E} \sup_{s,t\in G} \frac{|X_s - X_t|^p}{d(s,t)^\beta} < \infty \,. \tag{1.10}$$

Thus, chaining not only proves that the process $(X_t)_{t\in T}$ has a continuous version, it also provides the very good estimate (1.9). One reason for which everything is so easy in this case is that the size of the terms $X_{\pi_{n+1}(u)} - X_{\pi_n(u)}$ decreases like a geometric series.

## 1.3 More Chaining in $\mathbb{R}^m$

One may also consider conditions more general than (1.2), for example

$$\forall n \ge 0 \,,\ \forall s,t \in T \,,\ d(s,t) \le 3\sqrt{m} 2^{-n} \Rightarrow \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{c_n}\Big) \le d_n \,, \tag{1.11}$$

where $\varphi$ is a convex function $\ge 0$ with $\varphi(0) = 0$, and $c_n, d_n$ are numbers. The factor $3\sqrt{m}$ is to simplify the statement of the forthcoming inequality (1.15) and is not important. Equivalently, one may consider conditions such as

$$\forall s,t \in T \,,\ \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{\psi(d(s,t))}\Big) \le \theta(d(s,t)) \,. \tag{1.12}$$

where $\psi$ and $\theta$ are functions. We follow exactly the same method as previously, but instead of (1.7) we use now that for r.v.s $V_i \ge 0$ we have $\varphi(\max_i V_i) \le \sum_i \varphi(V_i)$, so that

$$\varphi(\mathsf{E} \max_i V_i) \le \mathsf{E}\varphi(\max_i V_i) \le \sum_i \mathsf{E}\varphi(V_i)$$

and hence

$$\mathsf{E} \max_i V_i \le \varphi^{-1}\Big(\sum_i \mathsf{E}\varphi(V_i)\Big) \,. \tag{1.13}$$

Therefore the r.v. $Y_n$ of (1.5) satisfies

$$\mathsf{E} Y_n \le c_n \varphi^{-1}(K(m) 2^{nm} d_n) \,, \tag{1.14}$$

and combining with (1.6),

$$\mathsf{E} \sup_{s,t\in G, d(s,t)\le 2^{-k}} |X_s - X_t| \le 3\sum_{n\ge k} c_n \varphi^{-1}(K(m)2^{nm} d_n) \ . \tag{1.15}$$

The series in (1.15) has no reason to converge like a geometric series, so we already are being more sophisticated than in the case of the Kolmogorov conditions.[1]

## 1.4 Chaining in a Metric Space: Dudley's Bound

Suppose now that we want to study the uniform convergence on $[0,1]$ of a random Fourier series $X_t = \sum_{k\ge 1} a_k g_k \cos(2\pi i k t)$ where $(g_k)$ are independent standard Gaussian r.v.s. It took a very long time to understand that the fruitful way to attack the question is to *forget* about the natural distance on $[0,1]$ and rather to consider the distance $d$ given by

$$d(s,t)^2 = \mathsf{E}(X_s - X_t)^2 = \sum_k a_k^2 (\cos(2i\pi ks) - \cos(2i\pi kt))^2 \ . \tag{1.16}$$

More generally one is lead to consider Gaussian processes indexed by an abstract set $T$.[2] We say that $(X_t)_{t\in T}$ is a Gaussian process when the family $(X_t)_{t\in T}$ is jointly Gaussian centered. Then, just as in (1.16), the process induces a canonical distance $d$ on $T$ given by $d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2}$. Starting with the next chapter we will control the r.v.s $|X_s - X_t|$ through their tail properties, and to avoid repetition, we use a different condition, which, in the case of Gaussian processes at least, is just another way to present the same situation.[3] We impose the condition

$$\forall\, s,t \in T \ , \ \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{d(s,t)}\Big) \le 1 \ , \tag{1.17}$$

where $\varphi$ is convex function with $\varphi(0) = 0, \varphi \ge 0$. Until much later, we need only the choice $\varphi(x) = \exp(x^2/4) - 1$.[4]

In the absence of further structure on our metric space, how do we choose the approximating sets $T_n$? Thinking back to the Kolmogorov conditions it is very natural to introduce the following definition.

[1] In the left-hand side of (1.15) we would like to do better than controlling the expectation, but one really needs some regularity of the function $\varphi$ for this. It suffices here to say that when $\varphi(x) = |x|^p$ for $p \ge 1$ we may replace the expectation by the norm of $L^p$, proceeding exactly as we did in the case of the Kolmogorov conditions.

[2] Let us stress the point. Even though the index set is a subset of $\mathbb{R}^m$ we have *no chance* to really understand the process unless we forget this irrelevant structure.

[3] If you find the presentation too abstract here, you may like to go directly to Chapter 2.

[4] The factor $1/4$ in the exponential is simply to ensure that (1.17) is satisfied by a Gaussian process for the canonical distance, because if $g$ is a standard Gaussian r.v. then $\mathsf{E}\exp(g^2/4) \le 2$.

**Definition 1.4.1.** *For $\epsilon > 0$ the covering number $N(T,d,\epsilon)$ of a metric space $(T,d)$ is the smallest integer $N$ such that $T$ can be covered by $N$ balls of radius $\epsilon$.*

Equivalently, $N(T,d,\varepsilon)$ is the smallest number $N$ such that there exists a set $V \subset T$ with $\operatorname{card} V \le N$ and such that each point of $T$ is within distance $\epsilon$ of $V$.

Let us denote by $\Delta(T) = \sup_{s,t \in T} d(s,t)$ the diameter of $T$, and observe that $N(T,d,\Delta(T)) = 1$. We construct our approximating points $T_n$ as follows. Consider the largest integer $n_0$ with $\Delta(T) \le 2^{-n_0}$. For $n \ge n_0$ consider a set $T_n \subset T$ with $\operatorname{card} T_n = N(T,d,2^{-n})$ such that each point of $T$ is within distance $2^{-n}$ of a point of $T_n$.

We then perform the chaining as in the case of the Kolmogorov conditions, using for $\pi_n(t)$ a point in $T_n$ with $d(x,\pi_n(x)) \le 2^{-n}$. Consider

$$U_n = \{(s,t)\ ;\ s \in T_n\ ,\ t \in T_{n-1}\ ,\ d(s,t) \le 3 \cdot 2^{-n}\}\ ,$$

so that

$$\operatorname{card} U_n \le \operatorname{card} T_n \operatorname{card} T_{n-1} \le \operatorname{card} T_n^2 = N(T,d,2^{-n})^2\ .$$

This crude bound in hard to improve in general and should be compared to (1.4). Using (1.13) the r.v.

$$Y_n = \max\{|X_s - X_t|\ ;\ (s,t) \in U_n\}$$

satisfies

$$\mathsf{E} Y_n \le 3 \cdot 2^{-n} \varphi^{-1}(N(T,d,2^{-n-1})^2)\ ,$$

and exactly as in the case of the Kolmogorov conditions we obtain

$$\mathsf{E} \sup_{d(s,t) \le 2^{-k}} |X_s - X_t| \le L \sum_{n \ge k} 2^{-n} \varphi^{-1}(N(T,d,2^{-n-1})^2)\ ,$$

where $L$ is a number (which may change between occurrences). We delay the exercise of writing this inequality in integral form as

$$\mathsf{E} \sup_{d(s,t) \le \delta} |X_s - X_t| \le L \int_0^\delta \varphi^{-1}(N(T,d,\epsilon)^2) \mathrm{d}\epsilon\ . \tag{1.18}$$

In the case of the function $\varphi(x) = \exp(x^2/4) - 1$, so that $\varphi^{-1}(x) = 2\sqrt{\log(1+x)}$, inequality (1.18) is easily shown to be equivalent to the following more elegant formulation:

**Theorem 1.4.2 (Dudley's bound).** *If $(X_t)_{t \in T}$ is a Gaussian process with natural distance $d$ then*

$$\mathsf{E} \sup_{d(s,t) \le \delta} |X_s - X_t| \le L \int_0^\delta \sqrt{\log N(T,d,\epsilon)} \mathrm{d}\epsilon\ . \tag{1.19}$$

This *very general* inequality is *by far the most useful result* on continuity of Gaussian processes.

**Exercise 1.4.3.** Prove that the previous bound gives the correct modulus of continuity for Brownian motion on $[0, 1]$.

The message of Chapter 2 is simple:

- However useful, Dudley's bound is not optimal in a number of fundamentally important situations.
- It requires no more work to obtain a better bound which is optimal in every situation.

## 1.5 Chaining in a Metric Space: Pisier's Bound

In the bound (1.18) occurs the term $N(T, d, \epsilon)^2$ rather than $N(T, d, \epsilon)$. This does not matter if $\varphi(x) = \exp(x^2/4) - 1$, but it does matter if $\varphi(x) = |x|^p$. We really do not have the right integral in the right-hand side. In this section we show how to correct this, demonstrating at the same time that even in a structure as general as a metric space not all arguments are trivial. This material can be skipped at first reading.

To improve the brutal chaining argument leading to (1.18), without loss of generality we assume that $T$ is finite. For $n \geq n_0$ we consider a map $\theta_n : T_{n+1} \to T_n$ such that $d(\theta_n(t), t) \leq 2^{-n}$ for each $t \in T_n$. Since we assume that $T$ is finite, we have $T = T_m$ when $m$ is large enough. We fix such an $m$, and we define $\pi_n(t) = t$ for each $t \in T$ and each $n \geq m$. Starting with $n = m$ we then define recursively $\pi_n(t) = \theta_n(\pi_{n+1}(t))$ for $n \geq n_0$. The point of this construction is that $\pi_{n+1}(t)$ determines $\pi_n(t)$ so that there are at most $N(T, d, 2^{-n-1})$ pairs $(\pi_{n+1}(t), \pi_n(t))$, and the bound (1.13) implies

$$\mathsf{E} \sup_{t \in T} |X_{\pi_{n+1}(t)} - X_{\pi_n(t)}| \leq 2^{-n} \varphi^{-1}(N(T, d, 2^{-n-1})) \ . \tag{1.20}$$

Using the chaining identity

$$X_t - X_{\pi_n(t)} = \sum_{k \geq n} X_{\pi_{k+1}(t)} - X_{\pi_k(t)} \ ,$$

we have proved the following.

**Lemma 1.5.1.** *We have*

$$\mathsf{E} \sup_{t \in T} |X_t - X_{\pi_n(t)}| \leq \sum_{k \geq n} 2^{-k} \varphi^{-1}(N(T, d, 2^{-k-1})) \ . \tag{1.21}$$

Taking $n = n_0$ this yields the following clean result (due to G. Pisier):

**Theorem 1.5.2.** *We have*

$$\mathsf{E}\sup_{s,t\in T}|X_s - X_t| \le L\int_0^{\Delta(T)} \varphi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon\ . \tag{1.22}$$

In order to get a modulus of continuity, a clever twist is required in the argument.

**Theorem 1.5.3.** *For any $\delta > 0, n \ge n_0$ we have*

$$\mathsf{E}\sup_{d(s,t)<\delta}|X_s - X_t| \le \delta\varphi^{-1}(N(T,d,2^{-n})^2) + 4\sum_{k\ge n} 2^{-k}\varphi^{-1}(N(T,d,2^{-k-1}))\ . \tag{1.23}$$

To ensure that the right-hand side is small, we fix $n$ large enough so that the sum is small, and then we take $\delta$ small enough that the first term of the right-hand side is small.

*Proof.* We fix $n$ and we set $Z = \sup_{t\in T}|X_t - X_{\pi_n(t)}|$. We define

$$V = \{(\pi_n(s), \pi_n(t))\ ;\ d(s,t) < \delta\} \subset T_n \times T_n\ .$$

For $(a,b) \in V$, let $(s(a,b), t(a,b)) \in T\times T$ such that $\delta(s(a,b), t(a,b)) < \delta$ and $a = \pi_n(s(a,b))$, $b = \pi_n(t(a,b))$. Thus

$$\sup_{(a,b)\in V}|X_a - X_b| \le \sup_{(a,b)\in V}|X_{s(a,b)} - X_{t(a,b)}| + 2Z\ ,$$

and using (1.13)

$$\mathsf{E}\sup_{(a,b)\in V}|X_{s(a,b)} - X_{t(a,b)}| \le \delta\varphi^{-1}(N(T,d,2^{-n})^2)\ ,$$

so that

$$\mathsf{E}\sup_{(a,b)\in V}|X_a - X_b| \le \delta\varphi^{-1}(N(T,d,2^{-n})^2) + 2\mathsf{E}Z\ . \tag{1.24}$$

Now,

$$\sup_{d(s,t)<\delta}|X_s - X_t| \le \sup_{(a,b)\in V}|X_a - X_b| + 2Z\ .$$

Taking expectation and using (1.24) and (1.21) finishes the proof. □

## 1.6 Does this Book Contain any Idea?

**Idea 1.** It is possible to organize chaining optimally using increasing sequences of partitions.

**Idea 2.** There is an automatic device to construct such sequences of partitions, using "functionals", quantities with measure the size of the subsets of the index set. This yields a complete understanding of boundedness of Gaussian processes.

**Idea 3.** Ellipsoids are much smaller than one should think, because they (and more generally sufficiently convex bodies) are thin around the edges. This explains the funny fractional powers of logarithms in certain matching theorems.

**Idea 4.** One may witness that a metric space is large by the fact that it contains large trees, or equivalently that it supports an extremely scattered probability measure.

**Idea 5.** Consider a set on which $T$ you are given a distance $d$ and a random distance $d_\omega$ such that it is rare that, given $s, t \in T$ the distance $d_\omega(s,t)$ is much smaller that $d(s,t)$. Then if the appropriate sense $(T,d)$ is large, it must be the case that $(T, d_\omega)$ is typically large. This principle enormously constrains the structure of certain bounded processes such as random Fourier series and infinitely divisible processes.

**Idea 6.** There are different ways a random series might converge. It might converge because chaining witnesses that there is cancellation between terms, or it might converge because the sum of the absolutes values of its terms already converges. Many processes built on random series can be split in two parts, each one converging according to one of previous phenomenon.

The book contains many more ideas, but you will have to read more to discover them.

## 1.7 Overview by Chapters

For us a stochastic process is a collection of random variables (r.v.s) $(X_t)_{t\in T}$, where $T$ is an index set, and our main objective is to find conditions under which the trajectories of such a process are bounded. A specific feature of the space $T = [0,1]^m$ occurring in the Kolmogorov conditions is that it is really "$m$-dimensional" and "the same around each point". *Much of the work done in this book is to handle situations where such a homogeneity does not occur,* and these situations do occur in classical problems.

### 1.7.1 Gaussian Processes and the Generic Chaining

This subsection gives an overview of Chapter 2. More generally, Subsection 1.7.$n$ gives the overview for Chapter $n+1$.

The most important question considered in this book is the boundedness of Gaussian processes. The key object is the metric space $(T,d)$ where $T$ is the index set and $d$ the intrinsic distance (0.1). As investigated in Section 2.6 this metric space is far from being arbitrary: it is isometric to a subset of a

Hilbert space. It is however a deadly trap to try to use this specific property of the metric space $(T, d)$. The proper approach is to just think of it as a general metric space.

In Section 2.2 we explain the basic idea of the "generic chaining", one of the key ideas of this work. Chaining is a succession of steps that provide successive approximations of the index space $(T, d)$. In the Kolmogorov chaining, for each $n$ the "variation of the process during the $n$-th step is controlled uniformly over all possible chains". Generic chaining allows that the "variation of the process during the $n$-th step may depend on which chain we follow". Once the argument is properly organized, it is not any more complicated than the classical argument. It is in fact exactly the same. Yet, while Dudley's classical bound is not always sharp, the bound obtained through the generic chaining is optimal.

It is technically convenient to formulate the generic chaining bound using special sequences of partitions of the metric space $(T, d)$, that we shall call *admissible sequences* throughout the book. To key make the generic chaining bound useful is then to be able to construct admissible sequences. These admissible sequences measure an aspect of the "size" of the metric space. In Section 2.3 we introduce another method to measure the "size" of the metric space, through the behavior of certain "functionals", that are simply numbers attached to each subset of the entire space. The fundamental fact is that the two measures of the size of the metric space one obtains through either admissible sequences or through functionals are equivalent in full generality. This is proved in Section 2.3 for the easy part (that the admissible sequence approach provides a larger measure of size than the functional approach) and in Section 2.4 for the converse. This converse is, in effect, an algorithm to construct sequences of partitions in a metric space given a functional. Functionals are of considerable use throughout the book.

In Section 2.5 we prove that the generic bound can be reversed for Gaussian processes, therefore providing a characterization of their sample-boundedness. Generic chaining entirely explains the size of Gaussian processes, and the dream of Section 2.7 is that a similar situation will occur for many processes.

In Section 2.6 we explain why a Gaussian process in a sense *is* nothing but a subset of Hilbert space. Remarkably a number of basic questions remain unanswered, such as how to relate through geometry the size of a subset of Hilbert space seen as a Gaussian process with the corresponding size of its convex hull.

Dudley's bound fails to explain the size of the Gaussian processes indexed by ellipsoids in Hilbert space. This is investigated in Section 2.8. Ellipsoids will play a basic role in Chapter 3.

### 1.7.2 Matching Theorems

Chapter 3 makes the point that the generic chaining (or some equivalent form of it) is already required to really understand the irregularities occurring in the distribution of $N$ points $(X_i)_{i \le N}$ independently and uniformly distributed in the unit square. These irregularities are measured by the "cost" of pairing (=matching) these points with $N$ fixed points that are very uniformly spread, for various notions of cost.

These optimal results involve mysterious powers of $\log N$. We are able to trace them back to the geometry of ellipsoids in Hilbert space, so we start the chapter with an investigation of these ellipsoids in Section 3.2. The philosophy of the main result, the Ellipsoid Theorem, is that an ellipsoid is in some sense somewhat smaller than it appears at first. This is due to convexity: an ellipsoid gets "thinner" when one gets away from its center. The Ellipsoid Theorem is a special case of a more general result (with the same proof) about the structure of sufficiently convex bodies, one that will have important applications in Chapter 16.

In Section 3.3 we provide general background on matchings. In Section 3.5 we investigate the case where the cost of a matching is measured by the average distance between paired points. We prove the result of Ajtai, Komlós, Tusnády, that the expected cost of an optimal matching is at most $L\sqrt{\log N}/\sqrt{N}$ where $L$ is a number. The factor $1/\sqrt{N}$ is simply a scaling factor, but the fractional power of log is optimal as shown in Section 3.6. In Section 3.7 we investigate the case where the cost of a matching is measured instead by the maximal distance between paired points. We prove the theorem of Leighton and Shor that the expected cost of a matching is at most $L(\log N)^{3/4}/\sqrt{N}$, and the power of log is shown to be optimal in Section 3.8.

With the exception of Section 3.2, the results of Chapter 3 are not connected to any subsequent material before Chapter 14.

### 1.7.3 Mostly Trees

We describe different notions of trees, and show how one can measure the "size" of a metric space by the size of the largest trees it contains, in a way which is equivalent to the measures of size introduced in Chapter 2. This idea played an important part in the history of Gaussian processes. Its appeal is mostly that trees are easy to visualize. Building a large tree in a metric space is an efficient method to bound its size from below. We then learn the powerful method of measuring the size of a metric space by the existence of very scattered probability measures: "witnessing measures".

### 1.7.4 Bernoulli Processes

In Chapter 5 we investigate Bernoulli processes, where the individual random variables $X_t$ are linear combinations of independent random signs. Random signs are obviously important r.v.s, and occur frequently in connection with "symmetrization procedures", a very useful tool. Each Bernoulli process is associated with a Gaussian process in a canonical manner, when one replaces the random signs by independent standard Gaussian r.v.s. The Bernoulli process has better tails than the corresponding Gaussian process (it is "subgaussian") and is bounded whenever the corresponding Gaussian process is bounded. There is however a completely different reason for which a Bernoulli process might be bounded, namely that the sum of the absolute values of the coefficients of the random signs remain bounded independently of the index $t$. A natural question is then to decide whether these two extreme situations are the only fundamental reasons why a Bernoulli process can be bounded, in the sense that a suitable "mixture" of them occurs in every bounded Bernoulli process. This was the "Bernoulli Conjecture" (to be stated formally on page 111), which has been so brilliantly solved by W. Bednorz and R. Latała.

It is a long road to the solution of the Bernoulli conjecture, and we start to study the main tools to work with Bernoulli processes. A linear combination of independent random signs looks like a Gaussian r.v. when the coefficients of the random signs are small. We can expect that a Bernoulli process will look like a Gaussian process when these coefficients are suitably small. We also develop this fundamental idea: the key to understanding Bernoulli processes is to achieve control in the supremum norm.

The Bernoulli conjecture, on which the author worked so many years, greatly influenced the way he looked at various processes. In the case of empirical processes, this is explained in Section 5.8.

### 1.7.5 Random Fourier Series and Trigonometric Sums

The basic example of a random Fourier series is

$$X_t = \sum_{k \geq 1} \xi_k \exp(2\pi ikt) \ , \tag{1.25}$$

where $t \in [0,1]$ and the r.v.s $\xi_k$ are independent symmetric. In this chapter we provide a final answer to the convergence of such series.

The fundamental case where $\xi_k = a_k g_k$ for numbers $a_k$ and independent Gaussian r.v.s $(g_k)$ is of great historical importance. There is however another motivation for the study of such series. The generic chaining and related methods is well adapted to the case of "non-homogeneous index space". The study of certain of the processes we will consider in the next chapters is however already subtle even in the absence of the extra difficulty due to this lack of homogeneity. The setting of random Fourier series allows us to put aside the issue of lack of homogeneity and to concentrate on the other difficulties. It provides an ideal setting to understand a basic fact: : many processes can be exactly controlled, not by using one or two distances, but by using an entire family of distances. This concept of "family of distance" will play a major role later. Another highlight of the chapter is a technical result allowing to perform efficient chaining for Bernoulli processes.

### 1.7.6 Partition Scheme for Families of Distances

Once one has survived the initial surprise of the new idea that many processes are naturally associated to an entire family of distances, it is very pleasant to realize that the tools of Section 2.4 can be extended to this setting with essentially the same proof. This is the purpose of Section 7.1.

In Section 7.3 we apply these tools to the situation of "canonical processes" where the r.v.s $X_t$ are linear combinations of independent copies of symmetric r.v.s with density proportional to $\exp(-|x|^\alpha)$ where $\alpha \geq 1$ (and to considerably more general situations as discovered by R. Latała). In these situations, the size of the process can be completely described as a function of the geometry of the index space, a far reaching extension of the Gaussian case.

### 1.7.7 Proof of the Bernoulli Conjecture

Having learned how to manipulate "families of distances" we are now better prepared to prove the Bernoulli conjecture. The proof occupies most of Chapter 0. In the last section of we investigate how to get lower bounds on Bernoulli processes using "witnessing measures".

### 1.7.8 Infinitely Divisible Processes

We study these processes in a general setting: we make no assumption of stationarity of increments of any kind and our processes are to Lévy processes what a general Gaussian process is to Brownian motion.

As a prologue we study $p$-stable processes. These are conditionally Gaussian, and in Section 1.1 we use this property to provide lower bounds for such processes. Although these bounds are in general very far from being upper bounds, they are in a sense optimal (where there is "stationarity", these lower bounds can be reversed).

Our main tool to study more general infinitely divisible processes is their representation as conditionally Bernoulli processes. We may now use the Latała-Bednorz theorem to considerably simplify the author's previous results. For a large class of infinitely divisible processes, we prove lower bounds which are in a precise sense optimal. These lower bounds are not upper bounds in general, but they are upper bounds for "the part of boundedness of the process which is due to cancellation". Thus, whatever bound might be true for the "remainder of the process" owes nothing to cancellation. The results are described in complete detail with all definitions in Section 1.4.

### 1.7.9 Unfulfilled Dreams

In Chapter 10 we outlay a long range research program, concerning a natural class of processes, selector processes, for which it could be true that "chaining explains all the part of the boundedness which is due to cancellation", and in Section 10.2 we state a precise conjecture to that effect. Even if this conjecture is true, there would remain to describe the "part of the boundedness which owes nothing to cancellation", and for this part also we propose sweeping conjectures. The underlying hope behind these conjectures is that, ultimately, a bound for a selector process always arises from the use of the 'union bound' $\mathsf{P}(\cup_n A_n) \le \sum_n \mathsf{P}(A_n)$ in a simple situation, the use of basic principles such as linearity and positivity, or combinations of these.

### 1.7.10 Empirical Processes

We focus on a special yet fundamental topic: the control of the supremum of the empirical process over a class of functions.

We demonstrate again the power of the chaining scheme of Section 0.8 by providing a sharper version of Ossiander's bracketing theorem with a very simple proof. We then illustrate various techniques by presenting proofs of two deep recent results.

### 1.7.11 Gaussian Chaos

Our satisfactory understanding of the properties of Gaussian processes should bring information about processes that are, in various senses, related to Gaussian processes. Such is the case of order two Gaussian chaos (which are essentially second degree polynomials of Gaussian random variables). It seems at present a hopelessly difficult task to give lower and upper bounds of the same order for these processes, but in Section 12.1 we obtain a number of results in the right direction. Chaos processes are very instructive because there exists other methods than chaining to control their size (a situation which we do not expect to occur for processes defined as sums of a random series).

In Section 12.2 we study the tails of a single multiple order Gaussian chaos, and present (yet another) deep result of R. Latała which provides a rather complete description of these tails.

### 1.7.12 Convergence of Orthogonal Series; Majorizing Measures

The old problem of characterizing the sequences $(a_m)$ such that for each orthonormal sequence $(\varphi_m)$ the series $\sum_{m\ge1} a_m \varphi_m$ converges a.s. was solved by A. Paszkiewicz. Using a more abstract point of view, we present a very much simplified proof of his results (due essentially to W. Bednorz). This leads us to the question of discussing when a certain condition on the "increments" of a process implies its boundedness. When the increment condition is of "polynomial type", this is more difficult than in the case of Gaussian processes, and requires the notion of "majorizing measure". We present several elegant results of this theory, in their seemingly final form recently obtained by W. Bednorz.

### 1.7.13 Shor's Matching Theorem

This chapter continues Chapter 3. We prove a deep improvement of the Ajtai, Komlós, Tusnády theorem due to P. Shor. Unfortunately, due mostly to our lack of geometrical understanding, the best conceivable matching theorem, which would encompass this result as well as those of Chapter 3, and much more, remains as a challenging problem, "the ultimate matching conjecture" (a conjecture which is solved in the next chapter in dimension $\geq 3$).

### 1.7.14 The Ultimate Matching Conjecture in Dimension Three

In this case, which is easier than the case of dimension two (but still apparently rather non-trivial), we are able to obtain the seemingly final result about matchings, a strong version of "the ultimate matching conjecture". There are no more fractional powers of $\log N$ here, but in a random sample of $N$ points uniformly distributed in $[0,1]^3$, local irregularities occur at all scales between $N^{-1/3}$ and $(\log N)^{1/3}N^{-1/3}$, and our result can be seen as a precise global description of these irregularities.

### 1.7.15 Applications to Banach Space Theory

Chapter 16 gives applications to Banach space theory. The sections of this Chapter are largely independent of each other, but all reflect past interests of the author. In Section 16.1.2, we study the cotype of operators from $\ell_N^\infty$ into a Banach space. In Section 16.1.5, we prove a comparison principle between Rademacher (=Bernoulli) and Gaussian averages of vectors in a finite dimensional Banach space, and we use it to compute the Rademacher cotype-2 of a finite dimensional space using only a few vectors. In Section 16.2.1 we discover how to classify the elements of the unit ball of $L^1$ "according to the size of the level sets". In Section 16.2.2 we explain, given a Banach space $E$ with an 1-unconditional basis $(e_i)$, how to "compute" the quantity $\mathsf{E}\|\sum_i g_i e_i\|$ when $g_i$ are independent Gaussian r.v.s, a further variation on the fundamental theme of the interplay between the $L^1, L^2$ and $L^\infty$ norms. In Section 16.3.1 we study the norm of the restriction of an operator from $\ell_N^q$ to the subspace generated by a randomly chosen small proportion of the coordinate vectors, and in Section 16.3.2 we use these results to obtain a sharpened version of the celebrated results of J. Bourgain on the $\Lambda_p$ problem. In Section 16.3.3, given a uniformly bounded orthonormal system, we study how large a subset we can find on the span of which the $L^2$ and $L^1$ norms are close to each other. In Section 16.3.4, given $1 < p < 2$ and a $k$-dimensional subspace of $L^p$ we investigate for which values of $N$ we can embed nearly isometrically this subspace as a subspace of $\ell^p_N$. We prove that we may choose $N$ as small as about $k \log k (\log \log k)^2$. A recent proof by G. Schechtman of a theorem of Y. Gordon concludes this chapter in Section 16.3.5.

# Part I

# The Generic Chaining

# 2. Gaussian Processes and the Generic Chaining

## 2.1 Overview

The overview of this chapter is given in Chapter 1, Subsection 1.7.1. More generally, Subsection 1.7.$n$ is the overview of Chapter $n+1$.

## 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:

$$\forall u > 0\ ,\ \mathsf{P}(|X_s - X_t| \geq u) \leq 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right)\ . \tag{2.1}$$

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.

$$\forall t \in T\ ,\ \ \mathsf{E}X_t = 0\ . \tag{2.2}$$

We will measure the "size of the process $(X_t)_{t\in T}$" by the quantity $\mathsf{E}\sup_{t\in T} X_t$. Why this quantity is a good measure of the "size of the process" is explained in Lemma 2.2.1 below.

When $T$ is uncountable it is not obvious what the quantity $\mathsf{E}\sup_{t\in T} X_t$ means. We *define* it by the following formula:

$$\mathsf{E}\sup_{t\in T} X_t = \sup\Big\{\mathsf{E}\sup_{t\in F} X_t\ ;\ F\subset T\ ,\ F \text{ finite}\Big\}\ , \tag{2.3}$$

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.

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. □

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

$$\mathsf{E}\sup_{t\in T}|X_t| \le 2\mathsf{E}\sup_{t\in T} X_t + \mathsf{E}|X_{t_0}| \le 3\mathsf{E}\sup_{t\in T}|X_t| \ . \tag{2.4}$$

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.[1] 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.[2]

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

We actually often need to control the tails of the r.v. $\sup_{s,t\in T}|X_s - X_t|$, not only its first moment. Emphasis is given to the first moment because 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 (2.3), does not decrease generality.

Given any $t_0$ in $T$, the centering hypothesis (2.2) implies

$$\mathsf{E}\sup_{t\in T} X_t = \mathsf{E}\sup_{t\in T}(X_t - X_{t_0}) \ . \tag{2.5}$$

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,

---

[1] It would have taken supra-human dedication for the author to write in detail all the solutions, so there is no real warranty that each exercise is really feasible or even correct.

[2] 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".

$$\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y > u)\,\mathrm{d}u\,. \tag{2.6}$$

Thus it is natural to look for bounds of

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

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

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \ge u\Big) \le \sum_{t\in T}\mathsf{P}(X_t - X_{t_0} \ge u)\,. \tag{2.8}$$

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 $T$,

$$\Delta(T) = \sup_{t_1,t_2\in T} d(t_1,t_2)\,. \tag{2.9}$$

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 (2.1) and (2.8) imply

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \ge u\Big) \le 2\,\mathrm{card}\,T\exp\Big(-\frac{u^2}{2\Delta(T)^2}\Big)\,. \tag{2.10}$$

Let us now record a simple yet important computation, that will allow us to use the information (2.10).

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

$$\forall\, u > 0\,,\ \mathsf{P}(Y \ge u) \le A\exp\Big(-\frac{u^2}{B^2}\Big) \tag{2.11}$$

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

$$\mathsf{E}Y \le LB\sqrt{\log A}\,. \tag{2.12}$$

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 one certainly needs to get used to it, 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 (2.6) and we observe that since $\mathsf{P}(Y \ge u) \le 1$, then for any number $u_0$ we have

$$
\begin{aligned}
\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y \ge u)\mathrm{d}u = \int_0^{u_0} \mathsf{P}(Y \ge u)\mathrm{d}u + \int_{u_0}^\infty \mathsf{P}(Y \ge u)\mathrm{d}u \\
&\le u_0 + \int_{u_0}^\infty A\exp\Big(-\frac{u^2}{B^2}\Big)\mathrm{d}u \\
&\le u_0 + \frac{1}{u_0}\int_{u_0}^\infty uA\exp\Big(-\frac{u^2}{B^2}\Big)\mathrm{d}u \\
&= u_0 + \frac{AB^2}{2u_0}\exp\Big(-\frac{u_0^2}{B^2}\Big) , \qquad (2.13)
\end{aligned}
$$

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

Combining (2.12) and (2.10) we obtain that (considering separately the case where $\operatorname{card} T = 1$)

$$\mathsf{E}\sup_{t\in T} X_t \le L\Delta(T)\sqrt{\log \operatorname{card} T} . \qquad (2.14)$$

The following special case is fundamental.

**Lemma 2.2.4.** *If $(g_k)_{k\ge1}$ are standard Gaussian r.v.s then*

$$\mathsf{E}\sup_{k\le N} g_k \le L\sqrt{\log N} . \qquad (2.15)$$

**Exercise 2.2.5.** (a) Prove that (2.15) holds as soon as the r.v.s $g_k$ satisfy

$$\mathsf{P}(g_k \ge t) \le 2\exp\Big(-\frac{t^2}{2}\Big) \qquad (2.16)$$

for $t > 0$.
(b) For $N \ge 2$ construct $N$ centered r.v.s $(g_k)_{k\le N}$ satisfying (2.16), 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.)
(d) After learning (2.17) 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 (2.15) 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 $(A_k)_{k\ge1}$. Prove that

$$\mathsf{P}\Big(\bigcup_{k\le N} A_k\Big) \ge 1 - \exp\Big(-\sum_{k\le N}\mathsf{P}(A_k)\Big) . \qquad (2.17)$$

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

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

$$\sum_{k \leq N} \mathsf{P}(Y_k \geq u) \geq 1 \; . \tag{2.18}$$

Prove that

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

Hint: use (2.17) to prove that $\mathsf{P}(\sup_{k \leq N} Y_k \geq u) \geq 1/L$.
(b) We assume (2.18), but now $Y_k$ need not be $\geq 0$. Prove that

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

Hint: observe that for each event $\Omega$ 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, we consider in detail the consequences of an "exponential decay of tails" such as in (2.11). 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

$$\forall\, u > 0 \; , \; \mathsf{P}(Y \geq u) \leq 2 \exp\Big(-\frac{u}{B}\Big) \; . \tag{2.19}$$

Prove that

$$\mathsf{E} \exp\left(\frac{Y}{2B}\right) \leq L \; . \tag{2.20}$$

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 (2.20) that for $p \geq 1$ one has

$$(\mathsf{E}Y^p)^{1/p} \leq LpB \; . \tag{2.21}$$

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

$$\forall\, u > 0 \; , \; \mathsf{P}(Y \geq u) \leq 2 \exp\Big(-\frac{u^2}{B^2}\Big) \; , \tag{2.22}$$

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

$$(\mathsf{E}Y^p)^{1/p} \leq LB\sqrt{p} \; . \tag{2.23}$$

In words, (2.21) states that "as $p$ increases, the $L^p$ norm of an exponentially integrable r.v. does not grow faster than $p$," and (2.23) 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, (2.23) 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 (2.23) that for $p \geq 1$ one has $(\mathsf{E}|g|^p)^{1/p} \leq L\sqrt{p}$. Prove one has also

$$(\mathsf{E}|g|^p)^{1/p} \geq \frac{\sqrt{p}}{L} \ . \tag{2.24}$$

One knows how to compute exactly $\mathsf{E}|g|^p$, from which one can deduce (2.24). You are however asked to provide a proof in the spirit of this work by deducing (2.24) 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 (2.14) 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

$$X_t - X_{t_0} = X_t - X_{\pi_1(t)} + X_{\pi_1(t)} - X_{t_0} \ . \tag{2.25}$$

The idea is that it will be effective to use (2.8) 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 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

$$X_t - X_{t_0} = \sum_{n\geq 1} \big(X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\big) \ , \tag{2.26}$$

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 (2.26) 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

$$\operatorname{card} T_n \leq N_n \tag{2.27}$$

where

$$N_0 = 1\,;\; N_n = 2^{2^n} \quad \text{if} \quad n \geq 1\,. \tag{2.28}$$

The notation (2.28) will be used throughout the book. It is at this stage that the procedure to control $T_n$ differs form the traditional one, and it is the 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 it 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

$$d(t, \pi_n(t)) = d(t, T_n) = \inf_{s\in T_n} d(t,s)\,. \tag{2.29}$$

For $u > 0$, (2.1) 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 $\Omega_u$ the event defined by

$$\forall n \geq 1\,,\, \forall t\,,\, |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \leq u 2^{n/2} d(\pi_n(t), \pi_{n-1}(t))\,, \tag{2.30}$$

we obtain

$$\mathsf{P}(\Omega_u^c) \leq p(u) := \sum_{n\geq 1} 2\cdot 2^{2^{n+1}} \exp(-u^2 2^{n-1})\,. \tag{2.31}$$

Here again, at the crucial step, we have used the "union bound": we bound the probability that one of the events (2.30) fails by the sum of the probabilities that the individual events fail. When $\Omega_u$ occurs, (2.26) yields

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

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

$$\mathsf{P}\Big(\sup_{t \in T} |X_t - X_{t_0}| > uS\Big) \le L \exp\Big(-\frac{u^2}{2}\Big) \,. \tag{2.32}$$

In particular (2.32) implies

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

The triangle inequality and (2.6) 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

$$\mathsf{E} \sup_{t \in T} X_t \le L \sup_{t \in T} \sum_{n \ge 0} 2^{n/2} d(t, T_n) \,. \tag{2.33}$$

Now, how do we construct the sets $T_n$? It is obvious that we should try to make the right-hand side of (2.33) small, but this is obvious only because we have used an approach which naturally leads to this bound. The "traditional chaining method" (as used in Section 1.4) chooses the sets $T_n$ so that $\sup_{t \in T} d(t, T_n)$ is as small as possible for $\operatorname{card} T_n \le N_n$, where

$$d(t, T_n) = \inf_{s \in T_n} d(t, s) \,. \tag{2.34}$$

We define

$$e_n(T) = e_n(T, d) = \inf \sup_t d(t, T_n)\,, \tag{2.35}$$

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 not consistent with the conventions of Operator Theory, which uses $e_{2^n}$ to denote what we call $e_n$.[3] When $T$ is infinite, the numbers $e_n(T)$ are also defined by (2.35) but are not always finite (e.g. when $T = \mathbb{R}$).

Let us note that, since $N_0 = 1$,

$$\frac{\Delta(T)}{2} \le e_0(T) \le \Delta(T)\,. \tag{2.36}$$

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 $t$, (2.33) implies the following.

**Proposition 2.2.10 (Dudley's entropy bound [32]).** *Under the increment condition (2.1), it holds*

$$\mathsf{E} \sup_{t \in T} X_t \le L \sum_{n \ge 0} 2^{n/2} e_n(T)\,. \tag{2.37}$$

We proved this bound only when $T$ is finite, but using (2.3) 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 $e_n(U) \le 2e_n(T)$.*

The point here is that in the definition of $e_n(U)$ we insist that the balls are centered in $U$, not in $T$.

*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 $2a$. □

**Exercise 2.2.12.** Prove that the factor 2 in the inequality $e_n(U) \le 2e_n(T)$ cannot be improved even if $n = 0$.

Dudley's entropy bound is usually formulated using the covering numbers of Definition 1.4.1. These 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,

[3] We can't help it if Operator Theory gets it wrong.

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

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

$$\sqrt{\log 2}\sum_{n\ge 0} 2^{n/2}(e_n(T) - e_{n+1}(T)) \le \int_0^{e_0(T)} \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\ . \tag{2.38}$$

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 (2.38) 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

$$\mathsf{E}\sup_{t\in T} X_t \le L \int_0^\infty \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\ . \tag{2.39}$$

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 2.2.13.** 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 (2.37) is not an improvement over (2.39).

**Exercise 2.2.14.** Assume that for each $0 < \epsilon < A$ we have $\log N(t,d,\epsilon) \le (A/\epsilon)^\alpha$. Prove that $e_n(T) \le K(\alpha) A 2^{-n/\alpha}$.

Here $K(\alpha)$ is a number depending only on $\alpha$. 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 $\alpha$ 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 2.8.5.

**Exercise 2.2.15.** (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 $\mathbb{R}^m$,

$$N(A, d, \epsilon) \geq \frac{\mathrm{Vol}(A)}{\mathrm{Vol}(\epsilon B)} \tag{2.40}$$

and

$$N(A, d, 2\epsilon) \leq N^*(A, d, 2\epsilon) \leq \frac{\mathrm{Vol}(A + \epsilon B)}{\mathrm{Vol}(\epsilon B)} \ . \tag{2.41}$$

(c) Conclude that

$$\left(\frac{1}{\epsilon}\right)^m \leq N(B, d, \epsilon) \leq \left(\frac{2+\epsilon}{\epsilon}\right)^m \ . \tag{2.42}$$

(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 (2.40) and (2.41) to a more abstract setting, but with the same proofs. Consider a metric space $(T, d)$ and a positive measure $\mu$ on $T$ such all balls of a given radius have the same measure, $\mu(B(t, \epsilon)) = \varphi(\epsilon)$ for each $\epsilon > 0$ and each $t \in T$. 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 51. Yet other examples based on fundamental geometry (ellipsoids in $\mathbb{R}^m$) are explained in Section 2.8.

**Exercise 2.2.16.** 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

$$T = \Big\{(t_i)_{i\le m}\ ;\ t_i \ge 0\ ,\ \sum_{i\le m} t_i = 1\Big\}\ , \tag{2.43}$$

the convex hull of the canonical basis. By (2.15) 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 (2.37) 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 2.2.15 with $T = A = T_k$ and $\mu$ 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 2.4.9 below will show that in $\mathbb{R}^m$ the situation cannot be worse than this.

The bound (2.33) seems to be genuinely better than the bound (2.37) because when going from (2.33) to (2.37) 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 (2.33) is probably the most important idea of this work. 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 (2.33) instead of choosing them as in (2.35). 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 2.2.16, to find sets $T_n$ such that the right-hand side of (2.33) 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 (2.33) is an actual improvement over (2.37).

**Exercise 2.2.17.** (a) Consider a finite metric space $(T, d)$. Assume that it contains a point $t_0$ with the property that for $n \ge 0$ we have $\operatorname{card}(T \setminus B(t_0, 2^{-n/2})) \le N_n - 1$. Prove that $T$ contains sets $T_n$ with $\operatorname{card} T_n \le N_n$ and $\sup_{t\in T}\sum_{n\ge 0} 2^{n/2} d(t, T_n) \le L$. 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 $\mathbb{R}^M$.

The idea behind the bound (2.33) admits a technically more convenient formulation.

**Definition 2.2.18.** *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 $\mathcal{A}_n$. Admissible sequences of partitions will be constructed recursively, by breaking each element $C$ of $\mathcal{A}_n$ into at most $N_n$ pieces, obtaining then a partition $\mathcal{A}_{n+1}$ of $T$ consisting of at most $N_n^2 \leq N_{n+1}$ pieces.

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 (2.28) 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 2.2.19.** *(The generic chaining bound). Under the increment condition (2.1) (and if $\mathsf{E} X_t = 0$ for each $t$) then for each admissible sequence $(\mathcal{A}_n)$ we have*

$$\mathsf{E} \sup_{t \in T} X_t \leq L \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} \Delta(A_n(t)) \,. \tag{2.44}$$

Here as always, $\Delta(A_n(t))$ denotes the diameter of $A_n(t)$ for $d$. One could think that (2.44) could be much worse than (2.33), 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 (2.33). □

**Definition 2.2.20.** *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 2.2.21.** Prove that if $d \le Bd'$ then $\gamma_2(T,d) \le B\gamma_2(T,d')$.

**Exercise 2.2.22.** (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

$$2^{n/2} e_n(T) \le L\gamma_2(T,d) \; . \tag{2.45}$$

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 (2.45) with Exercise 2.4.8 below.

Combining Theorem 2.2.19 with Definition 2.2.20 yields

**Theorem 2.2.23.** *Under (2.1) and (2.2) we have*

$$\mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T,d) \; . \tag{2.46}$$

To make (2.46) 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 2.4.

Let us point out, recalling (2.32), and observing that

$$|X_s - X_t| \le |X_s - X_{t_0}| + |X_t - X_{t_0}| \; , \tag{2.47}$$

we have actually proved

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge Lu\gamma_2(T,d)\Big) \le 2\exp(-u^2) \; . \tag{2.48}$$

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 the reader is advised to write every detail on how this is deduced from (2.32): the different exponents in (2.32) and (2.48) are made possible by the fact that the constant $L$ is not the same in these inequalities.

We note that (2.48) implies a lot more than (2.46). Indeed, for each $p \ge 1$, using (2.23)[4]

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

and in particular

$$\mathsf{E}\big(\sup_{s,t} |X_s - X_t|\big)^2 \le L\gamma_2(T,d)^2 \; . \tag{2.50}$$

[4] A tiny bit of extra work, as done e.g. in [145] shows the more precise result that $(\mathsf{E}\sup_{s,t} |X_s - X_t|^p)^{1/p} \le K\gamma_2(T,d) + L\sqrt{p}\Delta(T)$.

## 2.3 Functionals

How do we efficiently construct admissible sequences of partitions? The quantity $\gamma_2(T,d)$ reflects a highly non-trivial geometric characteristic of the metric space. We will never, ever, enjoy a free lunch. This geometry must be understood in order to build competent admissible sequences. Although this will become clear only gradually, one crucial way to bring up this geometry is through *functionals*. 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 $F(H) \geq 0$, and if it is increasing, i.e.

$$H \subset H' \subset T \Rightarrow F(H) \leq F(H') \,. \tag{2.51}$$

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. A first example is given by $F(H) = \Delta(H)$. In the same direction, a fundamental example of a functional is

$$F(H) = \gamma_2(H,d) \,. \tag{2.52}$$

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 $T$.

For our purposes the relevant property of functionals is by no means intuitively obvious yet (but we shall soon see that the functional (2.52) does enjoy this property). Let us first try to explain it in words: if set 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 property will be called a "growth condition".

Let us address a secondary point before we give definitions. We denote by $B(t,r)$ the ball centered at $t$ of radius $r$, and we note that

$$\Delta(B(t,r)) \leq 2r \,.$$

This factor 2 is a nuisance. It is qualitatively the same to say that a set is contained in a ball of small radius or has small diameter, but quantitatively we have to account for this factor 2. In countless constructions we will produce sets $A$ which are "small" because they are contained in a ball of small radius $r$. Either we keep track of this property, which is cumbersome, or we control the size of $A$ through its diameter and we deal with this inelegant factor 2. We have chosen here the second method.[5]

What do we mean by "small pieces far from each other"? There is a scale $a > 0$ at which this happens, and a parameter $r \geq 8$ which gives us some room. The pieces are small at that scale: they are contained in balls with radius $2a/r$.[6] The balls are far from each other: any two centers of such

[5] The opposite choice was made in [145].

[6] This coefficient 2 is motived by the considerations of the previous paragraph.

balls are at mutual distance $\geq a$. The reason why we require $r \geq 8$ is that we want the following: two points taken in different balls with radius $2a/r$ whose centers are at distance $\geq a$ cannot be too close to each other. This would not be true for, say, $r = 4$, so we give ourselves some room, and take $r \geq 8$. Here is the formal definition.

**Definition 2.3.1.** *Given $a > 0$ and an integer $r \geq 8$ we say that subsets $H_1, \ldots, H_m$ of $T$ are $(a, r)$-separated if*

$$\forall \ell \leq m\,,\ H_\ell \subset B(t_\ell, 2a/r)\ , \tag{2.53}$$

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

$$\forall \ell\,,\ \ell' \leq m\,,\ \ell \neq \ell' \Rightarrow a \leq d(t_\ell, t_{\ell'}) \leq 2ar\ . \tag{2.54}$$

A secondary feature of this definition is that the small pieces $H_\ell$ 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 last inequality in condition (2.54).

**Exercise 2.3.2.** Find interesting examples of metric spaces for which there are no points $t_1, \ldots, t_m$ as in (2.54), for all values of $n$ (respectively all large enough values of $n$).

Now, what means "the union of the pieces is significantly larger than the *smallest* of these pieces"? In this first version of the growth condition, it 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.[7] 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. 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 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 $2^{n/2}$.

**Definition 2.3.3.** *We say that the functional $F$ satisfies the* growth condition *with parameters $r \geq 8$ and $c^* > 0$ if for any integer $n \geq 1$ and any $a > 0$ the following holds true, where $m = N_n$. For each collection of subsets $H_1, \ldots, H_m$ of $T$ that are $(a, r)$-separated it holds that*

$$F\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq c^* a 2^{n/2} + \min_{\ell \leq m} F(H_\ell)\ . \tag{2.55}$$

[7] We remind the reader that the function $\sqrt{\log y}$ arises from the fact it is the inverse of the function $\exp(-x^2)$.

**Exercise 2.3.4.** Find examples of spaces $(T, d)$ where the growth condition holds while $F(H) = 0$ for each each $H \subset T$. Hint: use Exercise 2.3.2.

The following illustrates how we might use the first part of (2.54).

**Exercise 2.3.5.** 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 functional satisfies the growth condition of Definition 2.3.3, it suffices to consider the values of $n$ for which $N_{n+1} \le (1 + 2/r)^k$. Hint: it follows from (2.42) that for larger values of $n$ there are no points $t_1, \ldots, t_m$ as in (2.54).

You may find it hard to give simple examples of functionals which satisfy the growth condition (2.55). It will become gradually apparent that this condition imposes strong restrictions on the metric space $(T, d)$ and in particular a control from above of the quantity $\gamma_2(T, d)$. It bears repeating that $\gamma_2(T, d)$ reflects the geometry of the space $(T, d)$. Once this geometry is understood, it is usually possible to guess a good choice for the functional $F$. Many examples will be given in subsequent chapters.

As we show now, we really have no choice. Functionals with the growth property are intimately connected with the quantity $\gamma_2(T, d)$.

**Proposition 2.3.6.** *Assume* $r \ge 16$. *Then the functional* $F(H) = \gamma_2(H, d)$ *satisfies the growth condition with parameters* $r$ *and* $c^* = 8$.

*Proof.* Let $m = N_n$ and consider points $(t_\ell)_{\ell \le m}$ of $T$ with $d(t_\ell, t_{\ell'}) > a$ if $\ell \ne \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/8)$, and the set $H = \bigcup_{\ell \le m} H_\ell$. We have to prove that

$$\gamma_2(H, d) \ge \frac{1}{8} a 2^{n/2} + \min_{\ell \le m} \gamma_2(H_\ell, d) \; . \tag{2.56}$$

Consider an admissible sequence of partitions of $(\mathcal{A}_n)$ of $H$, and consider the set

$$I = \{\ell \le m \; ; \; \exists A \in \mathcal{A}_{n-1} \; ; \; A \subset H_\ell\} \; .$$

Thus there is a one-to-one map from $I$ to $\mathcal{A}_{n-1}$ and since $\operatorname{card} I = N_n > \operatorname{card} \mathcal{A}_{n-1}$ there exists $\ell \notin I$. The next goal is to prove that for $t \in H_\ell$ one has

$$\sum_{k \ge 0} 2^{k/2} \Delta(A_k(t)) \ge \frac{1}{4} a 2^{(n-1)/2} + \sum_{k \ge 0} 2^{k/2} \Delta(A_k(t) \cap H_\ell) \; . \tag{2.57}$$

Each term of the sum in the left is $\ge$ to the corresponding term on the right, so it suffices to prove that

$$\Delta(A_{n-1}(t)) \ge \Delta(A_{n-1}(t) \cap H_\ell) + \frac{1}{4} a \; . \tag{2.58}$$

For $t \in H_\ell$, since $\ell \notin I$, 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 $d(t, B(t_{\ell'}, a/8)) \geq a/2$, this implies that $\Delta(A_n(t)) \geq a/2$. This proves (2.58) since $\Delta(A_n(t) \cap H_\ell) \leq \Delta(H_\ell) \leq a/4$.

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

$$\sup_{t \in H_\ell} \sum_{k \geq 0} 2^{k/2} \Delta(A_k(t) \cap H_\ell) \geq \gamma_2(H_\ell, d) \ .$$

Hence, taking the supremum over $t$ in $H_\ell$ in (2.57) we get

$$\sup_{t \in H_\ell} \sum_{k \geq n} 2^{k/2} \Delta(A_k(t)) \geq \frac{1}{2} a 2^{n/2} + \gamma_2(H_\ell, d) \geq \frac{1}{2} a 2^{n/2} + \min_\ell \gamma_2(H_\ell, d) \ .$$

Since the admissible sequence $(\mathcal{A}_n)$ is arbitrary, we have proved (2.56). □

## 2.4 Partitioning Schemes

In this section we use functionals satisfying the growth condition to construct admissible sequences of partitions. The basic result is as follows.

**Theorem 2.4.1.** *Assume that there exist on $T$ a functional which satisfies the growth condition of Definition 2.3.3 with parameters $r$ and $c^*$. Then*

$$\gamma_2(T, d) \leq \frac{Lr}{c^*} F(T) + Lr\Delta(T) \ . \tag{2.59}$$

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 (2.59). For this reason, it could be called "the fundamental partitioning theorem."

**Exercise 2.4.2.** Consider a metric space $T$ consisting of exactly two points. Prove that the functional given by $F(H) = 0$ for each $H \subset T$ satisfies the growth condition of Definition 2.3.3 for $r = 8$ and any $c^* > 0$. Explain why we cannot replace (2.59) by the inequality $\gamma_2(T, d) \leq LrF(T)/c^*$.

Let us first stress the following trivial fact which will be used many times.

**Lemma 2.4.3.** *Consider an integer $N$. If we cannot cover $T$ by $< N$ balls of radius $a$ then there exists points $(t_\ell)_{\ell \leq N}$ with $d(t_\ell, t_{\ell'}) \geq a$ for $\ell \neq \ell'$. In particular if $e_n(T) > a$ we can find points $(t_\ell)_{\ell \leq N_n}$ with $d(t_\ell, t_{\ell'}) \geq a$ for $\ell \neq \ell'$.*

*Proof.* We pick the points $t_\ell$ recursively with $d(t_\ell, t_{\ell'}) \geq a$ for $\ell' < \ell$. By hypothesis the balls of radius $a$ centered on the previously constructed points do not cover the space if there are $< N$ of them so that the construction continues until we have constructed $N$ points. □

The admissible sequence of partitions witnessing (2.59) will be constructed by recursive application of the following basic principle.

**Lemma 2.4.4.** *Under the conditions of Theorem 2.4.1 consider* $B \subset T$ *with* $\Delta(B) \le 2r^{-j}$ *for a certain* $j \in \mathbb{Z}$ *and* $n \ge 0$. *Let* $m = N_n$. *Then we can find a partition* $(A_\ell)_{\ell \le m}$ *into sets which have* either *of the following properties:*

$$\Delta(A_\ell) \le 2r^{-j-1} , \tag{2.60}$$

or else

$$t \in A_\ell \Rightarrow F(B \cap B(t, 2r^{-j-2})) \le F(B) - c^* 2^{n/2} r^{-j-1} . \tag{2.61}$$

*Proof.* Consider the set

$$C = \{t \in B \; ; \; F(B \cap B(t, 2r^{-j-2})) > F(B) - c^* 2^{n/2} r^{-j-1}\} .$$

We prove that we can cover $C$ by $< m$ balls of radius $r^{-j-1}$, and thus by sets $(A_\ell)_{\ell < m}$ which satisfy (2.60). Indeed, otherwise, by Lemma 2.4.3, we may find $(t_\ell)_{\ell \le m}$ in $C$ with $d(t_\ell, t_{\ell'}) \ge r^{-j-1}$ for $\ell \ne \ell'$. This would contradict (2.55) used for $a = r^{-j-1}$ and for the sets $H_\ell := B \cap B(t_\ell, 2r^{-j-2})$. We then set $A_m = B \setminus C$ which satisfies (2.61). □

Before we start the proof of Theorem 2.4.1 we need the following technical fact which will be used many times: the sum of a geometric series can basically be bounded by either the first or the last term of the series.

**Lemma 2.4.5.** *Consider numbers* $(a_n)_{n \ge 0}$, $a_n \ge 0$, *and assume* $\sup_n a_n < \infty$. *Consider* $\alpha > 1$ *and define*

$$I = \left\{k \ge 0 \; ; \; \forall n \ge 0 \, , n \ne k \, , \; a_n < a_k \alpha^{|k-n|}\right\} . \tag{2.62}$$

*Then*

$$\sum_{n \ge 0} a_n \le \frac{2\alpha}{\alpha - 1} \sum_{k \in I} a_k . \tag{2.63}$$

*Proof.* Let us write $n \prec k$ when $a_k \ge a_n \alpha^{|n-k|}$. This relation is a partial order: if $n \prec k$ and $k \prec p$ then $a_p \ge a_n \alpha^{|p-k|+|k-n|} \ge a_n \alpha^{|p-n|}$, 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 $a_n \le a_k \alpha^{-|n-k|}$, and therefore

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

*Proof of Theorem 2.4.1.* We construct an admissible sequence of partitions $\mathcal{A}_n$, and for $A \in \mathcal{A}_n$ and integer $j_n(A) \in \mathbb{Z}$ with

$$\Delta(A) \le 2r^{-j_n(A)} \,. \tag{2.64}$$

We start with $\mathcal{A}_0 = \{T\}$ and $j_0(T)$ the largest integer $j \in \mathbb{Z}$ with $\Delta(T) \le 2r^{-j_0}$, so that $r^{-j_0} \le r\Delta(T)$. Having constructed $\mathcal{A}_n$ we construct $\mathcal{A}_{n+1}$ as follows. For each $B \in \mathcal{A}_n$, we use Lemma 2.4.4 with $j = j_n(B)$ to split $B$ into sets $(A_\ell)_{\ell \le N_n}$. If $A_\ell$ satisfies (2.60) we set $j_{n+1}(A_\ell) = j_n(B) + 1$ and otherwise we set $j_{n+1}(A_\ell) = j_n(A_\ell)$.

The sequence thus constructed is admissible, since each set $B$ in $\mathcal{A}_n$ is split in at most $N_n$ sets and since $N_n^2 \le N_{n+1}$. We note also be construction that if $B \in \mathcal{A}_n$ and $A \subset B$, $A \in \mathcal{A}_{n+1}$ then

- either $j_{n+1}(A) = j_n(B) + 1$
- or else $j_{n+1}(A) = j_n(B)$ and

$$t \in A \Rightarrow F(B \cap B(t, 2r^{-j_{n+1}(A)-2})) \le F(B) - c^* 2^{n/2} r^{-j_{n+1}(A)-1} \,. \tag{2.65}$$

Let us then fix $t \in T$. We want to prove that

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

We set $j(n) = j_n(A_n(t))$ and $a(n) = 2^{n/2} r^{-j(n)}$. Thus $j(n) \le j(n+1) \le j(n) + 1$. Since $a(0) = r^{-j_0} \le Lr\Delta(T)$, and since $\Delta(A_n(t)) \le 2r^{-j(n)}$, it suffices to show that

$$\sum_{n \ge 1} a(n) \le \frac{Lr}{c^*} F(T) \,. \tag{2.66}$$

We first prove that the sequence $(a(n))$ is bounded. Indeed if $j(n-1) = j(n) - 1$ we have $a(n-1) \le a(n)$ so it suffices to consider the case $j(n-1) = j(n)$ and then $a(n) \le KF(T)$ by (2.65).

Consider then the set $I$ as provided by Lemma 2.4.5 for $\alpha = \sqrt{2}$, so that it suffices to prove

$$\sum_{n \in I} a(n) \le \frac{Lr}{c^*} F(T) \,. \tag{2.67}$$

Obviously $I \subset J = \{n \ge 0, j(n+1) = j(n)+1\}$ because if $j(n+1) = j(n)$ then $a(n+1) = \sqrt{2} a(n)$. Let us enumerate the elements[8] of $J$ as $n_1 < n_2 < \ldots$. Let us observe that $j(n) = j(n_k) + 1$ for $n_k < n \le n_{k+1}$. Let us consider $n_k \in J$ and $n^* = n_{k+1} + 1$, so that

$$j(n^*) = j(n_{k+1} + 1) = j(n_{k+1}) + 1 = j(n_k + 1) + 1 = j(n_k) + 2 \,.$$

[8] We assume here that $J$ is infinite, leaving the necessary simple modifications of the argument when $J$ is finite to the reader.

Thus (2.64) implies $\Delta(A_{n^*}(t)) \le 2r^{-j(n^*)} = 2r^{-j(n_k)-2}$ and therefore if $n = n_k - 1$,

$$A_{n^*}(t) \subset B \cap B(t, 2r^{-j_{n+1}(A_{n+1}(t))-2}) \; . \tag{2.68}$$

Consider now $n_k \in I$, so that $n_k \ge 1$. Then $j(n_k - 1) = j(n_k)$ for otherwise $a(n_k) = (\sqrt{2}/r)a(n_k - 1)$, contradicting the definition of $I$. We may then use (2.65) for $n = n_k - 1$, $B = A_n(t)$, $A = A_{n_k}(t)$ to conclude by (2.68)

$$a(n_k) \le \frac{Lr}{c^*}(F(A_n(t)) - F(A_{n^*}(t))) \; . \tag{2.69}$$

Let us set $f(n) = F(A_n(t))$, so that $f(0) = F(T)$ and $f(n+1) \le f(n)$ since $A_{n+1}(t) \subset A_n(t)$. For $k \ge 2$ and since $n = n_k - 1 \ge n_{k-1}$ and $n^* \le n_{k+2}$ we deduce from (5.47) that

$$a(n_k) \le \frac{Lr}{c^*}(f(n) - f(n^*)) \le \frac{Lr}{c^*}(f(n_{k-1}) - f(n_{k+2})) \; .$$

Summation of these inequalities over the values of $k$ for which $n_k \in I$ together with the fact that $a(n_1) \le Lrf(0)/c^*$ concludes the proof of (2.67) and of the theorem. □

**Exercise 2.4.6.** We say that a sequence $(F_n)_{n\ge 0}$ of functionals on $(T, d)$ satisfies the growth condition with parameters $r \ge 4$ and $c^* > 0$ if

$$\forall n \ge 0 \; , \; F_{n+1} \le F_n$$

and if for any integer $n \ge 0$ and any $a > 0$ the following holds true, where $m = N_n$. For each collection of subsets $H_1, \ldots, H_m$ of $T$ that are $(a, r)$-separated it holds

$$F_n\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge c^* a 2^{n/2} + \min_{\ell\le m} F_{n+1}(H_\ell) \; . \tag{2.70}$$

Prove that then

$$\gamma_2(T, d) \le \frac{Lr}{c^*} F_0(T) + Lr\Delta(T) \; . \tag{2.71}$$

Hint: copy the previous arguments by replacing everywhere $F(A)$ by $F_n(A)$ when $A \in \mathcal{A}_n$.

**Proposition 2.4.7.** *Consider a metric space $(T, d)$, and for $n \ge 0$, consider subsets $T_n$ of $T$ with* $\operatorname{card} T_0 = 1$ *and* $\operatorname{card} T_n \le N_n$ *for $n \ge 1$. Consider a number $S$ and let*

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

*Then $\gamma_2(U, d) \le LS$.*

*Proof.* For $H \subset U$ we define $F(H) = \inf \sup_{t \in H} \sum_{n \geq 0} 2^{n/2} d(t, V_n)$ where the infimum is taken over all choices of $V_n \subset T$ with $\operatorname{card} V_n \leq N_n$. It is important here not to assume that $V_n \subset H$ to ensure that $F$ is increasing. We then prove that $F$ satisfies the growth condition by an argument very similar to that of Proposition 2.3.6. □

**Exercise 2.4.8.** A rather obvious consequence of the previous result is that for any metric space $(T, d)$ we have

$$\gamma_2(T, d) \leq L \sum_{n \geq 0} 2^{n/2} e_n(T) . \tag{2.72}$$

Find a simple direct proof. Hint: you do have to construct the partitions.

**Exercise 2.4.9.** Use (2.45) and Exercise 2.2.15 (d) to prove that if $T \subset \mathbb{R}^m$ then

$$\sum_{n \geq 0} 2^{n/2} e_n(T) \leq L \log(m+1) \gamma_2(T, d) . \tag{2.73}$$

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

One reason to present the scheme of proof of Theorem 2.4.1 exactly in the present form is that it may help the reader to follow the proof of the Bednorz-Latala theorem, Theorem 5.2.5, which is one of the highlights of this work.

A slightly different partitioning scheme has recently been discovered by R. van Handel [150], and we describe a variant of it now. We consider a metric space $(T, d)$ and an integer $r \geq 8$. We assume that for $j \in \mathbb{Z}$ we are given a function $s_j(t) \geq 0$ on $T$. We make the following assumption.

$$\begin{aligned} &\text{For each subset } A \text{ of } T, \text{ each } j \in \mathbb{Z} \text{ with } \Delta(A) \leq 2r^{-j} \text{ and each} \\ &n \geq 1, \text{ then either } e_n(A) \leq r^{-j-1} \textit{ or else } \text{ there exists } t \in A \\ &\text{with } s_j(t) \geq 2^{n/2} r^{-j-1} . \end{aligned} \tag{2.74}$$

We will show later how to construct such functions $s_j(t)$ using a functional which satisfies the growth condition. Other constructions are considered in [150].

**Theorem 2.4.10.** *Assume that (2.74) holds. Then*

$$\gamma_2(T, d) \leq Lr \Big( \Delta(T) + \sup_{t \in T} \sum_{j \in \mathbb{Z}} s_j(t) \Big) . \tag{2.75}$$

The right-hand side is the supremum over $t$ of a sum of terms. It need not always be the same terms which will contribute the most for different values of $t$.

*Proof of Theorem 2.4.10.* Consider the largest $j_0 \in \mathbb{Z}$ with $\Delta(T) \le 2r^{-j_0}$, so that $2r^{-j_0} \le r\Delta(T)$. We construct by induction an increasing sequence of partitions $\mathcal{A}_n$ with $\operatorname{card}\mathcal{A}_n \le N_n$, and for $A \in \mathcal{A}_n$ we construct an integer $j_n(A) \in \mathbb{Z}$ with $\Delta(A) \le 2r^{-j_n(A)}$. We start with $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$ and $j_0(T) = j_1(T) = j_0$.

Once $\mathcal{A}_n$ has been constructed, we further split every element $B \in \mathcal{A}_n$. The idea is to first split $B$ into sets which are basically level sets for the function $s_j(t)$ in order to achieve the crucial relation (2.79) below, and then to further split each of these sets according to its metric entropy. More precisely, let $S = \sup_{t\in T} \sum_{j\in\mathbb{Z}} s_j(t)$ and set $j = j_n(B)$. We define the sets $A_k$ for $1 \le k \le n$ by setting for $k < n$

$$A_k = \{t \in A\ ;\ 2^{-k}S < s_j(t) \le 2^{-k+1}S\}\ , \tag{2.76}$$

and

$$A_n = \{t \in A\ ;\ s_j(t) \le 2^{-n+1}S\}\ . \tag{2.77}$$

The purpose of this construction is to ensure the following:

$$k \le n\ ;\ t, t' \in A_k \Rightarrow s_j(t') \le 2(s_j(t) + 2^{-n}S)\ . \tag{2.78}$$

This is obvious if one distinguishes the cases $k < n$ and $k = n$. For each set $A_k$, $k \le n$ we use the following procedure.

- If $e_{n-1}(A_k) \le r^{-j-1}$ we split $A_k$ into $N_{n-1}$ pieces of diameter $\le 2r^{-j-1}$. We decide that each of these pieces $A$ is an element of $\mathcal{A}_{n+1}$, for which we set $j_{n+1}(A) = j + 1$.
- Otherwise we decide that $A_k \in \mathcal{A}_{n+1}$ and we set $j_{n+1}(A_k) = j$. From (2.74) there exist $t' \in A_k$ for which $s_j(t') \ge 2^{(n-1)/2}r^{-j-1}$. Then by (2.78) we have
$$\forall t \in A_k\ ;\ 2^{(n-1)/2}r^{-j-1} \le 2(s_j(t) + 2^{-n}S)\ . \tag{2.79}$$

In summary, if $B \in \mathcal{A}_n$ and $A \in \mathcal{A}_{n+1}$, $A \subset B$ then

- either $j_{n+1}(A) = j_n(B) + 1$
- or else $j_{n+1}(A) = j_n(B)$ and, from (2.79)

$$\forall t \in A\ ;\ 2^{(n-1)/2}r^{-j_{n+1}(A)-1} \le 2(s_{j_n(B)}(t) + 2^{-n}S)\ . \tag{2.80}$$

This completes the construction. Since $\operatorname{card}\mathcal{A}_{n+1} \le N_{n-1}nN_n \le N_{n+1}$ the sequence $(\mathcal{A}_n)$ is admissible. Next, we fix $t \in T$. We set $j_n = j_n(A_n(t))$, and we observe that by construction that $j_n \le j_{n+1} \le j_n + 1$. We set $a(n) = 2^{n/2}r^{-j_n(t)}$ and we prove that

$$\sum_{n\ge 0} a(n) \le LrS\ . \tag{2.81}$$

This completes the proof since $\Delta(A_n(t)) \le 2r^{-j_n(A_n(t))}$. Consider then the set $I$ provided by Lemma 2.4.5, so that since $r^{-j_0} \le 2r\Delta(T)$ it suffices to prove that

$$\sum_{n\in I\setminus\{0\}} a(n) \le LrS\ . \tag{2.82}$$

For $n \in I \setminus \{0\}$, it holds $j_{n-1} = j_n < j_{n+1}$ (for otherwise this contradicts the definition of $I$). In particular the integers $j_n$ for $n \in I$ are all different so that $\sum_{n\ge 0} s_{j_n}(t) \le S$. Using (2.80) for $n-1$ instead of $n$ and since $j(n-1) = j(n)$ we get

$$a(n) \le Lr(s_{j_n}(t) + 2^{-n}S)\ ,$$

and summing these relations we have obtained the desired result. □

The following connects Theorems 2.4.1 and 2.4.10.

**Proposition 2.4.11.** *Assume that the functional $F$ satisfies the growth condition with parameters $r$ and $c^*$. Then the functions*

$$s_j(t) = \frac{1}{c^*}\big(F(B(t, 2r^{-j+1}) - F(B(t, 2r^{-j-2}))\big)$$

*satisfy (2.74).*

*Proof.* Consider a subset $A$ of $T$, $j \in \mathbb{Z}$ with $\Delta(A) \le 2r^{-j}$ and $n \ge 1$. Let $m = N_n$. If $e_n(A) > r^{-j-1}$ then by Lemma 2.4.3 we may find $(t_\ell)_{\ell\le m}$ in $A$ with $d(t_\ell, t_{\ell'}) \ge r^{-j-1}$ for $\ell \ne \ell'$. Consider the set $H_\ell = B(t_\ell, 2r^{-j-2})$ so that by (2.55) used for $a = r^{-j-1}$ it holds that

$$\min_{\ell\le m} F(H_\ell) \le c^* r^{-j-1} 2^{n/2} + F\Big(\bigcup_{\ell\le m} H_\ell\Big)\ .$$

Let us now consider $\ell' \le m$ such that $F(H_{\ell'})$ achieves the minimum in the left-hand side. Observe that $H_\ell \subset B(t_{\ell'}, 2r^{-j+1})$ for each $\ell$, so that $F(\bigcup_{\ell\le m} H_\ell) \le F(B(t_{\ell'}, 2r^{-j+1}))$, and then $s_j(t_{\ell'}) \ge 2^{n/2} r^{-j-1}$. □

Despite the fact that the proof of Theorem 2.4.10 is somewhat simpler than the proof of Theorem 2.4.1, in the various generalizations of this principle we will mostly follow the scheme of proof of Theorem 2.4.1 for a simple reason: it should help the reader that these various generalizations follow a common pattern, and it not clear at this point whether the method of Theorem 2.4.10 can be adapted to the proof of Theorem 5.2.5.

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

**Lemma 2.4.12.** *Consider $\alpha > 0$, an integer $\tau \ge 0$ and an increasing sequence of partitions $(\mathcal{B}_n)_{n\ge 0}$ with $\operatorname{card}\mathcal{B}_n \le N_{n+\tau}$. Let*

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

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

$$\sup_{t\in T}\sum_{n\geq 0} 2^{n/\alpha}\Delta(A_n(t)) \leq 2^{\tau/\alpha}(S + K(\alpha)\Delta(T)) \ . \tag{2.83}$$

Of course (for the last time) here $K(\alpha)$ denotes a number depending on $\alpha$ 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 2.4.13.** Prove that (2.83) might fail if one replaces the right-hand side by $K(\alpha,\tau)S$. Hint: $S$ does not control $\Delta(T)$.

## 2.5 Gaussian Processes: The Majorizing Measure Theorem

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

$$d(s,t) = \left(\mathsf{E}(X_s - X_t)^2\right)^{1/2} \ . \tag{2.84}$$

Recall the functional $\gamma_2$ of Definition 2.2.20.

**Theorem 2.5.1.** (The Majorizing Measure Theorem.) *For some universal constant* $L$ *it holds*

$$\frac{1}{L}\gamma_2(T,d) \leq \mathsf{E}\sup_{t\in T} X_t \leq L\gamma_2(T,d) \ . \tag{2.85}$$

The reason for the name is explained in Section 4.1. We will meditate on this statement in Section 2.7.

The right-hand side inequality in (2.85) follows from Theorem 2.2.23. To prove the lower bound we will use Theorem 2.4.1 and the functional

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

For this we need to prove that this functional satisfies the growth condition with $c^*$ a universal constant and to bound $\Delta(T)$. We strive to give a proof that relies on general principles, and lends itself to generalizations.

**Lemma 2.5.2.** (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*

$$\mathsf{E}\sup_{p\le m} X_{t_p} \ge \frac{a}{L_1}\sqrt{\log m}\,. \tag{2.86}$$

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

**Exercise 2.5.3.** Prove that Lemma 2.5.2 is equivalent to the following statement. If $(X_t)_{t\in T}$ is a Gaussian process, and $d$ is the canonical distance, then

$$e_n(T,d) \le L2^{-n/2}\mathsf{E}\sup_{t\in T} X_t\,. \tag{2.87}$$

Compare with Exercise 2.2.22.

A proof of Sudakov minoration may be found in [61], 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 2.5.4.** Use Lemma 12.2.6 and Lemma 16.3.32 to prove that for a Gaussian process $(X_t)_{t\in T}$ we have $e_n(T,d) \le L2^{-n/2}\mathsf{E}\sup_{t\in T}|X_t|$. Then use Exercise 2.2.2 to deduce (2.87).

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 (2.86) means that the bound (2.14) is of the correct order in this situation.

**Exercise 2.5.5.** Prove (2.86) when the r.v.s $X_{t_p}$ are independent. Hint: use Exercise 2.2.7 (b).

**Exercise 2.5.6.** 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 (5.10) 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 $X_{t_\ell}$ are independent, but find examples showing that this naive approach does not work in general to prove (2.86).

The following is a very important property of Gaussian processes, and one of the keys to Theorem 2.5.1. It is a facet of the theory of concentration of measure, a leading idea of modern probability theory. We refer the reader to [60] to learn about this.

**Lemma 2.5.7.** *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*

$$\mathsf{P}\Big(\Big|\sup_{t\in U} X_t - \mathsf{E}\sup_{t\in U} X_t\Big| \ge u\Big) \le 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big)\,. \tag{2.88}$$

In words, 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 2.5.8.** 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 $\sigma$, e.g. the variance of $\sup_t X_t$ is about $\sigma^2$. 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 $\sup_{t\in T} X_t$ has fluctuations of order 1 whatever the value of $n$.

**Proposition 2.5.9.** *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*

$$\mathsf{E}\sup_{t\in H} X_t \ge \frac{a}{L_1}\sqrt{\log m} - L_2\sigma\sqrt{\log m} + \min_{\ell\le m} \mathsf{E}\sup_{t\in H_\ell} X_t . \tag{2.89}$$

When $\sigma \le a/(2L_1L_2)$, (2.89) implies

$$\mathsf{E}\sup_{t\in H} X_t \ge \frac{a}{2L_1}\sqrt{\log m} + \min_{\ell\le m} \mathsf{E}\sup_{t\in H_\ell} X_t , \tag{2.90}$$

which can be seen as a generalization of (2.86).

*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 (2.88) 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

$$\mathsf{P}(V \ge u) \le 2m\exp\Big(-\frac{u^2}{2\sigma^2}\Big) , \tag{2.91}$$

and (2.12) 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

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

so that

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

We then take expectations and use (2.86). □

**Exercise 2.5.10.** Prove that (2.90) 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 2.5.11.** Prove that

$$\mathsf{E} \sup_{t\in H} X_t \leq La\sqrt{\log m} + \max_{\ell\leq m} \mathsf{E} \sup_{t\in H_\ell} X_t \ . \tag{2.92}$$

Try to find improvements on this bound. Hint: peek at (16.76) below.

*Proof of Theorem 2.5.1.* We fix $r \geq 2L_1L_2$. To prove the growth condition for the functional $F$ we simply observe that (2.90) implies that (2.55) holds for $c^* = 1/L$. Using Theorem 2.4.1 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 2.5.1 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 2.5.1 is that there is simply no other way to bound a Gaussian process than to control the quantity $\gamma_2(T, d)$. However, to control this quantity in a specific situation, we must in some way gain understanding of the underlying geometry.

The following is a noteworthy consequence of Theorem 2.5.1.

**Theorem 2.5.12.** *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 (2.84) 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 (2.50) with the left-hand side of (2.85). □

Let us now turn to a simple (and classical) example that illustrates well the difference between Dudley's bound (2.39) and the bound (2.33). 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 2.2.17. Consider an independent sequence $(g_i)_{i\ge1}$ of standard Gaussian r.v.s and for $i \ge 2$ set

$$X_i = \frac{g_i}{\sqrt{\log i}} \,. \tag{2.93}$$

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

$$\frac{1}{\sqrt{\log(\min(p,q))}} \le d(p,q) \le \frac{2}{\sqrt{\log(\min(p,q))}} \,. \tag{2.94}$$

Consider $1 \le n \le s-2$ and $T_n \subset T$ with $\operatorname{card} T_n = N_n$. There exists $p \le N_n + 1$ with $p \notin T_n$, so that (2.94) implies $d(p, T_n) \ge 2^{-n/2}/L$ (where the distance from a point to a set is defined in (2.34)). This proves that $e_n(T) \ge 2^{-n/2}/L$. Therefore

$$\sum_n 2^{n/2} e_n(T) \ge \frac{s-2}{L} \,. \tag{2.95}$$

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

$$d(p, T_n) \le d(p, N_s) \le L2^{-m/2}$$

by (2.94) and since $p \ge N_m$ and $N_s \ge N_m$. Hence we have

$$\sum_n 2^{n/2} d(p, T_n) \le \sum_{n\le m} L2^{n/2}2^{-m/2} \le L \,. \tag{2.96}$$

Comparing (2.95) and (2.96) proves that the bound (2.39) is worse than the bound (2.33) by a factor about $s$.

**Exercise 2.5.13.** Prove that when $T$ is finite, the bound (2.39) cannot be worse than (2.33) by a factor greater than about $\log\log \operatorname{card} T$. This shows that the previous example is in a sense extremal. Hint: use $2^{n/2} e_n(T) \le L\gamma_2(T, d)$ and $e_n(T) = 0$ if $N_n \ge \operatorname{card} T$.

**Exercise 2.5.14.** Prove that the estimate (2.73) is essentially optimal. Hint: if $m \ge \exp(10s)$, one can produce the situation of Example 2.2.17 (b) inside $\mathbb{R}^m$.

It follows from (2.96) and (2.33) that $\mathsf{E} \sup_{i\ge1} X_i < \infty$. A simpler proof of this fact is given in Proposition 2.6.2 below.

## 2.6 Gaussian Processes as Subsets of Hilbert Space

In this section we learn to think of a Gaussian process as *a subset of Hilbert space*. This will reveal our lack of understanding of basic geometric questions. Generalizing the idea of Exercise 2.2.16, 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

$$\|t\| = \|t\|_2 = \Big(\sum_{i\geq 1} t_i^2\Big)^{1/2} . \tag{2.97}$$

To each $t$ in $\ell^2$ we associate a Gaussian r.v.

$$X_t = \sum_{i\geq 1} t_i g_i \tag{2.98}$$

(the series converges in $L^2(\Omega)$). In this manner, for each subset $T$ of $\ell^2$ we can consider the Gaussian process $(X_t)_{t\in T}$. The distance induced on $T$ by the process coincides with the distance of $\ell^2$ since from (2.98) 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 $\Omega$ is the underlying probability space. The linear span of the variables $Y_t$ in $L^2(\Omega)$ is then separable and we may identify it with $\ell^2$ by choosing an orthonormal basis.

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

**Theorem 2.6.1.** *For a subset $T$ of $\ell^2$, we have*

$$\gamma_2(\operatorname{conv} T) \leq L\gamma_2(T) . \tag{2.99}$$

*Proof.* To prove (2.99) we observe that since $X_{a_1t_1+a_2t_2} = a_1X_{t_1} + a_2X_{t_2}$ we have

$$\sup_{t\in\operatorname{conv} T} X_t = \sup_{t\in T} X_t . \tag{2.100}$$

We then use (2.85) to write

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

A basic problem is that it is absolutely not obvious how to construct an admissible sequence of partitions on $\operatorname{conv} T$ witnessing (2.100). We will focus on a particularly striking instance of this problem. We recall the $\ell^2$ norm $\|\cdot\|$ of (2.97). Here is a simple fact.

**Proposition 2.6.2.** *Consider a set $T = \{t_k \,;\, k \geq 1\}$ where*

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

*Then $\mathsf{E}\sup_{t \in T} X_t \leq L$ and thus $\mathsf{E}\sup_{t \in \operatorname{conv} T} X_t \leq L$ by (2.100).*

*Proof.* We have

$$\mathsf{P}\Big(\sup_{k \geq 1} |X_{t_k}| \geq u\Big) \leq \sum_{k \geq 1} \mathsf{P}(|X_{t_k}| \geq u) \leq \sum_{k \geq 1} 2\exp\Big(-\frac{u^2}{2}\log(k+1)\Big) \quad (2.101)$$

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 (2.101) is at most $L\exp(-u^2/L)$. □

**Exercise 2.6.3.** Deduce Proposition 2.6.2 from (2.33). Hint: see Exercise 2.2.17 (a).

**Exercise 2.6.4.** Deduce from Proposition 2.6.2 that if $T$ is a subset of the unit ball of $L^2$ then

$$\gamma_2(\operatorname{conv} T) \leq L\sqrt{\log \operatorname{card} T}\,. \quad (2.102)$$

The simple proof of (2.102) hides the fact that it is a near miraculous result.

**Research problem 2.6.5.** Give a geometrical proof of (2.102).

The issue is that 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 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 [65].

**Research problem 2.6.6.** Consider a 2-smooth Banach space, and the distance $d$ induced by its norm. Is it true that for each subset $T$ of its unit ball $\gamma_2(\operatorname{conv} T, d) \leq K\sqrt{\log \operatorname{card} T}$? More generally, is it true that for each finite subset $T$ one has $\gamma_2(\operatorname{conv} T, d) \leq K\gamma_2(T, d)$? (Here $K$ may depend on the Banach space, but not on $T$.)

**Research problem 2.6.7.** Still more generally, is it true that for a finite subset $T$ of a $q$-smooth Banach space, one has $\gamma_q(\operatorname{conv} T) \leq K\gamma_q(T)$?

Even when the Banach space is $\ell^p$, I do not know the answer to these problems (unless $p = 2$!). (The Banach space $\ell^p$ 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 $\ell^p$. 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 2.2.16.

The following shows that the situation of Proposition 2.6.2 is in a sense generic.

**Theorem 2.6.8.** *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*

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

*Proof.* By Theorem 2.5.1 we can find an admissible sequence $(\mathcal{A}_n)$ of $T$ with

$$\forall t \in T\ , \sum_{n \geq 0} 2^{n/2}\Delta(A_n(t)) \leq L\mathsf{E}\sup_{t \in T} X_t := S\ . \tag{2.103}$$

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 $N_n N_{n-1} \leq N_{n+1}$ elements. Let $U = \bigcup_{k \geq 1} U_k$. Then $U$ contains at most $N_{n+2}$ elements of norm $\geq 2^{-n/2}$. We enumerate $U$ as $\{t_k ; k = 1, \ldots\}$ 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

$$t = \sum_{1 \leq n \leq m} \pi_n(t) - \pi_{n-1}(t) = \sum_{1 \leq n \leq m} a_n(t) u_n(t)\ , \tag{2.104}$$

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 (2.104) that

$$t \in 2S\,\mathrm{conv}(U \cup \{0\}) .$$

This concludes the proof. □

**Exercise 2.6.9.** What is the purpose of the condition $0 \in T$?

**Exercise 2.6.10.** 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 $\ell^2$, 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 2.6.8 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 2.6.11.** 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 $\ell^2$, with the following properties:

$$\forall\, k \ge 1\ ,\ \|t_k\|\sqrt{\log(k+1)} \le L\mathsf{E}\sup_{t\in T} X_t\ , \tag{2.105}$$

$$\|t_k\| \le L\delta\ , \tag{2.106}$$

$$T \subset \overline{\mathrm{conv}}\{t_k\ ;\ k \ge 1\}\ , \tag{2.107}$$

where $\overline{\mathrm{conv}}$ denotes the closed convex hull. Hint: copy the proof of Theorem 2.6.8, 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 next exercise is inspired by the paper [9] of S Artstein. It is more elaborate, and may be omitted at first reading. A Bernoulli r.v. $\varepsilon$ is such that $\mathsf{P}(\varepsilon = \pm 1) = 1/2$.[9]

**Exercise 2.6.12.** 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

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

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

[9] One must distinguish Bernoulli r.v.s $\varepsilon_i$ from positive numbers $\epsilon_k$!

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

We want to prove that there exists a number $L$ such that if

$$q \ge \delta^{-1} \mathsf{E} \sup_{t\in T} \sum_{i\le N} g_i t_i \ , \tag{2.108}$$

then with high probability

$$U_q(T) \subset B(0, L\delta\sqrt{q}) \ . \tag{2.109}$$

(a) Use the subgaussian inequality (5.1.1) to prove that if $\|x\| = 1$, then

$$\mathsf{E} \exp\bigg(\frac{1}{4}\Big(\sum_{i\le N} \varepsilon_{i,p} x_i\Big)^2\bigg) \le L \ . \tag{2.110}$$

(b) Use (2.110) and independence to prove that for $x \in \mathbb{R}^n$ and $v \ge 1$,

$$\mathsf{P}(\|U_q(x)\| \ge Lv\sqrt{q}\|x\|) \le \exp(-v^2 q) \ . \tag{2.111}$$

(c) Use (2.111) to prove that with probability close to 1, for each of the vectors $t_k$ of Exercise 2.6.11 one has $\|U_q(t_k)\| \le L\delta\sqrt{q}$ and conclude.

## 2.7 Dreams

We may reformulate the inequality (2.85)

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

of Theorem 2.5.1 by the statement

$$\textit{Chaining suffices to explain the size of a Gaussian process.} \tag{2.112}$$

We simply mean that the “natural” chaining bound for the size of a Gaussian process (i.e. the right-hand side inequality in (2.85)) is of correct order, *provided* one uses the best possible chaining. This is what the left-hand side of (2.85) shows. We may dream of removing the word “Gaussian” in that statement. The desire to achieve this lofty goal in as many situations as possible motivates a lot of the rest of the book.

Besides the generic chaining, we have found in Theorem 2.6.8 another optimal way to bound Gaussian processes: to put them into the convex hull of a “small” process. Since we do not really understand the geometry of going from a set to it convex hull it is better for the time being to consider this method as somewhat distinct from the generic chaining. Let us try to formulate it in a way which is suitable for generalizations. Given a countable set $\mathcal{V}$ of r.v.s let us define the (possibly infinite) quantity

$$S(\mathcal{V}) = \inf\Big\{ S > 0 \; ; \; \int_S^\infty \sum_{V \in \mathcal{V}} \mathsf{P}(|V| > u) \mathrm{d}u \le S \Big\} \; . \tag{2.113}$$

**Lemma 2.7.1.** *It holds that*

$$\mathsf{E} \sup_{V \in \operatorname{conv} \mathcal{V}} |V| \le 2S(\mathcal{V}) \; . \tag{2.114}$$

*Proof.* We combine (2.6) with the fact that for $S > S(\mathcal{V})$ it holds

$$\int_0^\infty \mathsf{P}\Big( \sup_{V \in \operatorname{conv} \mathcal{V}} |V| \ge u \Big) \le S + \int_S^\infty \sum_{|V| \in \mathcal{V}} \mathsf{P}(V > u) \mathrm{d}u \le 2S \; . \qquad \square$$

Thus (2.114) provides a method to bound stochastic processes. This method may look childish, but for Gaussian processes, the following reformulation of Theorem 2.6.8 shows that it is in fact optimal.

**Theorem 2.7.2.** *Consider a countable set $T$. Consider a Gaussian process $(X_t)_{t \in T}$ and assume that $X_{t_0} = 0$ for some $t_0 \in T$. Then there exists a countable set $\mathcal{V}$ of Gaussian r.v.s, each of which is a multiple of the difference of two variables $X_t$ with*

$$\forall t \in T \; ; \; X_t \in \operatorname{conv} \mathcal{V} \; , \tag{2.115}$$

$$S(\mathcal{V}) \le L \mathsf{E} \sup_{t \in T} X_t \; . \tag{2.116}$$

The proof of Theorem 2.7.2 is nearly obvious by using (2.101) to bound $S(\mathcal{V})$ for the set $\mathcal{V}$ consisting of the variables $X_{t_k}$ for the sequence $(t_k)$ constructed in Theorem 2.6.8. We may dream of proving statements such as Theorem 2.7.2 for many classes of processes.

Also worthy of detailing is another remarkable geometric consequence of Theorem 2.6.8 in a somewhat different direction. Consider an integer $N$ and let us provide $\ell^2_N$ ($= \mathbb{R}^N$ provided with the Euclidean distance) with the canonical Gaussian measure $\mu$, i.e. the law of the i.i.d. Gaussian sequence $(g_i)_{i \le N}$. Let us view an element $t$ of $\ell^2_N$ as a function on $\ell^2_N$ by the canonical duality, so $t$ is a r.v. $Y_t$ on the probability space $(\ell^2_N, \mu)$. 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 $\ell^2_N$, and assume that $T \subset \operatorname{conv}\{t_k ; k \ge 1\}$. Then for any $v > 0$ we have

$$\Big\{ \sup_{t \in T} t \ge v \Big\} \subset \bigcup_{k \ge 1} \{ t_k \ge v \} \; . \tag{2.117}$$

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 $k \geq 1$ and a certain $S$ we have $\|t_k\|\sqrt{\log(k+1)} \leq S$. Then (2.101) implies that for $u \geq 2$

$$\sum_{k\geq 1} \mu(\{t_k \geq Su\}) \leq L\exp(-u^2/L) .$$

Theorem 2.6.8 implies that may take $S \leq L\mathsf{E}\sup_t X_t$. Therefore for $v \geq L\mathsf{E}\sup_t X_t$, the fact that the set in the left-hand side of (2.117) is small (in the sense of probability) may 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.

We may dream that something similar occurs in many other settings.

## 2.8 A First Look at Ellipsoids

We have illustrated the gap between Dudley's bound (2.39) and the sharper bound (2.33), using the examples (2.43) and (2.93). These examples might look artificial, but here we demonstrate that the gap between Dudley's bound (2.39) and the generic chaining bound (2.33) already exists for *ellipsoids* in Hilbert space. Truly understanding ellipsoids will be fundamental in several subsequent questions, such as the matching theorems of Chapter 3.

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

$$\mathcal{E} = \left\{ t \in \ell^2\ ;\ \sum_{i\geq 1} \frac{t_i^2}{a_i^2} \leq 1 \right\} . \tag{2.118}$$

**Proposition 2.8.1.** *We have*

$$\frac{1}{L}\left(\sum_{i\geq 1} a_i^2\right)^{1/2} \leq \mathsf{E}\sup_{t\in\mathcal{E}} X_t \leq \left(\sum_{i\geq 1} a_i^2\right)^{1/2} . \tag{2.119}$$

*Proof.* The Cauchy-Schwarz inequality implies

$$Y := \sup_{t\in\mathcal{E}} X_t = \sup_{t\in\mathcal{E}} \sum_{i\geq 1} t_i g_i \leq \left(\sum_{i\geq 1} a_i^2 g_i^2\right)^{1/2} . \tag{2.120}$$

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 (2.119) follows from the Cauchy-Schwarz inequality:

$$\mathsf{E}Y \leq (\mathsf{E}Y^2)^{1/2} = \left(\sum_{i\geq 1} a_i^2\right)^{1/2} . \tag{2.121}$$

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,

$$\mathsf{E}X_t^2 = \sum_i t_i^2 \leq \max_i a_i^2 \sum_j \frac{t_j^2}{a_j^2} \leq \sigma^2 \ . \tag{2.122}$$

Then (2.88) 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 2.5.1,

$$\gamma_2(\mathcal{E}) \leq L\Big(\sum_{i\geq 1} a_i^2\Big)^{1/2} \ . \tag{2.123}$$

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 3.2.11, 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 (2.119) as

$$\frac{1}{L}\Big(\sum_{n\geq 0} 2^n a_{2^n}^2\Big)^{1/2} \leq \mathsf{E}\sup_{t\in\mathcal{E}} X_t \leq \Big(\sum_{n\geq 0} 2^n a_{2^n}^2\Big)^{1/2} \ . \tag{2.124}$$

Proposition 2.8.1 describes the size of ellipsoids with respect to Gaussian processes. Our next result describes their size with respect to Dudley's entropy bound (2.37).

**Proposition 2.8.2.** *We have*

$$\frac{1}{L}\sum_{n\geq 0} 2^{n/2} a_{2^n} \leq \sum_{n\geq 0} 2^{n/2} e_n(\mathcal{E}) \leq L\sum_{n\geq 0} 2^{n/2} a_{2^n} \ . \tag{2.125}$$

The right-hand sides in (2.124) and (2.125) 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 3.2.

**Exercise 2.8.3.** 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 (2.73).

The proof of (2.125) hinges on ideas which are at least 50 years old, and which relate to the methods of Exercise 2.2.15. 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 (2.40).

**Lemma 2.8.4.** *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

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

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 (2.41).

**Lemma 2.8.5.** *We have*

$$e_{n+3}(\mathcal{E}) \leq 3\max_{k\leq n}(a_{2^k}2^{k-n}) \ . \tag{2.126}$$

*Proof.* We keep the notation of the proof of Lemma 2.8.4. First we show that

$$e_{n+3}(\mathcal{E}) \leq e_{n+3}(\mathcal{E}_n) + a_{2^n} \ . \tag{2.127}$$

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

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

so that $(\sum_{i>2^n} t_i^2)^{1/2} \le a_{2^n}$ and, viewing $\mathcal{E}_n$ as a subset of $\mathcal{E}$, we have $d(t, \mathcal{E}_n) \le a_{2^n}$. Thus if we cover $\mathcal{E}_n$ by certain balls with radius $\epsilon$, the balls with the same centers but radius $\epsilon + a_{2^n}$ cover $\mathcal{E}$. This proves (2.127).

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

$$\text{any two points of } Z \text{ are at mutual distance} \ge 2\epsilon \tag{2.128}$$

$$\operatorname{card} Z \text{ is as large as possible under (2.128).} \tag{2.129}$$

Then by (2.129) the balls centered at points of $Z$ and with radius $\le 2\epsilon$ cover $\mathcal{E}_n$. Thus

$$\operatorname{card} Z \le N_{n+3} \Rightarrow e_{n+3}(\mathcal{E}_n) \le 2\epsilon \ . \tag{2.130}$$

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

$$\operatorname{card} Z \ \mathrm{Vol}(\epsilon B) \le \mathrm{Vol}(\mathcal{E}_n + \epsilon B) \ . \tag{2.131}$$

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

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

we have

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

Therefore

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

and comparing with (2.131) yields

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

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

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

since $\sum_{i\geq 0} i2^{-i} = 4$.

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

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

so that $e_{n+3}(\mathcal{E}_n) \leq 2\epsilon$. The conclusion follows from (2.127). □

*Proof of Proposition 2.8.2.* We have, using (2.126)

$$\begin{aligned}\sum_{n\geq 3} 2^{n/2} e_n(\mathcal{E}) = \sum_{n\geq 0} 2^{(n+3)/2} e_{n+3}(\mathcal{E}) &\leq L \sum_{n\geq 0} 2^{n/2} \Big(\sum_{k\leq n} 2^{k-n} a_{2^k}\Big) \\ &\leq L \sum_{k\geq 0} 2^k a_{2^k} \sum_{n\geq k} 2^{-n/2} \leq L \sum_{k\geq 0} 2^{k/2} a_{2^k} \ .\end{aligned}$$

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

## 2.9 Continuity of Gaussian Processes

*By far* the most important result concerning continuity of Gaussian processes is Dudley's bound (1.19). However since the finiteness of the right hand side of (1.19) is not necessary for the Gaussian process to be continuous, there are situations where this bound is not appropriate.[10] In the present section we show that a suitable form of the generic chaining allows to capture the exact modulus of continuity of a Gaussian process with respect to its canonical distance in full generality. Not surprisingly, the modulus of continuity is closely related to the rate at which the series $\sum_n 2^{n/2} \Delta(A_n(t))$ converges uniformly on $T$ for a suitable admissible sequence $(\mathcal{A}_n)$. Our first result shows how to obtain a modulus of continuity using the generic chaining.

**Lemma 2.9.1.** *Consider a metric space $(T,d)$ and a process $(X_t)_{t\in T}$ which satisfies the increment condition (2.1):*

$$\forall u > 0\ ,\ \mathsf{P}(|X_s - X_t| \geq u) \leq 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right) \ . \qquad (2.1)$$

*Assume that there exists a sequence $(T_n)$ of subsets of $T$ with $\operatorname{card} T_n \leq N_n$ such that for certain integer $m$, and a certain number $B$ one has*

$$\sup_{t\in T} \sum_{n\geq m} 2^{n/2} d(t, T_n) \leq B \ . \qquad (2.132)$$

[10] In practice however, as of today the Gaussian processes for which continuity is important can be handled through Dudley's bound, while for those which cannot be handled through this bound it is boundedness which matters. For this reason, the considerations of the present section are of purely theoretical interest and may be skipped at first reading.

*Consider* $\delta > 0$. *Then, for any* $u \geq 1$, *with probability* $\geq 1 - \exp(-u^2 2^m)$ *we have*

$$\forall\, s,t \in T\ ,\ d(s,t) \leq \delta \Rightarrow |X_s - X_t| \leq Lu(2^{m/2}\delta + B)\ . \tag{2.133}$$

*Proof.* We assume $T$ finite for simplicity. For $n \geq m$ and $t \in T$ denote by $\pi_n(t)$ an element of $T_n$ such that $d(t, \pi_n(t)) = d(t, T_n)$. Consider the event $\Omega(u)$ defined by

$$\forall\, n \geq m+1\ ,\ \forall\, t \in T_n\ ,\ |X_{\pi_{n-1}(t)} - X_{\pi_n(t)}| \leq Lu2^{n/2} d(\pi_{n-1}(t), \pi_n(t))\ ,$$

and

$$\forall\, s', t' \in T_m\ ,\ |X_{s'} - X_{t'}| \leq Lud(s', t')2^{m/2}\ . \tag{2.134}$$

Then, as usual, we have $\mathsf{P}(\Omega(u)) \geq 1 - \exp(-u^2 2^m)$. Now, when $\Omega(u)$ occurs, for any $t \in T$ and any $k \geq 0$, using chaining as usual and (2.132) we get

$$|X_t - X_{\pi_m(t)}| \leq LuB\ . \tag{2.135}$$

Moreover, using (2.132) again,

$$d(t, \pi_m(t)) \leq d(t, T_m) \leq B2^{-m/2}\ ,$$

so that, using (2.135),

$$\begin{aligned} d(s,t) \leq \delta &\Rightarrow d(\pi_m(s), \pi_m(t)) \leq \delta + 2B2^{-m/2} \\ &\Rightarrow |X_{\pi_m(s)} - X_{\pi_m(t)}| \leq Lu(\delta 2^{m/2} + B)\ . \end{aligned}$$

Combining with (2.135) proves that $|X_s - X_t| \leq Lu(\delta 2^{m/2} + B)$ and completes the proof. □

**Exercise 2.9.2.** Deduce Dudley's bound (1.19) from Lemma 2.9.1.

We now turn to our main result, which exactly describes the modulus of continuity of a Gaussian process in term of certain admissible sequences. It implies in particular the remarkable fact (discovered by X. Fernique) that for Gaussian processes the "local modulus of continuity" (as in (2.136)) is also "global".

**Theorem 2.9.3.** *Consider a Gaussian process* $(X_t)_{t \in T}$, *with canonical associated distance* $d$ *given by (0.1). Assume that* $S = \mathsf{E}\sup_t X_t < \infty$. *For* $k \geq 1$ *consider* $\delta_k > 0$ *and assume that*

$$\forall\, t \in T\ ;\ \mathsf{E} \sup_{\{s \in T; d(s,t) \leq \delta_k\}} |X_s - X_t| \leq 2^{-k} S\ . \tag{2.136}$$

*Let* $n_0 = 0$ *and for* $k \geq 1$ *consider an integer* $n_k$ *for which*

$$L_1 S 2^{-n_k/2-k} \leq \delta_k \; . \tag{2.137}$$

*Then we can find an admissible sequence $(\mathcal{A}_n)$ of partitions of $T$ such that*

$$\forall\, k \geq 0 \; ; \; \sup_{t\in T} \sum_{n\geq n_k} 2^{n/2} \Delta(A_n(t)) \leq L S 2^{-k} \; . \tag{2.138}$$

*Conversely, given integers $n_k$ and an admissible sequence $(\mathcal{A}_n)$ as in (2.138), and defining now $\delta_k^* = S2^{-n_k/2-k}$, with probability $\geq 1-\exp(-u^2)$ we have*

$$\sup_{\{s,t\in T; d(s,t)\leq \delta_k^*\}} |X_s - X_t| \leq L u 2^{-k} S \; . \tag{2.139}$$

The abstract formulation here might make it hard at first to grab the power of the statement. The numbers $\delta_k$ describe the (uniform) modulus of continuity of the process. The numbers $n_k$ describe the uniform convergence (over $t$) of the series $\sum_{n\geq 0} 2^{n/2}\Delta(A_n(t))$. They relate to each other by the relation $\delta_k \sim S2^{-n_k/2-k}$. The first part of the theorem assumes only the "local" modulus of continuity (2.136), while the converse provides a uniform modulus of continuity (2.139).

*Proof.* Let us set $L_1 = 2L_0$ where $L_0$ is the constant of (2.85). By induction over $k$ we construct an admissible sequence $(\mathcal{A}_n)_{n\leq n_k}$ such that

$$1 \leq p \leq k \Rightarrow \sup_{t\in T} \sum_{n_{p-1}<n\leq n_p} 2^{n/2}\Delta(A_n(t)) \leq 2L_0 S 2^{-p} \; . \tag{2.140}$$

For $k=1$ the existence of the sequence $(\mathcal{A}_n)_{n<n_1}$ follows from the left-hand side of (2.85), so we turn to the induction step from $k$ to $k+1$. Using (2.140) for $p=k$ we deduce that for each $t\in T$, $2^{n_k/2}\Delta(A_{n_k}(t)) \leq 2L_0 S 2^{-k} = L_1 S 2^{-k}$, so that, using (2.137), $\Delta(A_{n_k}(t)) \leq L_1 S 2^{-n_k/2-k} \leq \delta_k$. Consequently, for any element $C$ of $\mathcal{A}_{n_k}$ we have $\Delta(C)\leq \delta_k$, so that considering any element $t$ of $C$ we have

$$\mathsf{E} \sup_{s\in C} X_s = \mathsf{E}\sup_{s\in C}(X_s - X_t) \leq \mathsf{E} \sup_{\{s\in T; d(s,t)\leq\delta_k\}} |X_s - X_t| \leq S2^{-k} \; .$$

Using again (2.85) we obtain for each $C\in\mathcal{A}_{n_k}$ an admissible sequence $(\mathcal{A}_{C,n})_{n\geq 0}$ for which

$$\forall\, t\in C \; , \; \sum_{n\geq 0} 2^{n/2}\Delta(A_{C,n}(t)) \leq L_0 S 2^{-k} \; . \tag{2.141}$$

For $n_k < n \leq n_{k+1}$ we simply define $\mathcal{A}_n$ as the collection of all sets in one of the partitions $A_{C,n-1}$ where $C\in\mathcal{A}_{n_k}$, so that $\operatorname{card}\mathcal{A}_n \leq N_{n-1}\operatorname{card}\mathcal{A}_{n_k} \leq N_{n-1}^2 \leq N_n$, and since $A_n(t)\subset A_{C,n-1}(t)$ it follows from (2.141) that

$$\sup_{t\in T}\sum_{n_k<n\leq n_{k+1}} 2^{n/2}\Delta(A_n(t)) \leq \sum_{n\geq n_k} 2^{n/2}\Delta(A_{C,n-1}(t)) \leq 2L_0 S 2^{-k} \; .$$

This completes the induction and the construction of the sequence $(\mathcal{A}_n)$ since (2.140) implies (2.138).

It remains to prove the "conversely" part. For this for each $n \geq 0$ we simply consider a subset $T_n$ of $T$ such that

$$\forall A \in \mathcal{A}_n \ , \ \mathrm{card}(T_n \cap A) = 1 \ .$$

We then use Lemma 2.9.1 for $m = n_k$ and $B = S2^{-k}$. □

## 2.10 Notes and Comments

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 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 (2.37). (Although, as he pointed out, R. Dudley did not state (2.37), he performed all the essential steps and (2.37) deserves to be called Dudley's bound.) A few years later, X. Fernique proved that in the "stationary case" Dudley's inequality can be reversed [35], i.e. he proved in that case the lower bound of Theorem 2.5.1. This historically important result was central to the work of Marcus and Pisier [69], [70] who build on it to solve all the classical problems on random Fourier series. Some of their results will be presented in Section 0.2. Interestingly, now that the right approach has been found, the proof of Fernique's result is not really easier than that of Theorem 2.5.1.

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 bound was essentially optimal. Gilles Pisier suggested in 1983 that I should work on this conjecture. In my first attempt I proved fast that Fernique's conjecture held in the case where the metric space $(T, d)$ is ultrametric. I learned 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 struggled in the dark for months. At some point I tried a combination of suitable point masses, and easily found a direct construction of the majorizing measure on ellipsoids.

This made it believable that Fernique's conjecture was true, but I still tried to disprove it. Then I realized that I did not understand why a direct approach 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 a matter of three days to prove Fernique's conjecture. Gilles Pisier made two comments about this discovery. The first one was "you are lucky", by which he 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 difficult to grasp, and was dismissed by the main body of probabilists as a mere curiosity. (I myself found 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 my brain.

It is on purpose that I did not mention Slepian's lemma. This 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 I made was to discover (ages ago) the scheme of proof of Proposition 2.5.9 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 (2.85) 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 $L = 1$.)

# 3. Matching Theorems

We remind the reader that, before attacking any chapter, she should find useful to read the overview of this chapter, which is provided in the appropriate subsection of Chapter 1. Here this overview should help to understand the overall approach and especially the ultimate goal of the first section.

## 3.1 Partitioning Scheme, II

Consider parameters $\alpha, p \geq 1$.

**Theorem 3.1.1.** *Consider a metric space $(T,d)$ and a number $r \geq 4$. Assume that for $j \in \mathbb{Z}$ we are given functions $s_j \geq 0$ on $T$ with the following property:*

*Whenever we consider a subset $A$ of $T, j \in \mathbb{Z}$ with $\Delta(A) \leq 2r^{-j}$ then either $e_n(A) \leq r^{-j-1}$* ***or else*** *there exist $t \in A$ with $s_j(t) \geq (2^{n/\alpha} r^{-j-1})^p$ .*

*Then we can find an admissible sequence $(\mathcal{A}_n)$ of partitions such that*

$$\forall t \in T \; ; \; \sum_{n \geq 0} (2^{n/\alpha} \Delta(A_n(t)))^p \leq K(\alpha, p, r) \sup_{t \in T} \sum_{j \in \mathbb{Z}} s_j(t) \; . \tag{3.1}$$

The proof is identical to that of Theorem 2.4.10 which corresponds to the case $\alpha = 2$ and $p = 1$.

## 3.2 The Ellipsoid Theorem

As pointed out after Proposition 2.8.2, an ellipsoid $\mathcal{E}$ is in some sense quite smaller than what one would predict by looking only at the numbers $e_n(\mathcal{E})$. We will trace the roots of this phenomenon to a simple geometric property, namely that an ellipsoid is "sufficiently convex", and we will formulate a general version of this principle for sufficiently convex bodies. The case of ellipsoids already suffices to provide tight upper bounds on certain matchings, which is the main goal of the present chapter. The general case is at the root of certain very deep facts of Banach space theory, such as Bourgain's celebrated solution of the $\Lambda_p$-problem in Sections 16.3.1 and 16.3.2.

The ellipsoid $\mathcal{E}$ of (2.118):

$$\mathcal{E} = \left\{ t \in \ell^2 \; ; \; \sum_{i \geq 1} \frac{t_i^2}{a_i^2} \leq 1 \right\} \tag{2.118}$$

is the unit ball of the norm

$$\|x\|_{\mathcal{E}} := \left( \sum_{i \geq 1} \frac{x_i^2}{a_i^2} \right)^{1/2} . \tag{3.2}$$

**Lemma 3.2.1.** *We have*

$$\|x\|_{\mathcal{E}} \, , \, \|y\|_{\mathcal{E}} \leq 1 \Rightarrow \left\| \frac{x+y}{2} \right\|_{\mathcal{E}} \leq 1 - \frac{\|x-y\|_{\mathcal{E}}^2}{8} . \tag{3.3}$$

*Proof.* The parallelogram identity implies

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

so that

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

and

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

Since (3.3) is the only property of ellipsoids we will use, it clarifies matters to state the following definition.

**Definition 3.2.2.** *Consider a number $p \geq 2$. A norm $\|\cdot\|$ in a Banach space is called $p$-convex if for a certain number $\eta > 0$ we have*

$$\|x\| \, , \, \|y\| \leq 1 \Rightarrow \left\| \frac{x+y}{2} \right\| \leq 1 - \eta \|x-y\|^p . \tag{3.4}$$

Thus (3.3) implies that the Banach space $\ell^2$ provided with the norm $\|\cdot\|_{\mathcal{E}}$ is 2-convex. For $1 < q < \infty$ the classical Banach space $L^q$ is $p$-convex where $p = \max(2, q)$. The reader is referred to [65] for this result and any other classical facts about Banach spaces. Let us observe that, taking $y = -x$ we must have

$$2^p \eta \leq 1 . \tag{3.5}$$

In this section we shall study the metric space $(T, d)$ where $T$ is the unit ball of a $p$-convex Banach space $B$, and where $d$ is the distance induced on $B$ by another norm $\|\cdot\|_\sim$. This concerns in particular the case where $T$ is the ellipsoid (2.118) and $\|\cdot\|_\sim$ is the $\ell^2$ norm.

Given a metric space $(T, d)$, we consider the functionals

$$\gamma_{\alpha,\beta}(T,d) = \Big(\inf \sup_{t\in T} \sum_{n\geq 0} \big(2^{n/\alpha} \Delta(A_n(t),d)\big)^\beta\Big)^{1/\beta}, \tag{3.6}$$

where $\alpha$ and $\beta$ are positive numbers, and where the infimum is over all admissible sequences $(\mathcal{A}_n)$. Thus, with the notation of Definition 2.2.20, we have $\gamma_{\alpha,1}(T,d) = \gamma_\alpha(T,d)$. For matchings, the important functionals are $\gamma_{2,2}(T,d)$ and $\gamma_{1,2}(T,d)$ (but it requires no extra effort to consider the general case). The importance of these functionals is that in certain conditions they nicely relate to $\gamma_2(T,d)$ through Hölder's inequality. We explain right now how this is done, even though this spoils the surprise of how the terms $\sqrt{\log N}$ occur in Section 3.5.

**Lemma 3.2.3.** *Consider a finite metric space $T$, and assume that* $\operatorname{card} T \leq N_m$. *Then*

$$\gamma_2(T,d) \leq \sqrt{m}\gamma_{2,2}(T,d)\ . \tag{3.7}$$

*Proof.* Since $T$ is finite there exists an admissible sequence $(\mathcal{A}_n)$ of $T$ for which

$$\forall t \in T\ , \sum_{n\geq 0} 2^n \Delta(A_n(t),d)^2 \leq \gamma_{2,2}(T,d)^2\ . \tag{3.8}$$

Since $\operatorname{card} T \leq N_m$, we may assume that $A_m(t) = \{t\}$ for each $t$, so that in (3.8) the sum is really over $n \leq m-1$. Since $\sum_{0\leq n\leq m-1} a_n \leq \sqrt{m}(\sum_{0\leq n\leq m} a_n^2)^{1/2}$ by the Cauchy-Schwarz inequality, it follows that

$$\forall t \in T\ , \sum_{n\geq 0} 2^{n/2} \Delta(A_n(t),d) \leq \sqrt{m}\gamma_{2,2}(T,d)\ . \qquad \square$$

How to relate the functionals $\gamma_{1,2}$ and $\gamma_2$ by a similar argument is shown in Lemma 3.7.5 below.

We may wonder how it is possible, using something as simple as the Cauchy-Schwarz inequality in Lemma 3.2.3 that we can ever get essentially exact results. At a general level the answer is obvious: it is because we use this inequality in the case of near equality. That this is indeed the case for the ellipsoids of Corollary 3.2.7 below is a non-trivial fact about the geometry of these ellipsoids.

**Theorem 3.2.4.** *If $T$ is the unit ball of a $p$-convex Banach space, if $\eta$ is as in (3.4) and if the distance $d$ on $T$ is induced by another norm, then*

$$\gamma_{\alpha,p}(T,d) \leq K(\alpha,p,\eta) \sup_{n\geq 0} 2^{n/\alpha} e_n(T,d)\ . \tag{3.9}$$

The following exercise stresses the point of this theorem.

**Exercise 3.2.5.** (a) Prove that for a general metric space $(T,d)$, it is true that

$$\gamma_{\alpha,p}(T,d) \le K(\alpha)\Big(\sum_{n\ge 0}\big(2^{n/\alpha}e_n(T,d)\big)^p\Big)^{1/p}, \tag{3.10}$$

and that

$$\sup_n 2^{n/\alpha}e_n(T,d) \le K(\alpha)\gamma_{\alpha,p}(T,d) . \tag{3.11}$$

(b) Prove that it is essentially impossible in general to improve on (3.10).

In words, the content of Theorem 3.2.4 is that the size of $T$, as measured by the functional $\gamma_{\alpha,p}$ is smaller than what one would expect when knowing only the numbers $e_n(T,d)$.

**Corollary 3.2.6.** (The Ellipsoid Theorem.) *Consider the ellipsoid $\mathcal{E}$ of (2.118) and $\alpha \ge 1$. Then*

$$\gamma_{\alpha,2}(\mathcal{E}) \le K(\alpha)\sup_{\epsilon>0}\epsilon(\mathrm{card}\{i\ ;\ a_i \ge \epsilon\})^{1/\alpha}. \tag{3.12}$$

*Proof.* Without loss of generality we may assume that the sequence $(a_i)$ is non-increasing. We apply Theorem 3.2.4 to the case $\|\cdot\| = \|\cdot\|_{\mathcal{E}}$, and where $d$ is the distance of $\ell^2$, and we get, using (2.126) in the last inequality,

$$\gamma_{\alpha,2}(\mathcal{E}) \le K(\alpha)\sup_n 2^{n/\alpha}e_n(\mathcal{E}) \le K(\alpha)\sup_n 2^{n/\alpha}a_{2^n} .$$

Now, the choice $\epsilon = a_{2^n}$ implies

$$2^{n/\alpha}a_{2^n} \le \sup_{\epsilon>0}\epsilon(\mathrm{card}\{i\ ;\ a_i \ge \epsilon\})^{1/\alpha}. \qquad \square$$

The restriction $\alpha \ge 1$ is inessential and can be removed by a suitable modification of (2.126). The important cases are $\alpha = 1$ and $\alpha = 2$.

**Corollary 3.2.7.** *Consider a countable set $J$, numbers $(b_i)_{i\in J}$ and the ellipsoid*

$$\mathcal{E} = \Big\{x \in \ell^2(J)\ ;\ \sum_{j\in J} b_j^2 x_j^2 \le 1\Big\} .$$

*Then*

$$\gamma_{\alpha,2}(\mathcal{E}) \le K(\alpha)\sup_{u>0}\frac{1}{u}(\mathrm{card}\{j \in J\ ;\ |b_j| \le u\})^{1/\alpha} .$$

*Proof.* Without loss of generality we can assume that $J = \mathbb{N}$. We then set $a_i = 1/b_i$, we apply Corollary 3.2.6, and we set $\epsilon = 1/u$. $\square$

We give right away a striking application of this result. This application is at the root of the results of Section 3.7.

**Proposition 3.2.8.** *Consider the set $\mathcal{L}$ of functions $f : [0,1] \to \mathbb{R}$ such that $f(0) = f(1) = 0$, $f$ is continuous on $[0,1]$, $f$ is differentiable outside a finite set and $\sup|f'| \le 1$. Then $\gamma_{1,2}(\mathcal{L}, d_2) \le L$, where $d_2(f,g) = \|f-g\|_2 = \big(\int_{[0,1]}(f-g)^2\mathrm{d}\lambda\big)^{1/2}$.*

*Proof.* The very beautiful idea (due to Coffman and Shor [31]) is to use the Fourier transform to represent $\mathcal{L}$ as a subset of an ellipsoid. The Fourier coefficients are defined for $p \in \mathbb{Z}$ by

$$c_p(f) = \int_0^1 \exp(2\pi i p x) f(x) \mathrm{d}x \ .$$

The key fact is the Plancherel formula,

$$\|f\|_2 = \Big(\sum_{p\in\mathbb{Z}} |c_p(f)|^2\Big)^{1/2} \ , \tag{3.13}$$

which states that the Fourier transform is an isometry from $L^2([0,1])$ into $\ell^2_{\mathbb{C}}(\mathbb{Z})$. Thus, if

$$\mathcal{D} = \{(c_p(f))_{p\in\mathbb{Z}} \ ; \ f \in \mathcal{L}\} \ ,$$

it suffices to prove that $\gamma_{1,2}(\mathcal{D}, d) < \infty$ where $d$ is the distance induced by $\ell^2_{\mathbb{C}}(\mathbb{Z})$. By integration by parts, and since $f(0) = f(1) = 0$, $c_p(f') = -2\pi i p c_p(f)$, so that, using (3.13) for $f'$, we get

$$\sum_{p\in\mathbb{Z}} p^2 |c_p(f)|^2 \le \sum_{p\in\mathbb{Z}} |c_p(f')|^2 \le \|f'\|_2 \ ,$$

and since $|c_0(f)| \le \|f\|_2 \le 1$, for $f \in \mathcal{L}$ we have

$$|c_0(f)|^2 + \sum_{p\in\mathbb{Z}} p^2 |c_p(f)|^2 \le 2 \ ,$$

so that $\mathcal{D}$ is a subset of the complex ellipsoid $\mathcal{E}$ in $\ell^2_{\mathbb{C}}(\mathbb{Z})$ defined by

$$\sum_{p\in\mathbb{Z}} \max(1, p^2) |c_p|^2 \le 2 \ .$$

Viewing each complex number $c_p$ as a pair $(x_p, y_p)$ of real numbers with $|c_p|^2 = x_p^2 + y_p^2$ yields that $\mathcal{E}$ is (isometric to) the real ellipsoid defined by

$$\sum_{p\in\mathbb{Z}} \max(1, p^2)(x_p^2 + y_p^2) \le 2 \ ,$$

and the result follows from Corollary 3.2.7. □

**Exercise 3.2.9.** (a) For $k \ge 1$ consider the space $T = \{0,1\}^{2^k}$. Writing $t = (t_i)_{i\le 2^k}$ a point of $T$, consider on $T$ the distance $d(t,t') = 2^{-j}$, where $j = \min\{i \le 2^k ; t_i \ne t'_j\}$. Consider the set $\mathcal{L}$ of 1-Lipschitz functions on $(T,d)$ which are zero at $t = (0,\dots,0)$. Prove that $\gamma_{1,2}(\mathcal{L}, d_\infty) \le L\sqrt{k}$, where $d_\infty$ denotes the distance induced by the uniform norm. Hint: use (3.10) and Lemma 3.5.13 below.
(b) Let $\mu$ denote the uniform probability $\mu$ on $T$ and $d_2$ the distance induced by $L^2(\mu)$. It can be shown that $\gamma_{1,2}(\mathcal{L}, d_2) \ge \sqrt{k}/L$. (This could be challenging.) Meditate upon the difference with Proposition 3.2.8.

*Proof of Theorem 3.2.4.* We denote by $\|\cdot\|$ the norm of the $p$-convex Banach space of which $T$ is the unit ball. For $t \in T$ and $j \in \mathbb{Z}$ we set

$$c_j(t) = \inf\{\|v\| \; ; \; v \in B_d(t, r^{-j}) \cap T\} \; ,$$

where the index $d$ emphasizes that the ball is for the distance $d$ rather than for the norm. Let us set

$$B = \sup_{n \geq 0} 2^{n/\alpha} e_n(T, d) \; . \tag{3.14}$$

The proof relies on Theorem 3.1.1 for the functions

$$s_j(t) = KB^p(c_{j+2}(t) - c_{j-1}(t)) \; ,$$

for a suitable value of $K$. It is clear that

$$\forall t \in T \; , \; \sum_{j \in \mathbb{Z}} s_j(t) \leq 3KB^p \; ,$$

and the issue is to prove that (3.1) holds for $W = KB^p$. Consider then a set $A \subset T$ with $\Delta(A) \leq 2r^{-j}$ and assume that $e_n(A) \geq a := r^{-j-1}$. Let $m = N_n$, and consider points $(t_\ell)_{\ell \leq m}$ in $A$, such that $d(t_\ell, t_{\ell'}) \geq a$ whenever $\ell \neq \ell'$. Consider $H_\ell = T \cap B_d(t_\ell, a/r)$. Set

$$u = \inf\Big\{\|v\| \; ; \; v \in \operatorname{conv} \bigcup_{\ell \leq m} H_\ell\Big\} \, , \tag{3.15}$$

and consider $u'$ such that

$$2 > u' > \max_{\ell \leq m} \inf\{\|v\| \; ; \; v \in H_\ell\} = \max_{\ell \leq m} c_{j+2}(t_\ell) \; . \tag{3.16}$$

For $\ell \leq m$ consider $v_\ell \in H_\ell$ with $\|v_\ell\| \leq u'$. It follows from (3.4) that for $\ell, \ell' \leq m$,

$$\Big\|\frac{v_\ell + v_{\ell'}}{2u'}\Big\| \leq 1 - \eta \Big\|\frac{v_\ell - v_{\ell'}}{u'}\Big\|^p \; . \tag{3.17}$$

Moreover, since $(v_\ell + v_{\ell'})/2 \in \operatorname{conv} \bigcup_{\ell \leq m} H_\ell$, we have $u \leq \|v_\ell + v'_\ell\|/2$, and (3.17) implies

$$\frac{u}{u'} \leq 1 - \eta \Big\|\frac{v_\ell - v_{\ell'}}{u'}\Big\|^p \, ,$$

so that, using that $u' \leq 2$ in the second inequality,

$$\|v_\ell - v_{\ell'}\| \leq u' \Big(\frac{u' - u}{\eta u'}\Big)^{1/p} \leq R := 2\Big(\frac{u' - u}{\eta}\Big)^{1/p} \; ,$$

and hence the points $w_\ell := R^{-1}(v_\ell - v_1)$ belong to $T$. Now, since $H_\ell \subset B_d(t_\ell, a/r)$ we have $v_\ell \in B_d(t_\ell, a/r)$. Since $r \geq 4$, we have $d(v_\ell, v_{\ell'}) \geq a/2$ for $\ell \neq \ell'$, and, since the distance $d$ arises from a norm, we have $d(w_\ell, w_{\ell'}) \geq$

$R^{-1}a/2$ for $\ell \neq \ell'$. Therefore $e_n(T,d) \geq R^{-1}a/4$, so that from (5.22) it holds $2^{n/\alpha}R^{-1}a/4 \leq B$, and hence

$$(2^{n/\alpha}r^{j-1})^p \leq KB^p(u'-u) .$$

Since this holds for any $u'$ as in (3.16), there exists $\ell$ such that

$$(2^{n/\alpha}r^{j-1})^p \leq KB^p(c_{j+2}(t_\ell)-u) . \tag{3.18}$$

Since

$$\bigcup_{\ell \leq m} H_\ell \subset B(t_\ell, r^{-j+1}) \cap T$$

it holds $u \geq c_{j-1}(t_\ell)$ and (5.40) proves the result. □

**Exercise 3.2.10.** Write the previous proof using a certain functional with an appropriate growth condition.

The following generalization of Theorem 3.2.4 yields very precise results when applied to ellipoids. It will not be used in the sequel, so we refer to [145] for a proof.

**Theorem 3.2.11.** *Consider $\beta$ , $\beta'$ , $p > 0$ with*

$$\frac{1}{\beta} = \frac{1}{\beta'} + \frac{1}{p} . \tag{3.19}$$

*Then, under the conditions of Theorem 3.2.4 we have*

$$\gamma_{\alpha,\beta}(T,d) \leq K(p,\eta,\alpha)\Big(\sum_n (2^{n/\alpha}e_n(T,d))^{\beta'}\Big)^{1/\beta'} .$$

**Exercise 3.2.12.** Use Theorem 3.2.11 to obtain a geometrical proof of (2.123). Hint: Choose $\alpha = 2, \beta = 1, \beta' = p = 2$ and use (2.126).

## 3.3 Matchings

The rest of this chapter is devoted to the following problem. Consider $N$ r.v.s $X_1, \ldots, X_N$ independently uniformly distributed in the unit cube $[0,1]^d$, where $d \geq 2$. Consider a typical realization of these points. How evenly distributed in $[0,1]^d$ are the points $X_1, \ldots, X_N$? To measure this, we will match the points $(X_i)_{i \leq N}$ with *non-random* "evenly distributed" points $(Y_i)_{i \leq N}$, that is, we will find a permutation $\pi$ of $\{1, \ldots, N\}$ such that the points $X_i$ and $Y_{\pi(i)}$ are "close". There are different ways to measure "closeness". For example one may wish that the sum of the distances $d(X_i, Y_{\pi(i)})$ be as small as possible (Section 3.5), that the maximum distance $d(X_i, Y_{\pi(i)})$ be as small as possible (Section 3.7), or one can use more complicated measures of "closeness" (Section 14.1). The case where $d = 2$ is very special, and is the object of the present chapter. The case $d \geq 3$ will be studied in Chapter 15. The reader having never thought of the matter might think that the points $X_1, \ldots, X_N$ are very evenly distributed. A moment thinking reveals this is not quite the case, for example, with probability close to one, one is bound to find a little square of area about $N^{-1} \log N$ that contains no point $X_i$. This is a very local irregularity. In a somewhat informal manner one can say that this irregularity occurs at scale $\sqrt{\log N}/\sqrt{N}$. The specific feature of the case $d = 2$ is that in some sense there are irregularities at all scales $2^{-j}$ for $1 \leq j \leq L^{-1} \log N$, and that these are all of the same order. Such a statement is by no means obvious at this stage. In the same direction, a rather deep fact about matchings is that

$$\text{\textit{obstacles to matchings at different scales may combine}}$$
$$\text{\textit{in dimension } 2 \textit{ but not in dimension }} \geq 3\,. \tag{3.20}$$

It is difficult to state a real theorem to this effect, but this is actually seen with great clarity in the proofs. The crucial estimates involve controlling sums (depending on a parameter), each term of which representing a different scale. In dimension 2, many terms contribute to the final sum (which therefore results in the contribution of many different scales), while in higher dimension only a few terms contribute. (The case of higher dimension remains non-trivial because *which* terms contribute depend on the value of the parameter.) Of course these statements are very mysterious at this stage, but we expect that a serious study of the methods involved will gradually bring the reader to share this view.

What does it mean to say that the non-random points $(Y_i)_{i \leq N}$ are evenly distributed? When $N$ is a square, $N = n^2$, everybody will agree that the $N$ points $(k/n, \ell/n), 1 \leq k\,, \ell \leq n$ are evenly distributed. More generally we will say that the non-random points $(Y_i)_{i \leq N}$ are *evenly spread* if one can cover $[0,1]^2$ with $N$ rectangles with disjoint interiors, such that each rectangle $R$ has an area $1/N$, contains exactly one point $Y_i$, and is such that $R \subset B(Y_i, 10/\sqrt{N})$. To construct such points one may proceed as follows. Consider the largest integer $k$ with $k^2 \leq N$, and observe that $k(k+3) \geq (k+1)^2 \geq N$, so that there exists integers $(n_i)_{i \leq k}$ with $k \leq n_i \leq k+3$ and $\sum_{i \leq k} n_i = N$. Cut the unit square into $k$ vertical strips, in a way that the $i$-th strip has width $n_i/N$ and to this $i$-th strip attribute $n_i$ points placed at even intervals $1/n_i$.[1]

---

[1] A more elegant approach that dispenses from this slightly awkward construction. It is the concept of "transportation cost". One attributes mass $1/N$ to each point $X_i$, and one measures the "cost of transporting" the resulting probability measure to the uniform probability on $[0,1]^2$. (In the presentation one thus replaces the evenly spread points $Y_i$ by a more canonical object, the uniform probability on $[0,1]^2$.) This approach does not make the proofs any easier, so we shall not use it despite its aesthetic appeal.

The basic tool to construct matchings is the following classical fact. The proof, based on the Hahn-Banach theorem, is given in Section A.1.

**Proposition 3.3.1.** *Consider a matrix* $C = (c_{ij})_{i,j \le N}$. *Let*

$$M(C) = \inf \sum_{i \le N} c_{i\pi(i)} ,$$

*where the infimum is over all permutations* $\pi$ *of* $\{1, \ldots, N\}$. *Then*

$$M(C) = \sup \sum_{i \le N} (w_i + w'_i) , \tag{3.21}$$

*where the supremum is over all families* $(w_i)_{i \le N}$ , $(w'_i)_{i \le N}$ *that satisfy*

$$\forall i, j \le N , \; w_i + w'_j \le c_{ij} . \tag{3.22}$$

Thus, if $c_{ij}$ is the cost of matching $i$ with $j$, $M(C)$ is the minimal cost of a matching, and is given by the "duality formula" (3.21).

The following is a well-known, and rather useful result of combinatorics. We deduce it from Proposition 3.3.1 in Section A.1, but other proofs exist, based on different ideas, see e.g. [10] § 2.

**Corollary 3.3.2 (Hall's Marriage Lemma).** *Assume that to each* $i \le N$ *we associate a subset* $A(i)$ *of* $\{1, \ldots, N\}$ *and that, for each subset* $I$ *of* $\{1, \ldots, N\}$ *we have*

$$\operatorname{card}\Big(\bigcup_{i \in I} A(i)\Big) \ge \operatorname{card} I . \tag{3.23}$$

*Then we can find a permutation* $\pi$ *of* $\{1, \ldots, N\}$ *for which*

$$\forall i \le N , \; \pi(i) \in A(i) .$$

Another well-known application of Proposition 3.3.1 is the following "duality formula".

**Proposition 3.3.3.** *Consider points* $(X_i)_{i \le N}$ *and* $(Y_i)_{i \le N}$ *in a metric space* $(T, d)$. *Then*

$$\inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) = \sup_{f \in \mathcal{C}} \sum_{i \le N} (f(X_i) - f(Y_i)) , \tag{3.24}$$

*where* $\mathcal{C}$ *denotes the class of 1-Lipschitz functions on* $(T, d)$, *i.e. functions* $f$ *for which* $|f(x) - f(y)| \le d(x, y)$.

*Proof.* Given any permutation $\pi$ and any 1-Lipschitz function $f$ we have

$$\sum_{i \le N} f(X_i) - f(Y_i) = \sum_{i \le N} (f(X_i) - f(Y_{\pi(i)})) \le \sum_{i \le N} d(X_i, Y_{\pi(i)}) .$$

This proves the inequality $\geq$ in (3.24). To prove the converse, we use (3.21) with $c_{ij} = d(X_i, Y_j)$, so that

$$\inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) = \sup \sum_{i \le N} (w_i + w'_i) \; , \tag{3.25}$$

where the supremum is over all families $(w_i)$ and $(w'_i)$ for which

$$\forall i \, , \, j \le N \, , \, w_i + w'_j \le d(X_i, Y_j) \; . \tag{3.26}$$

Given a family $(w'_i)_{i \le N}$, consider the function

$$f(x) = \min_{j \le N} (-w'_j + d(x, Y_j)) \; . \tag{3.27}$$

It is 1-Lipschitz, since it is the minimum of functions which are themselves 1-Lipschitz. By definition we have $f(Y_j) \le -w'_j$ and by (3.26) for $i \le N$ we have $w_i \le f(X_i)$, so that

$$\sum_{i \le N} (w_i + w'_i) \le \sum_{i \le N} (f(X_i) - f(Y_i)) \; . \qquad \square$$

## 3.4 Discrepancy Bounds

Generally speaking, the study of expressions of this type

$$\sup_{f \in \mathcal{F}} \Big| \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\mu) \Big| \tag{3.28}$$

for a class of functions $\mathcal{F}$ will be important in the present book, and in particular in Chapter 11. A bound on such a quantity is called a *discrepancy bound* because it bounds uniformly on $\mathcal{F}$ the "discrepancy" between the true measure $\int f \mathrm{d}\mu$ and the "empirical measure" $N^{-1} \sum_{i \le N} f(X_i)$. Finding such a bound simply requires finding a bound for the supremum of the process $(|Z_f|)_{f \in \mathcal{F}}$, where the r.v.s $Z_f$ is given by[2]

$$Z_f = \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\mu) \; , \tag{3.29}$$

a topic at the very center of our attention.

A relation between discrepancy bounds and matching theorems can be guessed from Proposition 3.3.3 and will be made explicit in the next section. In this book *every* matching theorem will be proved through a discrepancy bound.

[2] Please remember this notation which is used throughout this chapter.

## 3.5 The Ajtai-Komlós-Tusnády Matching Theorem

**Theorem 3.5.1 ([5]).** *If the points* $(Y_i)_{i\le N}$ *are evenly spread and the points* $(X_i)_{i\le N}$ *are i.i.d. uniform on* $[0,1]^2$*, then (for* $N \ge 2$*)*

$$\mathsf{E}\ \inf_{\pi} \sum_{i\le N} d(X_i, Y_{\pi(i)}) \le L\sqrt{N\log N}\ , \tag{3.30}$$

*where the infimum is over all permutations of* $\{1,\ldots,N\}$ *and where* $d$ *is the Euclidean distance.*

The term $\sqrt{N}$ is just a scaling effect. There are $N$ terms $d(X_i, Y_{\pi(i)})$ each of which should be about $1/\sqrt{N}$. The non-trivial part of the theorem is the factor $\sqrt{\log N}$. In Section 3.6 we shall show that (3.30) can be reversed, i.e.

$$\mathsf{E}\ \inf_{\pi} \sum_{i\le N} d(X_i, Y_{\pi(i)}) \ge \frac{1}{L}\sqrt{N\log N}\ . \tag{3.31}$$

Let us state the "discrepancy bound" at the root of Theorem 3.5.1. Consider the class $\mathcal{C}$ of 1-Lipschitz functions on $[0,1]^2$, i.e. of functions $f$ that satisfy

$$\forall x, y \in [0,1]^2\ ,\ |f(x) - f(y)| \le d(x,y)\ ,$$

where $d$ denotes the Euclidean distance. We denote by $\lambda$ the uniform measure on $[0,1]^2$.

**Theorem 3.5.2.** *We have*

$$\mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\le N}\Big(f(X_i) - \int f\mathrm{d}\lambda\Big)\Big| \le L\sqrt{N\log N}\ . \tag{3.32}$$

**Research problem 3.5.3.** Prove that the following limit

$$\lim_{N\to\infty}\frac{1}{\sqrt{N\log N}}\mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\le N}\Big(f(X_i) - \int f\mathrm{d}\lambda\Big)\Big|$$

exists.

At the present time there does not seem to exist the beginning of a general approach for attacking a problem of this type, and certainly the methods of the present book are not appropriate for this. Quite amazingly however, the corresponding problem has been solved in the case where the transportation cost is measured by the square of the distance, see [6]. The methods seem rather specific to the case of the square of a distance.

Theorem 3.5.2 is obviously interesting in its own right, and proving it is the goal of this section. Before we discuss it, let us put matchings behind us.

*Proof of Theorem 3.5.1.* We recall (3.24), i.e.

$$\inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) = \sup_{f \in \mathcal{C}} \sum_{i \le N} (f(X_i) - f(Y_i)) , \tag{3.33}$$

and we simply write

$$\sum_{i \le N} (f(X_i) - f(Y_i)) \le \Big| \sum_{i \le N} (f(X_i) - \int f d\lambda) \Big| + \Big| \sum_{i \le N} (f(Y_i) - \int f \mathrm{d}\lambda) \Big| . \tag{3.34}$$

Next, we claim that

$$\Big| \sum_{i \le N} (f(Y_i) - \int f \mathrm{d}\lambda) \Big| \le L\sqrt{N} . \tag{3.35}$$

We recall that since $(Y_i)_{i \le N}$ are evenly spread one can cover $[0,1]^2$ with $N$ rectangles $R_i$ with disjoint interiors, such that each rectangle $R_i$ has an area $1/N$ and is such that $Y_i \in R_i \subset B(Y_i, 10/\sqrt{N})$. Consequently

$$\Big| \sum_{i \le N} (f(Y_i) - \int f \mathrm{d}\lambda) \Big| \le \sum_{i \le N} \Big| (f(Y_i) - N \int_{R_i} f \mathrm{d}\lambda) \Big| ,$$

and since $f$ is Lipschitz each term in the right-hand side is $\le L/\sqrt{N}$. This proves the claim.

Now, using (3.33) and taking expectation

$$\begin{aligned} \mathsf{E} \inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) &\le L\sqrt{N} + \mathsf{E} \sup_{f \in \mathcal{C}} \Big| \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\lambda) \Big| \\ &\le L\sqrt{N \log N} \end{aligned}$$

by (3.32). □

To prove Theorem 3.5.2 the overall strategy is clear. We think of the left-hand side as $\mathsf{E} \sup_{f \in \mathcal{C}} |Z_f|$, where $Z_f$ is the random variable of (3.29). We then find nice tail properties for these r.v.s, and we use the methods of Chapter 2. In the end (and because we are dealing with a deep fact) we shall have to prove some delicate "smallness" property of the class $\mathcal{C}$. This smallness property will ultimately be derived from the ellipsoid theorem. The (very beautiful) strategy for the hard part of the estimates relies on a kind of 2-dimensional version of Proposition 3.2.8 and is outlined on page 80.

Whereas the delicate part of the estimates is beautifully taken care by the ellipsoid theorem, the proof is unfortunately marred by all kinds of accessory complications which must be taken care of, but for which there is plenty of room. In order to enjoy the beauty of the proof and to leave aside the complications, let us define $\mathcal{C}^*$ as the class of 1-Lipschitz functions on the unit square *which are 0 on the boundary of the square.* Rather than proving Theorem 3.5.2 itself, we will prove the following weaker result.[3]

[3] If your life really depends on understanding Theorem 3.5.2 itself please see [145].

**Theorem 3.5.4.** *We have*

$$\mathsf{E} \sup_{f \in \mathcal{C}^*} \Big| \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\lambda) \Big| \le L \sqrt{N \log N} \ . \tag{3.36}$$

The following fundamental classical result will allow us to control the tails of the r.v. $Z_f$ of (3.29). It will be used many times.

**Lemma 3.5.5 (Bernstein's inequality).** *Let $(Y_i)_{i \ge 1}$ be independent r.v.s with $\mathsf{E} Y_i = 0$ and consider a number $U$ with $|Y_i| \le U$ for each $i$. Then, for $v > 0$,*

$$\mathsf{P}\Big( \Big| \sum_{i \ge 1} Y_i \Big| \ge v \Big) \le 2 \exp\Big( - \min\Big( \frac{v^2}{4 \sum_{i \ge 1} \mathsf{E} Y_i^2}, \frac{v}{2U} \Big) \Big) \ . \tag{3.37}$$

*Proof.* For $|x| \le 1$, we have

$$|e^x - 1 - x| \le x^2 \sum_{k \ge 2} \frac{1}{k!} = x^2 (e - 2) \le x^2$$

and thus, since $\mathsf{E} Y_i = 0$, for $U |\lambda| \le 1$, we have

$$|\mathsf{E} \exp \lambda Y_i - 1| \le \lambda^2 \mathsf{E} Y_i^2 \ .$$

Therefore $\mathsf{E} \exp \lambda Y_i \le 1 + \lambda^2 \mathsf{E} Y_i^2 \le \exp \lambda^2 \mathsf{E} Y_i^2$, and thus

$$\mathsf{E} \exp \lambda \sum_{i \ge 1} Y_i = \prod_{i \ge 1} \mathsf{E} \exp \lambda Y_i \le \exp \lambda^2 \sum_{i \ge 1} \mathsf{E} Y_i^2 \ .$$

Now, for $0 \le \lambda \le 1/U$ we have

$$\begin{aligned} \mathsf{P}\Big( \sum_{i \ge 1} Y_i \ge v \Big) &\le \exp(-\lambda v) \mathsf{E} \exp \lambda \sum_{i \ge 1} Y_i \\ &\le \exp\Big( \lambda^2 \sum_{i \ge 1} \mathsf{E} Y_i^2 - \lambda v \Big) \ . \end{aligned}$$

If $Uv \le 2 \sum_{i \ge 1} \mathsf{E} Y_i^2$, we take $\lambda = v / (2 \sum_{i \ge 1} \mathsf{E} Y_i^2)$, obtaining a bound $\exp(-v^2 / (4 \sum_{i \ge 1} \mathsf{E} Y_i^2))$. If $Uv > 2 \sum_{i \ge 1} \mathsf{E} Y_i^2$, we take $\lambda = 1/U$, and we note that

$$\frac{1}{U^2} \sum_{i \ge 1} \mathsf{E} Y_i^2 - \frac{v}{U} \le \frac{Uv}{2U^2} - \frac{v}{U} \le - \frac{v}{2U} \ ,$$

so that $\mathsf{P}(\sum_{i \ge 1} Y_i \ge v) \le \exp(-\min(v^2 / 4 \sum_{i \ge 1} \mathsf{E} Y_i^2, v/2U))$. Changing $Y_i$ into $-Y_i$ we obtain the same bound for $P(\sum_{i \ge 1} Y_i \le -v)$. □

**Corollary 3.5.6.** *For each $v > 0$ we have*

$$\mathsf{P}(|Z_f| \geq v) \leq 2 \exp\left(-\min\left(\frac{v^2}{4N\|f\|_2^2}, \frac{v}{4\|f\|_\infty}\right)\right) , \tag{3.38}$$

*where $\|f\|_p$ denotes the norm of $f$ in $L_p(\lambda)$.*

*Proof.* We use Bernstein's inequality with $Y_i = f(X_i) - \int f \mathrm{d}\lambda$ if $i \leq N$ and $Y_i = 0$ if $i > N$. We then observe that $\mathsf{E}Y_i^2 \leq \mathsf{E}f^2 = \|f\|_2^2$ and $|Y_i| \leq 2 \sup |f| = 2\|f\|_\infty$. □

Let us then pretend for a while that in (3.38) the bound was instead $2\exp(-v^2/(4N\|f\|_2^2))$. Thus we would be back to the problem we consider first, bounding the supremum of a stochastic process under the increment condition (2.1), where the distance on $\mathcal{C}$ is given by the $d(f_1, f_2) = \sqrt{2N}\|f_1 - f_2\|_2$. The first thing to point out is that Theorem 3.5.2 is a prime example of a natural situation where using covering numbers does not yield the correct result, where we recall that for a metric space $(T, d)$, the covering number $N(T, d, \epsilon)$ denotes the smallest number of balls of radius $\epsilon$ that are needed to cover $T$. This is closely related to the fact that, as explained in Section 2.8, covering numbers do not describe well the size of ellipsoids. It is hard to formulate a theorem to the effect that covering numbers do not suffice, but root of the problem is described in the next exercise, and a more precise version can be found later in Exercise 3.5.15.

**Exercise 3.5.7.** (a) Prove that for each $0 < \epsilon \leq 1$

$$\log N(\mathcal{C}^*, d_2, \epsilon) \geq \frac{1}{L\epsilon^2} . \tag{3.39}$$

Hint: Consider an integer $n \geq 0$, and divide $[0,1]^2$ into $2^{2n}$ equal squares of area $2^{-2n}$. For every such square $C$ consider a number $\varepsilon_C = \pm 1$. Consider then the function $f \in \mathcal{C}$ such that $f(x) = \varepsilon_C d(x, B)$ for $x \in C$, where $B$ denotes the boundary of $C$. Prove that by appropriate choices of the signs $\varepsilon_C$ one may find at least $\exp(2^{2n}/L)$ such functions which are at mutual distance $\geq 2^{-n}/L$.
(b) Prove that $\gamma_2(\mathcal{C}^*, d_2) = \infty$. This might be challenging. Hint: try to use the previous construction on different parts of the square at different scales.

Since covering numbers do not suffice, we must then appeal to Theorem 2.2.19. It follows from Exercise 3.5.7 above that $\gamma_2(\mathcal{C}^*, d_2) = \infty$, but we will replace $\mathcal{C}^*$ by a sufficiently large finite subset $\mathcal{F}$, for which we shall need the crucial estimate $\gamma_2(\mathcal{F}, d_2) \leq L\sqrt{\log N}$. As in Proposition 3.2.8, one may then parametrize $\mathcal{C}^*$ as a subset of a certain ellipsoid using the Fourier transform, and then Corollary 3.2.7 yields $\gamma_{2,2}(\mathcal{C}^*, d_2) \leq L$. Finally the simple use of Cauchy-Schwarz inequality in (3.7) yields $\gamma_2(\mathcal{F}, d_2) \leq L\sqrt{\log\log \operatorname{card} \mathcal{F}}$, which is the desired estimate.

The main ingredient in controlling the $\ell^2$ distance is the following 2-dimensional version of Proposition 3.2.8, where we use the functional $\gamma_{2,2}$ of (3.6), and where the underlying distance is the distance induced by $L^2([0,1]^2)$.

**Proposition 3.5.8.** *We have* $\gamma_{2,2}(\mathcal{C}^*, d_2) < \infty$.

*Proof.* We represent $\mathcal{C}^*$ as a subset of an ellipsoid using the Fourier transform. The Fourier transform associates to each function $f$ on $L^2([0,1]^2)$ the complex numbers $c_{p,q}(f)$ given by

$$c_{p,q}(f) = \iint_{[0,1]^2} f(x_1, x_2) \exp 2i\pi(px_1 + qx_2) \mathrm{d}x_1 \mathrm{d}x_2 \ . \tag{3.40}$$

The Plancherel formula

$$\|f\|_2 = \Big(\sum_{p,q\in\mathbb{Z}} |c_{p,q}(f)|^2\Big)^{1/2} \tag{3.41}$$

asserts that Fourier transform is an isometry, so that if

$$\mathcal{D} = \{(c_{p,q}(f))_{p,q\in\mathbb{Z}} \ ; \ f \in \mathcal{C}^*)\} \ ,$$

it suffices to show that $\gamma_{2,2}(\mathcal{D}, d) < \infty$ where $d$ is the distance in the complex Hilbert space $\ell^2_{\mathbb{C}}(\mathbb{Z}\times\mathbb{Z})$. Using (3.40) and integration by parts we get

$$-2i\pi p c_{p,q}(f) = c_{p,q}\Big(\frac{\partial f}{\partial x}\Big) \ .$$

Using (3.41) for $\partial f/\partial x$, and since $\|\partial f/\partial x\|_2 \le 1$ we get $\sum_{p,q\in\mathbb{Z}} p^2 |c_{p,q}(f)|^2 \le 1/4\pi^2$. Proceeding similarly for $\partial f/\partial y$, we get

$$\mathcal{D} \subset \mathcal{E} = \Big\{(c_{p,q}) \in \ell^2_{\mathbb{C}}(\mathbb{Z}\times\mathbb{Z}) \ ; \ |c_{0,0}| \le 1 \ , \sum_{p,q\in\mathbb{Z}} (p^2+q^2)|c_{p,q}|^2 \le 1\Big\} \ .$$

We view each complex number $c_{p,q}$ as a pair $(x_{p,q}, y_{p,q})$ of real numbers, and $|c_{p,q}|^2 = x^2_{p,q} + y^2_{p,q}$, so that

$$\begin{aligned} \mathcal{E} = \ &\Big\{\big((x_{p,q}), (y_{p,q})\big) \in \ell^2(\mathbb{Z}\times\mathbb{Z}) \times \ell^2(\mathbb{Z}\times\mathbb{Z}) \ ; \\ &x^2_{0,0} + y^2_{0,0} \le 1 \ , \sum_{p,q\in\mathbb{Z}} (p^2+q^2)(x^2_{p,q} + y^2_{p,q}) \le 1\Big\} \ . \end{aligned} \tag{3.42}$$

For $u \ge 1$, we have

$$\mathrm{card}\big\{(p,q) \in \mathbb{Z}\times\mathbb{Z} \ ; \ p^2 + q^2 \le u^2\big\} \le (2u+1)^2 \le Lu^2 \ .$$

We then deduce from Corollary 3.2.7 that $\gamma_{2,2}(\mathcal{E}, d) < \infty$. □

Let us now came back to Earth and deal with the actual bound (3.38). For this we develop an appropriate version of Theorem 2.2.19. It will be used many times. The ease with which one deals with two distances is remarkable.

**Theorem 3.5.9.** *Consider a set $T$ provided with two distances $d_1$ and $d_2$. Consider a centered process $(X_t)_{t\in T}$ which satisfies*

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

$$\mathsf{P}(|X_s - X_t| \ge u) \le 2\exp\left(-\min\left(\frac{u^2}{d_2(s,t)^2}, \frac{u}{d_1(s,t)}\right)\right)\ . \tag{3.43}$$

*Then*

$$\mathsf{E}\sup_{s,t\in T}|X_s - X_t| \le L(\gamma_1(T,d_1) + \gamma_2(T,d_2))\ . \tag{3.44}$$

This theorem will be applied when $d_1$ is the $\ell_\infty$ distance, but it sounds funny, when considering two distances, to call them $d_2$ and $d_\infty$.

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

$$\forall t \in T\ ,\ \sum_{n\ge 0} 2^n \Delta_1(B_n(t)) \le 2\gamma_1(T,d_1) \tag{3.45}$$

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

$$\forall t \in T\ ,\ \sum_{n\ge 0} 2^{n/2} \Delta_2(C_n(t)) \le 2\gamma_2(T,d_2)\ . \tag{3.46}$$

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 $n \ge 1$, 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

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

and the sequence $(\mathcal{A}_n)$ is admissible.[4] For each $n \ge 0$ 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 (3.43) 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 $u \ge \sqrt{u}$,

$$\begin{aligned}\mathsf{P}\Big(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| &\ge u\big(2^n d_1(\pi_n(t),\pi_{n-1}(t)) + 2^{n/2} d_2(\pi_n(t),\pi_{n-1}(t))\big)\Big) \\ &\le 2\exp(-u2^n)\ ,\end{aligned} \tag{3.47}$$

so that, proceeding as in (2.31), with probability $\ge 1 - L\exp(-u)$ we have

[4] Observe how the inequality $N_n^2 \le N_{n+1}$ makes it convenient to work with the sequence $N_n$.

$$\forall n \;,\; \forall t \;,\; |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \le u\big(2^n d_1(\pi_n(t), \pi_{n-1}(t)) + 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t))\big) \;. \tag{3.48}$$

Now, under (3.48) 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 (3.45) in the last inequality,

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

Proceeding similarly for $d_2$ shows that under (3.48) 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 (2.47),

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge Lu(\gamma_1(T, d_1) + \gamma_2(T, d_2))\Big) \le L\exp(-u) \;, \tag{3.49}$$

which using (2.6) implies the result. □

**Exercise 3.5.10.** Consider a space $T$ equipped with two different distances $d_1$ and $d_2$. Prove that

$$\gamma_2(T, d_1 + d_2) \le L(\gamma_2(T, d_1) + \gamma_2(T, d_2)) \;. \tag{3.50}$$

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 3.5.9 the partitions generated by $\mathcal{A}_n$ and $\mathcal{B}_n$.

**Exercise 3.5.11.** (R. Latała, S. Mendelson) Consider a process $(X_t)_{t\in T}$ and for a subset $A$ of $T$ and $n \ge 0$ 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 $(\mathcal{A}_n)_{n\ge 0}$.
(a) Prove that

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

Hint: Use chaining and (1.13) for $\varphi(x) = |x|^{2^n}$.
(b) Explain why this result implies Theorem 3.5.9.

We can now state a general bound, from which we will deduce Theorem 3.5.2.

**Theorem 3.5.12.** *Consider a class $\mathcal{F}$ of functions on $[0,1]^2$ and assume that $0 \in \mathcal{F}$. Then*

$$\mathsf{E} \sup_{f \in \mathcal{F}} \Big| \sum_{i \le N} \Big( f(X_i) - \int f \mathrm{d}\lambda \Big) \Big| \le L\big(\sqrt{N}\gamma_2(\mathcal{F}, d_2) + \gamma_1(\mathcal{F}, d_\infty)\big) , \tag{3.51}$$

*where $d_2$ and $d_\infty$ are the distances induced on $\mathcal{F}$ by the norms of $L^2$ and $L^\infty$ respectively.*

*Proof.* Combining Corollary 3.5.6 with Theorem 3.5.9 we get, since $0 \in \mathcal{F}$,

$$\mathsf{E} \sup_{f \in \mathcal{F}} |Z_f| \le \mathsf{E} \sup_{f, f' \in \mathcal{F}} |Z_f - Z_{f'}| \le L\big(\gamma_2(\mathcal{F}, 2\sqrt{N} d_2) + \gamma_1(\mathcal{F}, 4 d_\infty)\big) . \tag{3.52}$$

Finally, $\gamma_2(\mathcal{F}, 2\sqrt{N} d_2) = 2\sqrt{N}\gamma_2(\mathcal{F}, d_2)$ and $\gamma_1(\mathcal{F}, 4d_\infty) = 4\gamma_1(\mathcal{F}, d_\infty)$. □

There is plenty of room to control $\gamma_1(\mathcal{F}, d_\infty)$, our next task. We first state a general principle, which was already known to Kolmogorov.

**Lemma 3.5.13.** *Consider a metric space $(U, d)$ and assume that for certain numbers $B$ and $\alpha \ge 1$ and each $0 < \epsilon < B$ we have*

$$N(U, d, \epsilon) \le \Big(\frac{B}{\epsilon}\Big)^\alpha . \tag{3.53}$$

*Consider the set $\mathcal{B}$ of 1-Lipschitz functions $f$ on $U$ with $\|f\|_\infty \le B$. Then for each $\epsilon > 0$ we have*

$$\log N(\mathcal{B}, d_\infty, \epsilon) \le K\Big(\frac{B}{\epsilon}\Big)^\alpha , \tag{3.54}$$

*where $K$ depends only on $\alpha$. In particular,*

$$e_n(\mathcal{B}, d_\infty) \le KB2^{-n/\alpha} . \tag{3.55}$$

*Proof.* By homogeneity we may and do assume that $B = 1$. For each $n \ge 0$ consider a set $V_n \subset U$ with $\operatorname{card} V_n \le 2^{n\alpha}$ such that any point of $U$ is within distance $2^{-n}$ of a point of $V_n$. We define on $\mathcal{B}$ the distance $d_n$ by $d_n(f, g) = \max_{x \in V_n} |f(x) - g(x)|$. We prove first that

$$d(f, g) \le 2^{-n+1} + d_n(f, g) . \tag{3.56}$$

Indeed, for any $x \in U$ we can find $y \in V_n$ with $d(x, y) \le 2^{-n}$ and then $|f(x) - g(x)| \le 2^{-n+1} + |f(y) - g(y)| \le 2^{-n+1} + d_n(f, g)$. Next we prove that

$$N(\mathcal{B}, d_n, 2^{-n}) \le L^{\operatorname{card} V_n} N(\mathcal{B}, d_{n-1}, 2^{-n+1}) . \tag{3.57}$$

For this we fix $n$. Considering $g \in \mathcal{B}$, if $f \in \mathcal{B}$ satisfies $d_{n-1}(f,g) \le 2^{-n+1}$ then by using (3.56) for $n-1$ rather than $n$ we obtain to for each $x \in v_n$ we have $|f(x) - g(x)| \le 2^{-n+3}$. The usual volume argument shows that the ball of radius $2^{-n+1}$ for $d_{n-1}$ can be covered by $L^{\operatorname{card} V_n}$ balls of radius $2^{-n}$ and this proves (3.57). Since $\operatorname{card} V_n = 2^{\alpha n}$, iteration of the relation (3.57) proves that $\log N(\mathcal{B}, d_n, 2^{-n}) \le K 2^{\alpha n}$. Finally (3.56) implies that

$$\log N(\mathcal{B}, d, 2^{-n+2}) \le \log N(\mathcal{B}, d_{n-1}, 2^{-n-1}) \le K 2^{\alpha n}$$

and concludes the proof. □

We apply the previous lemma to $U = [0,1]^2$ which obviously satisfies (3.53) for $\alpha = 2$, so that (3.55) implies that for $n \ge 0$,

$$e_n(\mathcal{C}^*, d_\infty) \le L 2^{-n/2} . \tag{3.58}$$

**Proposition 3.5.14.** *We have*

$$\mathsf{E} \sup_{f \in \mathcal{C}^*} \Big| \sum_{i \le N} f(X_i) - \int f \mathrm{d}\lambda \Big| \le L \sqrt{N \log N} . \tag{3.59}$$

*Proof.* Consider the largest integer $m$ with $2^{-m} \ge 1/N$. By (3.58) we may find a subset $T$ of $\mathcal{C}^*$ with $\operatorname{card} T \le N_m$ and

$$\forall f \in \mathcal{C}^* ,\ d_\infty(f, T) \le L 2^{-m/2} \le L/\sqrt{N} .$$

Thus

$$\mathsf{E} \sup_{f \in \mathcal{C}^*} |Z_f| \le \mathsf{E} \sup_{f \in T} |Z_f| + L\sqrt{N} . \tag{3.60}$$

To prove (3.59) it suffices to show that

$$\mathsf{E} \sup_{f \in T} |Z_f| \le L \sqrt{N \log N} . \tag{3.61}$$

Proposition 3.5.8 and Lemma 3.2.3 imply $\gamma_2(T, d_2) \le L\sqrt{m} \le L\sqrt{\log N}$. Now, as in (2.72) we have

$$\gamma_1(T, d_\infty) \le L \sum_{n \ge 0} 2^n e_n(T, d_\infty) .$$

Since $e_n(T, d_\infty) = 0$ for $n \ge m$, (3.58) yields $\gamma_1(T, d_\infty) \le L 2^{m/2} \le L\sqrt{N}$. Thus (3.61) follows from Theorem 3.5.12 and this completes the proof. □

**Exercise 3.5.15.** Use Exercise 3.5.7 to convince yourself that covering numbers cannot yield better than the estimate $\gamma_2(T, d_2) \le L \log N$.

**Exercise 3.5.16.** Consider the space $T = \{0,1\}^{\mathbb{N}}$ provided with the distance $d(t,t') = 2^{-j/2}$, where $j = \min\{i \geq 1; t_i \neq t'_i\}$ for $t = (t_i)_{i\geq 1}$. This space somewhat resembles the unit square, in the sense that $N(T,d,\epsilon) \leq L\epsilon^{-2}$ for $\epsilon \leq 1$. Prove that if $(X_i)_{i\leq N}$ are i.i.d. uniformly distributed in $T$ and $(Y_i)_{i\leq N}$ are uniformly spread (in a manner which is left to the reader to define precisely) then

$$\frac{1}{L}\sqrt{N}\log N \leq \mathsf{E}\inf_{\pi}\sum_{i\leq N} d(X_i, Y_{\pi}(i)) \leq L\sqrt{N}\log N \ , \tag{3.62}$$

where the infimum is over all the permutations of $\{1,\ldots,N\}$. Hint: for the upper bound, covering numbers suffice, e.g. in the form of (3.54). You may find the lower bound a real challenge.

## 3.6 Lower Bound for the Ajtai-Komlós-Tusnády Theorem

Recalling that $\mathcal{C}$ denotes the class of functions that are 1-Lipschitz on the unit square, and that $\mathcal{C}^*$ denotes the class of these functions that are 0 on the boundary of the square we shall prove the following, where $(X_i)_{i\leq N}$ are i.i.d. in $[0,1]^2$.

**Theorem 3.6.1.** *We have*

$$\mathsf{E}\sup_{f\in\mathcal{C}^*}\Big|\sum_{i\leq N}\Big(f(X_i) - \int f\mathrm{d}\lambda\Big)\Big| \geq \frac{1}{L}\sqrt{N\log N} \ . \tag{3.63}$$

In particular it follows from (3.35) that if the points $Y_i$ are evenly spread then (provided $N \geq L$),

$$\mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\leq N}(f(X_i) - f(Y_i))\Big| \geq \frac{1}{L}\sqrt{N\log N} \ ,$$

so (3.24) implies that the expected cost of matching the points $X_i$ and the points $Y_i$ is at least $\sqrt{N\log N}/L$.

We may assume $N$ large and we consider a number $r \in \mathbb{N}$ with $r \simeq \log N/100$. The idea of the proof is to recursively construct with high probability for $k \leq r$ the functions $f_k$ such that

$$\sum_{i\leq N}\Big(f_k(X_i) - \int f_k\mathrm{d}\lambda\Big) \geq \frac{\sqrt{N}}{L\sqrt{r}} \ , \tag{3.64}$$

and for any $q \leq r$,

$$\sum_{k\leq q} f_k \text{ is 1-Lipschitz} \ . \tag{3.65}$$

The function $g = \sum_{k \le r} f_k$ is then 1-Lipschitz and satisfies

$$\sum_{i \le N} \Big( g(X_i) - \int g \mathrm{d}\lambda \Big) \ge \frac{\sqrt{Nr}}{L}$$

and this completes the proof.

For $1 \le k \le r$ and $1 \le \ell \le 2^k$ we consider the function[5] $f'_{k,\ell}$ on $[0,1]$ defined as follows:

$$f'_{k,\ell}(x) = \begin{cases} 0 & \text{unless } x \in [(\ell-1)2^{-k}, \ell 2^{-k}[ \\ \frac{1}{2\sqrt{r}} & \text{for } x \in [(\ell-1)2^{-k}, (\ell-1/2)2^{-k}[ \\ -\frac{1}{2\sqrt{r}} & \text{for } x \in [(\ell-1/2)2^{-k}, \ell 2^{-k}[ \ . \end{cases} \tag{3.66}$$

We define

$$f_{k,\ell}(x) = \int_0^x f'_{k,\ell}(y) \mathrm{d}y \ . \tag{3.67}$$

We now list a few useful properties of these functions. In these formulas $\|.\|_2$ denotes the norm in $L^2([0,1])$, etc. The proofs of these assertions are completely straightforward and better left to the reader.

**Lemma 3.6.2.** *The following holds true:*

$$f'_{k,\ell}(x) = 0 \textit{ unless } x \in [(\ell-1)2^{-k}, \ell 2^{-k}[ \ . \tag{3.68}$$

$$\textit{The family } (f'_{k,\ell}) \textit{ is orthogonal in } L^2([0,1]) \ . \tag{3.69}$$

$$\|f'_{k,\ell}\|_2^2 = \frac{1}{4r} 2^{-k} \ . \tag{3.70}$$

$$\|f'_{k,\ell}\|_1 = \frac{1}{2\sqrt{r}} 2^{-k} \ . \tag{3.71}$$

$$\|f_{k,\ell}\|_1 = \frac{1}{8\sqrt{r}} 2^{-2k} \ . \tag{3.72}$$

$$\|f'_{k,\ell}\|_\infty = \frac{1}{2\sqrt{r}} \ ; \ \|f_{k,\ell}\|_\infty = \frac{1}{4\sqrt{r}} 2^{-k} \ . \tag{3.73}$$

$$\|f_{k,\ell}\|_2^2 = \frac{1}{48r} 2^{-3k} \ . \tag{3.74}$$

Given numbers $z_{k,\ell,\ell'} \in \{0, 1, -1\}$ we consider the function

$$f_k = \frac{\sqrt{r}}{16} 2^k \sum_{\ell, \ell' \le 2^k} z_{k,\ell,\ell'} f_{k,\ell} \otimes f_{k,\ell'} \ , \tag{3.75}$$

[5] These functions, together with the constant function equal to one, form the so-called Haar basis.

where $f_{k,\ell} \otimes f_{k,\ell'}(x,y) = f_{k,\ell}(x) f_{k,\ell'}(y)$. Since $f_{k,\ell} \otimes f_{k,\ell'}$ is zero outside the little square $[(\ell-1)2^{-k}, \ell 2^{-k}[\times[(\ell'-1)2^{-k}, \ell' 2^{-k}[$, and since these little squares are disjoint as $\ell$ and $\ell'$ vary, the function $f_k$ is easy to visualize.

One obstacle is that a function of the type $\sum_{k\le q} f_k$ as in (3.65) need not be always 1-Lipschitz. It shall require some care to ensure that we properly choose the coefficients $z_{k,\ell,\ell'}$ to ensure this condition. The next two lemmas prepare for this.

**Lemma 3.6.3.** *A function* $f = \sum_{k\le q} f_k$ *satisfies*

$$\Big\|\frac{\partial f}{\partial x}\Big\|_2 \le 2^{-7} . \tag{3.76}$$

*Proof.* First we write

$$\frac{\partial f}{\partial x}(x,y) = \frac{\sqrt{r}}{16} \sum_{k\le r} 2^k \sum_{\ell,\ell'\le 2^k} z_{k,\ell,\ell'} f'_{k,\ell}(x) f_{k,\ell'}(y) .$$

Using (3.69) and (3.70) we obtain, since $z^2_{k,\ell,\ell'} \le 1$,

$$\begin{aligned}\int \Big(\frac{\partial f}{\partial x}\Big)^2 \mathrm{d}x &= \frac{r}{(16)^2} \sum_{k\le q} 2^{2k} \sum_{\ell,\ell'\le 2^k} z^2_{k,\ell,\ell'} \|f'_{k,\ell}\|_2^2 f_{k,\ell'}(y)^2 \\ &\le \frac{1}{2^{10}} \sum_{k\le q} 2^k \sum_{\ell,\ell'\le 2^k} f_{k,\ell'}(y)^2 .\end{aligned}$$

Integrating in $y$ and using (3.74) yields

$$\Big\|\frac{\partial f}{\partial x}\Big\|_2^2 \le \frac{1}{2^{10}} \sum_{k\le r} \frac{1}{48r} \le 2^{-14} . \qquad \square$$

**Lemma 3.6.4.** *Consider a function of the type* $f = \sum_{k\le q} f_k$*, where* $f_k$ *is given by (3.64) and where* $z_{k,\ell,\ell'} \in \{0,1,-1\}$*. Then*

$$\Big|\frac{\partial^2 f}{\partial x \partial y}\Big| \le \frac{2^q}{2^5\sqrt{r}} . \tag{3.77}$$

*Proof.* We note that

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\sqrt{r}}{16} \sum_{k\le q} 2^k \sum_{\ell,\ell'\le 2^k} z_{k,\ell,\ell'} f'_{k,\ell} \otimes f'_{k,\ell'} ,$$

and we note from the second part of (3.73) that since the functions $(f'_{k,\ell})_{\ell\le 2^k}$ have disjoint supports, that second sum is $\le 1/(4r)$ at every point. Also, $\sum_{k\le q} 2^k \le 2^{q+1}$. $\square$

Having constructed $f_1, \ldots, f_q$, to construct $f_{q+1}$ we need to construct the coefficients $z_{q+1,\ell,\ell'}$. Let $g := \sum_{k \le q} f_k$, to obtain (3.65) we need to ensure that $g' := g + f_{q+1}$ is Lipschitz. For this let us consider the little squares of the type

$$[(\ell-1)2^{-(q+1)}, \ell 2^{-(q+1)}[ \times [(\ell'-1)2^{-(q+1)}, \ell' 2^{-(q+1)}[ \tag{3.78}$$

for $1 \le \ell, \ell' \le 2^{q+1}$, so that there are $2^{2(q+1)}$ such squares. To ensure that $g'$ is 1-Lipschitz, it suffices to ensure that it is 1-Lipschitz on each square (3.78). Let us say that the square (3.78) is *dangerous* if it contains a point for which either $|\partial g/\partial x| \ge 1/2$ or $|\partial g/\partial y| \ge 1/2$. (The danger is that on this square $g' = g + f_{q+1}$ might not be 1-Lipschitz.) We observe from the definition that all functions $f'_{k,\ell}$ for $k \le q$ are constant on the squares (3.78). So on such a square the quantity $\partial f_q/\partial x$ does not depend on $x$. Moreover it follows from (3.77) that if $(x,y)$ and $(x,y')$ belong to the same square (3.78) then

$$\Big|\frac{\partial g}{\partial x}(x,y) - \frac{\partial g}{\partial x}(x,y')\Big| \le |y-y'|\frac{2^{q-5}}{\sqrt{r}} \le \frac{2^{-6}}{\sqrt{r}} \ .$$

In particular if a square (3.78) contains a point at which $|\partial g/\partial x| \ge 1/2$, then at each point of this square we have $|\partial g/\partial x| \ge 1/4$. Consequently (3.76) implies, with room to spare, that at most 1/2 of the squares (3.78) are dangerous. For these squares, we choose $z_{q+1,\ell,\ell'} = 0$, so that on these squares $g'$ will be 1-Lipschitz. Let us say that a square (3.78) is *safe* if it is not dangerous, so that at each point of a safe square we have $|\partial g/\partial x| \le 1/2$ and $|\partial g/\partial y| \le 1/2$. Now (3.73) implies

$$\Big|\frac{\partial g'}{\partial x} - \frac{\partial g}{\partial x}\Big| = \Big|\frac{\sqrt{r}}{16}2^{q+1}\sum_{\ell,\ell' \le 2^{q+1}} z_{q+1,\ell,\ell'} f'_{q+1,\ell} \otimes f_{q+1,\ell'}\Big| \le \frac{1}{2^7\sqrt{r}}$$

and

$$\Big|\frac{\partial g'}{\partial y} - \frac{\partial g}{\partial y}\Big| = \Big|\frac{\sqrt{r}}{16}2^{q+1}\sum_{\ell,\ell' \le 2^{q+1}} z_{q+1,\ell,\ell'} f_{q+1,\ell} \otimes f'_{q+1,\ell'}\Big| \le \frac{1}{2^7\sqrt{r}} \ ,$$

so we are certain than on a safe square we have $|\partial g'/\partial x| \le 1/\sqrt{2}$ and $|\partial g'/\partial y| \le 1/\sqrt{2}$, and hence that $g'$ is 1-Lipschitz.

At least half of the squares are safe. For a safe square, we chose $z_{q+1,\ell,\ell'} = \pm 1$ such that

$$z_{q+1,\ell,\ell'} D_{\ell,\ell'} = |D_{\ell,\ell'}|$$

where

$$D_{\ell,\ell'} = \sum_{i \le N}\Big(f_{q+1,\ell} \otimes f_{q+1,\ell'}(X_i) - \int f_{q+1,\ell} \otimes f_{q+1,\ell'} \mathrm{d}\lambda\Big) \ . \tag{3.79}$$

It is straightforward to show that

$$\mathsf{E}D_{\ell,\ell'}^2 \geq \frac{2^{-6q}N}{Lr^2} \ . \tag{3.80}$$

Let us then pretend for a moment that the r.v.s $D_{\ell,\ell'}$ are Gaussian and independent as $\ell, \ell'$ vary. Then with overwhelming probability at least 3/4 of them will be such that $|D_{\ell,\ell'}| \geq 2^{-3q}\sqrt{N}/Lr$. Thus this inequality will hold for at least 1/2 of the safe squares, and (3.64) holds for $k = q+1$ as desired.

It is not exactly true that the r.v.s $D_{\ell,\ell'}$ are independent and Gaussian. Standard techniques exist to take care of this, namely Poissonization and normal approximation. There is all the room in the world because $r \leq \sqrt{\log N}/100$. As these considerations are not related to the rest of the material of this work they are better omitted.[6]

## 3.7 The Leighton-Shor Grid Matching Theorem

**Theorem 3.7.1 ([63]).** *If the points $(Y_i)_{i\leq N}$ are evenly spread and if $(X_i)_{i\leq N}$ are i.i.d. uniform over $[0,1]^2$, then (for $N \geq 2$), with probability at least $1 - L\exp(-(\log N)^{3/2}/L)$ we have*

$$\inf_\pi \sup_{i\leq N} d(X_i, Y_{\pi(i)}) \leq L\frac{(\log N)^{3/4}}{\sqrt{N}} \ , \tag{3.81}$$

*and thus*

$$\mathsf{E}\inf_\pi \sup_{i\leq N} d(X_i, Y_{\pi(i)}) \leq L\frac{(\log N)^{3/4}}{\sqrt{N}} \ . \tag{3.82}$$

To deduce (3.82) from (3.81) one simply uses any matching in the (rare) event that (3.81) fails. We shall prove in Section 3.8 that the inequality (3.82) can be reversed.

A first simple idea is that to prove Theorem 3.7.1 we do not care about what happens at a scale smaller than $(\log N)^{3/4}/\sqrt{N}$. Consider the largest integer $\ell_1$ with $2^{-\ell_1} \geq (\log N)^{3/4}/\sqrt{N}$ (so that in particular $2^{\ell_1} \leq \sqrt{N}$). We divide $[0,1]$ into little squares of side $2^{-\ell_1}$. For each such square, we are interested in how many points $(X_i)$ it contains, but we do not care where these points are located in the square. We shall (as is the case for each matching theorem) deduce Theorem 3.7.1 from a discrepancy theorem for a certain class of functions. What we really have in mind is the class of functions which are indicators of a union $A$ of little squares of side $2^{-\ell_1}$, and such that the boundary of $A$ has a given length. It turns out that we shall have to parametrize the boundaries of these sets by curves, so it is convenient to turn things around and to consider the class of sets $A$ that are the interiors of curves of given length.

[6] The beautiful argument presented here goes back to [5]. If you find it too informal, other approach may be found in [145].

To make things precise, let us define the *grid* $G$ of $[0,1]^2$ of mesh width $2^{-\ell_1}$ by

$$G = \left\{(x_1, x_2) \in [0,1]^2 \; ; \; 2^{\ell_1} x_1 \in \mathbb{N} \;\; \textbf{or} \;\; 2^{\ell_1} x_2 \in \mathbb{N}\right\} .$$

A *vertex* of the grid is a point $(x_1, x_2) \in [0,1]^2$ with $2^{\ell_1} x_1 \in \mathbb{N}$ **and** $2^{\ell_1} x_2 \in \mathbb{N}$. An *edge* of the grid is the segment between two vertices that are at distance $2^{-\ell_1}$ of each other. A *square* of the grid is a square of side $2^{-\ell_1}$ whose edges are edges of the grid. Thus, an edge of the grid is a subset of the grid, but a square of the grid is not a subset of the grid.

A *curve* is the image of a continuous map $\varphi : [0,1] \to \mathbb{R}^2$. We say that the curve is a *simple curve* if it is one-to-one on $[0,1[$. We say that the curve is *traced on* $G$ if $\varphi([0,1]) \subset G$, and that it is *closed* if $\varphi(0) = \varphi(1)$. If $C$ is a closed simple curve in $\mathbb{R}^2$, the set $\mathbb{R}^2 \setminus C$ has two connected components. One of these is bounded. It is called the interior of $C$ and is denoted by $\overset{o}{C}$.

The proof of Theorem 3.7.1 has probabilistic part and a deterministic part. The probabilistic part is as follows.

**Theorem 3.7.2.** *With probability at least* $1 - L\exp(-(\log N)^{3/2}/L)$, *the following occurs. Given any closed simple curve* $C$ *traced on* $G$, *we have*

$$\Big|\sum_{i \le N} \big(\mathbf{1}_{\overset{o}{C}}(X_i) - \lambda(\overset{o}{C})\big)\Big| \le L\ell(C)\sqrt{N}(\log N)^{3/4} , \tag{3.83}$$

*where* $\lambda(\overset{o}{C})$ *is the area of* $\overset{o}{C}$ *and* $\ell(C)$ *is the length of* $C$.

It will be easier to discuss the following result, which concerns curves of a given length going through a given vertex.

**Proposition 3.7.3.** *Consider a vertex* $\tau$ *of* $G$ *and* $k \in \mathbb{Z}$. *Define* $\mathcal{C}(\tau, k)$ *as the set of closed simple curves traced on* $G$ *that contain* $\tau$ *and have length* $\le 2^k$. *Then, if* $k \le \ell_1 + 2$, *with probability at least* $1 - L\exp(-(\log N)^{3/2}/L)$, *for each* $C \in \mathcal{C}(\tau, k)$ *we have*

$$\Big|\sum_{i \le N} \big(\mathbf{1}_{\overset{o}{C}}(X_i) - \lambda(\overset{o}{C})\big)\Big| \le L2^k\sqrt{N}(\log N)^{3/4} . \tag{3.84}$$

It would be easy to control the left-hand side if one considered only curves with a simple pattern, such as boundaries of rectangles. The point however is that the curves we consider can be very complicated, and the longer we allow them to be, the more so. Before we discuss Proposition 3.7.3 further, we show that it implies Theorem 3.7.2.

*Proof of Theorem 3.7.2.* Since there are at most $(2^{\ell_1} + 1)^2 \le LN$ choices for $\tau$, we can assume with probability at least

$$1 - L(2^{\ell_1} + 1)^2(2\ell_1 + 4)\exp(-(\log N)^{3/2}/L) \ge 1 - L'\exp\big(-(\log N)^{3/2}/L'\big)$$

that (3.84) occurs for all choices of $C \in \mathcal{C}(\tau, k)$, for any $\tau$ and any $k$ with $-\ell_1 \le k \le \ell_1 + 2$.

Consider a simple curve $C$ traced on $G$. Then, bounding the length of $C$ by the total length of the edges of $G$, we have

$$2^{-\ell_1} \le \ell(C) \le 2(2^{\ell_1}+1) \le 2^{\ell_1+2}\,,$$

so if $k$ is the smallest integer for which $\ell(C) \le 2^k$, then $-\ell_1 \le k \le \ell_1+2$, so that we can use (3.84) and since $2^k \le 2\ell(C)$ the proof is finished. □

Let us go back to the analysis of Proposition 3.7.3. We denote by $\mathcal{F}(=\mathcal{F}_k)$ the class of functions of the type $\mathbf{1}_{\overset{\circ}{C}}$, where $C \in \mathcal{C}(\tau,k)$. Then the left-hand side of (3.84) is

$$\sup_{\mathcal{F}}\Big|\sum_{i\le N}(f(X_i)-\int f\mathrm{d}\lambda)\Big|\,.$$

The key point again is the control on the size of $\mathcal{F}$ for the distance of $L^2(\lambda)$. The difficult part of this control is the following.

**Proposition 3.7.4.** *We have*

$$\gamma_2(\mathcal{F},d_2) \le L2^k(\log N)^{3/4}\,. \tag{3.85}$$

Probably our most urgent duty is to reveal how the exponent 3/4 occurs. It is through the following general principle.

**Lemma 3.7.5.** *Consider a finite metric space* $(T,d)$ *with* $\operatorname{card} T \le N_m$. *Then*

$$\gamma_2(T,\sqrt{d}) \le m^{3/4}\gamma_{1,2}(T,d)^{1/2}\,. \tag{3.86}$$

*Proof.* Since $T$ is finite there exists an admissible sequence $(\mathcal{A}_n)$ of $T$ such that

$$\forall t\in T\,,\ \sum_{n\ge 0}(2^n\Delta(A_n(t),d))^2 \le \gamma_{1,2}(T,d)^2\,. \tag{3.87}$$

Without loss of generality we can assume that $A_m(t)=\{t\}$ for each $t$, so that in (3.87) the sum is over $n\le m-1$. Now

$$\Delta(A,\sqrt{d}) \le \Delta(A,d)^{1/2}$$

so that, using Hölder's inequality,

$$\begin{aligned}\sum_{0\le n\le m-1}2^{n/2}\Delta(A_n(t),\sqrt{d}) &\le \sum_{0\le n\le m-1}(2^n\Delta(A_n(t),d))^{1/2}\\ &\le m^{3/4}\Big(\sum_{n\ge 0}\big(2^n\Delta(A_n(t),d)\big)^2\Big)^{1/4}\\ &\le m^{3/4}\gamma_{1,2}(T,d)^{1/2}\,,\end{aligned}$$

which concludes the proof. □

**Lemma 3.7.6.** *We have* $\operatorname{card}\mathcal{C}(\tau,k) \le 2^{2^{k+\ell_1+1}} = N_{k+\ell_1+1}$.

*Proof.* A curve $C \in \mathcal{C}(\tau,k)$ consists of at most $2^{k+\ell_1}$ edges of $G$. If we move through $C$, at each vertex of $G$ we have at most 4 choices for the next edge, so $\operatorname{card}\mathcal{C}(\tau,k) \le 4^{2^{k+\ell_1}} = N_{k+\ell_1+1}$. □

**Corollary 3.7.7.** *We have*

$$\gamma_1(\mathcal{F}, d_\infty) \le L2^{k+\ell_1} \le L2^k\sqrt{N}\ . \tag{3.88}$$

*Proof.* Generally speaking, a set $T$ of cardinality $\le N_m$ and diameter $\Delta$ satisfies $\gamma_1(T,d) \le L\Delta m$, as is shown by taking $\mathcal{A}_n = \{T\}$ for $n \le m$ and $A_m(t) = \{t\}$. □

On the set of closed simple curves traced on $G$, we define the distance $d_1$ by $d_1(C,C') = \lambda(\overset{o}{C}\,\Delta\,\overset{o}{C}{}')$. We will deduce Proposition 3.7.4 from the following.

**Proposition 3.7.8.** *We have*

$$\gamma_{1,2}(\mathcal{C}(\tau,k),\, d_1) \le L2^{2k}\ . \tag{3.89}$$

*Proof of Proposition 3.7.4.* On the set of simple curves traced on $G$ we consider the distance

$$d_2(C_1,C_2) = \left\|\mathbf{1}_{\overset{o}{C}_1} - \mathbf{1}_{\overset{o}{C}_2}\right\|_2 = (\lambda(\overset{o}{C}\,\Delta\,\overset{o}{C}{}'))^{1/2} = \big(d_1(C_1,C_2)\big)^{1/2}\ , \tag{3.90}$$

so that

$$\gamma_2(\mathcal{F},d_2) = \gamma_2(\mathcal{C}(\tau,k),d_2) = \gamma_2(\mathcal{C}(\tau,k),\sqrt{d_1})\ .$$

When $k \le \ell_1 + 2$ we have $m := k+\ell_1+1 \le L\log N$, so that combining Proposition 3.7.8 with Lemma 3.7.6 yields the desired result. □

Proposition 3.7.8 is basically obvious because the metric space $(\mathcal{C}(\tau,k),d_1)$ is a Lipschitz image of a subset of the set $\mathcal{L}$ of Proposition 3.2.8. The elementary proof of the following may be found in Section A.2.

**Lemma 3.7.9.** *There exists a map $W$ from a subset $T$ of $\mathcal{L}$ onto $\mathcal{C}(\tau,k)$ which for any $f_1, f_2 \in T$ satisfies*

$$d_1(W(f_0),W(f_1)) \le L2^{2k}\|f_0 - f_1\|_2\ . \tag{3.91}$$

Finally we check the obvious fact that the functionals $\gamma_{\alpha,\beta}$ behave as expected under Lipschitz maps.

**Lemma 3.7.10.** *Consider two metric spaces $(T,d)$ and $(U,d')$. If $f:(T,d)\to(U,d')$ is onto and satisfies*

$$\forall x,y\in T\ ,\ d'(f(x),f(y)) \le Ad(x,y)$$

*for a certain constant $A$, then*

$$\gamma_{\alpha,\beta}(U,d') \le K(\alpha,\beta)A\gamma_{\alpha,\beta}(T,d)\,.$$

*Proof.* This is really obvious when $f$ is one to one. We reduce to that case by considering a map $\varphi : U \to T$ with $f(\varphi(x)) = x$ and replacing $T$ by $\varphi(U)$. □

*Proof of Proposition 3.7.8.* Combine Proposition 3.2.8 with Lemmas 3.7.10 and 3.7.9. □

Using Proposition 3.7.8, Corollary 3.7.7 and Theorem 3.5.12 we have proved that

$$\mathsf{E} \sup_{C\in\mathcal{C}(\tau,k)} \Big|\sum_{i\le N} \big(\mathbf{1}_{\overset{o}{C}}(X_i) - \lambda(\overset{o}{C})\big)\Big| \le L2^k\sqrt{N}(\log N)^{3/4} \,. \tag{3.92}$$

Only one difficulty remains: the control in probability required by Proposition 3.7.3. Either one tries to do it by hand[7] or one brings out some big guns. The "by hand approach" is messy and one does not learn much from it, so we go for the guns, here in the form of concentration of measure. The function

$$f(x_1,\dots,x_n) = \sup_{C\in\mathcal{C}(\tau,k)} \Big|\sum_{i\le N} \big(\mathbf{1}_{\overset{o}{C}}(x_i) - \lambda(\overset{o}{C})\big)\Big|$$

of points $x_1,\dots,x_n \in [0,1]^2$ has the property that changing the value of a given variable $x_i$ can change the value of $f$ by at most one. One of the earliest "concentration of measure" results (for which we refer to [60]) asserts that for such a function the r.v. $W = f(X_1,\dots,X_n)$ satisfies a deviation inequality of the form

$$\mathsf{P}(|W - \mathsf{E}W| \ge u) \le \exp\Big(-\frac{u^2}{2N}\Big)$$

and combining this with (3.92) proves the desired result.

It remains to deduce Theorem 3.7.1 from Theorem 3.7.2. The argument is purely deterministic and unrelated to any other material in the present book. The basic idea is very simple, and to keep it simple we describe it in slightly imprecise terms. Consider a union $A$ of little squares of side $2^{-\ell_1}$ and the union $A'$ of all the little squares that touch $A$. We want to prove that $A'$ contains as many points $Y_i$ as $A$ contains points $X_i$, so that by Hall's Marriage Lemma each point $X_i$ can be matched to a point $Y_i$ in the same little square, or in a neighbor of it. Since the points $Y_i$ are evenly spread the number of such points in $A'$ is very nearly $N\lambda(A')$. There may be more than $N\lambda(A)$ points $X_i$ in $A$, but (3.83) tells us that the excess number of points cannot be more than a proportion of the length $\ell$ of the boundary of $A$. The marvelous fact is that we may also expect that $\lambda(A') - \lambda(A)$ is also proportional to $\ell$, so that we may hope that the excess number of points $X_i$ in $A$ should not exceed $N(\lambda(A') - \lambda(A))$, proving the result. The proportionality constant is not quite right to make the argument work, but this difficulty is bypassed simply by applying the same argument to a slightly coarser grid.

[7] This is done in [145].

When one tries to describe precisely what is meant by the previous argument, one has to check a number of details. This elementary task which requires patience is performed in Section A.3.

## 3.8 Lower Bound for the Leighton-Shor Theorem

**Theorem 3.8.1.** *If the points* $(X_i)_{i\le N}$ *are i.i.d. uniform over* $[0,1]^2$ *and the points* $(Y_i)_{i\le N}$ *are evenly spread, then*

$$\mathsf{E}\inf_{\pi}\max_{i\le N} d(X_i, Y_{\pi(i)}) \ge \frac{(\log N)^{3/4}}{L\sqrt{N}} \ . \tag{3.93}$$

We denote by $\mathcal{C}$ the class of functions $f : [0,1] \to [0,1]$ such that $f(0) = f(1) = 0$, $\int_0^1 f'^2 \mathrm{d}\lambda \le 1$, and by $\mathcal{S}$ the class of their subgraphs

$$S(f) = \{(x,y) \in [0,1]^2 \ ; \ y \le f(x)\} \ .$$

We will show that with high probability we may find $f \in \mathcal{C}$ with

$$\mathrm{card}\{i \le N \ ; \ X_i \in S(f)\} \ge \lambda(S(f)) + \frac{1}{L}\sqrt{N}(\log N)^{3/4} \ . \tag{3.94}$$

**Lemma 3.8.2.** *The set of points within distance* $\epsilon$ *of the graph of* $f$ *has an area* $\le L\epsilon$.

*Proof.* The graph of $f \in \mathcal{C}$ has length $\int_0^1 \sqrt{1+f'^2}\mathrm{d}\lambda \le 2$. One can find a subset of the graph of $f$ of cardinality $\le L/\epsilon$ such that each point of the graph is within distance $\epsilon$ of this set. A point within distance $\epsilon$ of the graph then belongs to one of $L/\epsilon$ balls of radius $2\epsilon$.

*Proof of Theorem 3.8.1.* Let us denote by $S(f)_\epsilon$ the $\epsilon$-neighborhood of $S(f)$ in $[0,1]^2$. We first observe that for any $f \in \mathcal{C}$ we have

$$\mathrm{card}\{i \le N \ ; \ Y_i \in S(f)_\epsilon\} \le N\lambda(S(h)) + L\epsilon + L\sqrt{N} \ . \tag{3.95}$$

This is because by definition of an evenly spread family, each point $Y_i$ belongs to a small rectangle $R_i$ of area $1/N$ and of diameter $\le 10/\sqrt{N}$, and a pessimistic bound for the left-hand side above is the number of such rectangles that meet $S(f)_\epsilon$. These rectangles are entirely contained in the set of points within distance $L/\sqrt{N}$ of $S(f)_\epsilon$, and by lemma 3.8.2 this set has area $\le S(h) + L\epsilon + L\sqrt{N}$, hence the bound (3.95).

Consequently (3.94) implies that for $\epsilon = (\log N)^{3/4}/(L\sqrt{N})$ it holds

$$\mathrm{card}\{i \le N \ ; \ Y_i \in S_\epsilon(f)\} < \mathrm{card}\{i \le N \ ; \ X_i \in S(f)\} \ ,$$

and therefore any matching must pair at least one point $X_i \in S(f)$ with a point $Y_j \notin S_\epsilon(f)$, so that $\max_{i\le N} d(X_i, Y_{\pi(i)}) \ge \epsilon$. □

Recalling the functions $f_{k,\ell}$ of (3.67), we consider now an integer $c \geq 2$ which will be determined later ($c = 2$ works) and we set $\tilde{f}_{k,\ell} = f_{ck,\ell}/8 \geq 0$, so that

$$\int \tilde{f}_{k,\ell} \mathrm{d}\lambda = \frac{2^{-2c-6}}{\sqrt{r}} .$$

Consider the functions of the type

$$f_k = \sum_{1 \leq \ell \leq 2^{ck}} x_{k,\ell} \tilde{f}_{k,\ell} , \tag{3.96}$$

where $x_{k,\ell} \in \{0, 1\}$. Then $f(0) = f(1) = 0$.

**Lemma 3.8.3.** *A function $f$ of the type (3.96) satisfies*

$$\int_0^1 f'(x)^2 \mathrm{d}x \leq 2^{-8} . \tag{3.97}$$

*Proof.* Using (3.67) and (3.69) we obtain

$$\int_0^1 f'(x)^2 \mathrm{d}x \leq \frac{1}{32} \sum_{k \leq r} \sum_{\ell \leq 2^{ck}} x_{k,\ell}^2 \frac{1}{4r} 2^{-ck} \leq 2^{-8} . \qquad \square$$

Consequently each function of the type (3.96) belongs to $\mathcal{C}$.

*Almost correct proof of (3.94).*[8] Given $N$ large we choose $r$ as the largest integer for which $2^{cr} \leq N^{1/100}$, so that $r \geq \log N/L$. We construct inductively with high probability the functions $f_k \geq 0$ of the type (3.96) such that, setting $g = \sum_{k \leq q} f_k$

$$\operatorname{card}\{i \leq N \; ; \; X_i \in S(g+f_{q+1}) \setminus S(g)\} \geq \frac{\sqrt{N}}{L r^{1/4}} + \lambda(S(g+f_{q+1}) \setminus S(g)). \tag{3.98}$$

Summation of these inequalities over $q < r$ then proves (3.94).

The basic principle is that given a subset $A$ of the square, with $1/N \ll \lambda(A) \leq 1/2$, the number of points $X_i$ which belong to $A$ has typical fluctuations of order $\sqrt{N\lambda(A)}$. The area between the graph of $g$ and the graph of $g + \tilde{f}_{q+1,\ell}$ is $2^{-2c(q+1)-6}/\sqrt{r}$, so with probability $\geq 1/4$ it will contain an excess of points $X_i$ (compared to its expected value) of at least $2^{-c(q+1)}\sqrt{N}/L r^{1/4}$. In that case we set $x_{q+1,\ell} = 1$ and otherwise we set $x_{q+1,\ell} = 0$. With high probability there will be at least a fixed proportion of the $2^{cq}$ possible values of $\ell$ for which $x_{q+1,\ell} = 1$, and that will achieve (3.98).

The argument however contains a fatal flaw, because the decision we make at a step of the argument is based on some assumption about the number of points in a certain region, and this influences what happens at later stages

[8] The proof we give is in the spirit of the original proof of [63]. It is perfectly correct. If you find it too informal, a more formal approch is presented in [145].

of the construction. Specifically when $x_{q+1,\ell} = 0$ we know that the region $S(g + f_{q+1}) \setminus S(g)$ has is a sense a deficit of points, and at later stages we perform constructions inside this region. We will sketch how to correct this problem. Assume that the points $X_i$ are determined by a Poisson point process of intensity $N$ so that what happens in non-overlapping regions is independent, and for $\ell \leq 2^{ck}$ we define

$$\hat{f}_{k,\ell} = \sum_{s>k} \tilde{f}_{s,\ell'} \ , \tag{3.99}$$

where for each $s$ the sum is over the values of $\ell'$ such that $[\ell' 2^{-cs}, (\ell' + 1)2^{-cs}] \subset [\ell 2^{-ck}, (\ell + 1)2^{-ck}]$. Thus

$$\|\hat{f}_{k,\ell}\|_1 \leq \frac{1}{\sqrt{r}} \sum_{s>k} 2^{c(s-k)} 2^{-2cs-6} \leq 2^{1-c} \|\tilde{f}_{k,\ell}\|_1 \ .$$

Consequently, the region

$$S(g + \tilde{f}_{q+1,\ell}) \setminus S(g + \hat{f}_{q+1,\ell})$$

has an area at least $2^{-2c(q+1)-7}/\sqrt{r}$ and we select $x_{q+1,\ell} = 1$ when *this* region has an excess of points $X_i$. These points will not be involved in making decisions at further stages of the construction. A new obstacle arises: there could a deficit of points in the region

$$S(g + \hat{f}_{q+1,\ell}) \setminus S(g) \ .$$

However the union of all such regions in the construction is determined by a procedure which does not look at the points $X_i$ in it, so it will contain the right number of points within only fluctuations of order $\sqrt{N}$.

## 3.9 Notes and Comments

The original proof of the Leighton-Shor theorem amount basically to perform by hand a kind generic chaining in this highly non-trivial case, an incredible tour de force. A first attempt was made in [98] to relate (an important consequence of) the Leighton-Shor theorem to general methods for bounding stochastic processes, but runs into technical complications. Then Coffman and Shor [31] introduced the use of Fourier transforms and brought to light the role of ellipsoids, after which it became clear that the structure of these ellipsoids plays a central part in these matching results, a point of view systematically expounded in [127].

Chapter 14 is a continuation of the present chapter. The more difficult material it contains is presented later for fear of scaring readers at this early stage. A notable feature of the result presented there is that ellipsoids do not suffice, a considerable source of complication.

One could wonder for which kind of probability distributions on the unit square Theorem 3.5.1 remains true. The intuition that the uniform distribution considered in Theorem 3.5.1 is the "worst possible" is correct as is proved in [123]. The proof is overall similar to that of Theorem 3.5.1 but one has to find an appropriate substitute for Fourier transforms. The situation is different for Theorem 3.7.1, as is shown by the trivial example of a distribution concentrated at exactly two points at distance $d$ (where the best matching typically requires moving some of the random points for a distance $d$).

The original results of [5] are proved using an interesting technique called the *transportation method*. A version of this method, which avoids many of the technical difficulties of the original approach is presented in [146]. With the notation of Theorem 3.5.1, it is proved in [146] (a stronger version of the fact) that with probability $\geq 9/10$ one has

$$\inf_{\pi} \frac{1}{N} \sum_{i \leq N} \exp\Big(\frac{N d(X_i, Y_{\pi(i)})^2}{K \log N}\Big) \leq 2 \ . \tag{3.100}$$

Since $\exp x \geq x$, (3.100) implies that $\sum_{i \leq N} d(X_i, Y_{\pi_i})^2 \leq \log N$ and hence using the Cauchy-Schwarz inequality

$$\sum_{i \leq N} d(X_i, Y_{\pi(i)}) \leq L \sqrt{N \log N} \ . \tag{3.101}$$

Moreover it implies also

$$\max_{i \leq N} d(X_i, Y_{\pi(i)}) \leq L \log N / \sqrt{N} \ . \tag{3.102}$$

It does not seem known whether one can achieve simultaneously (3.101) and $\max_{i \leq N} d(X_i, Y_{\pi(i)}) \leq L(\log N)^{3/4}/\sqrt{N}$. In this circle of idea, see the ultimate matching conjecture on page 369.

For results about matching for unbounded distributions, see the work of J. Yukich [158].

Methods similar to those of this chapter may be used to obtain non-trivial discrepancy theorems for various classes of functions, as investigated in [127]. Let us mention one such result. We denote by $\lambda$ the uniform probability on the unit cube $[0,1]^3$, and by $(X_i)_{i \leq N}$ independent uniformly distributed r.v.s valued in this unit cube.

**Theorem 3.9.1.** *Consider the class $\mathcal{C}$ of convex sets in $\mathbb{R}^3$. Then*

$$\mathsf{E} \sup_{C \in \mathcal{C}} |\operatorname{card}\{i \leq N \ ; \ X_i \in C\} - N\lambda(C)| \leq L\sqrt{N}(\log N)^{3/4} \ .$$

# 4. Mostly Trees

The concept of tree presented in Section 4.1 is historically important: the author discovered many of the results he presents while thinking in terms of trees. We know now how to present these results and their proofs without ever mentioning trees, arguably in a more elegant fashion, so that trees are not used explicitly elsewhere in this book. However it might be too early to dismiss this concept, at least as an instrument of discovery.

## 4.1 Trees

We shall describe different ways to measure the size of a metric space and show that they are all equivalent to the functional $\gamma_2(T,d)$.[1]

In a nutshell, a tree is a certain structure that requires a "lot of space" to be constructed, so that a metric space that contains large trees needs itself to be large. At the simplest level, it already takes some space to construct in a set $A$ sets $B_1, \ldots, B_n$ which are appropriately separated from each other. This is even more so if the sets $B_1, \ldots, B_n$ are themselves large (for example because they themselves contain many sets far from each other). Trees are a proper formulation of the iteration of this idea. The basic use of trees is to measure the size of a metric space by the size of the largest tree (of a certain type) which it contains. Different types of trees yield different measures of size.

A *tree* $\mathcal{T}$ of a metric space $(T,d)$ is a *finite* collection of subsets of $T$ with the following two properties.

$$\text{Given } A, B \text{ in } \mathcal{T}\,,\ \text{ if } A \cap B \neq \emptyset\,,\ \text{ then either } A \subset B \text{ or else } B \subset A\,. \tag{4.1}$$

$$\mathcal{T} \text{ has a largest element}\,. \tag{4.2}$$

The important condition here is (4.1), and (4.2) is just for convenience.

If $A, B \in \mathcal{T}$ and $B \subset A$, $B \neq A$, we say that $B$ is a *child* of $A$ if

$$C \in \mathcal{T}\,,\, B \subset C \subset A \Rightarrow C = B \text{ or } C = A\,. \tag{4.3}$$

[1] It is possible to consider more general notions corresponding to other functionals considered in the book, but for simplicity we consider only the case of $\gamma_2$.

We denote by $c(A)$ the number of children of $A$. Since our trees are finite, some of their sets will have no children. It is convenient to "shrink these sets to a single point", so we will consider only trees with the following property

$$\text{If } A \in \mathcal{T} \text{ and } c(A) = 0\,, \text{ then } A \text{ contains exactly one point}\,. \tag{4.4}$$

A fundamental property of trees is as follows. Consider trees $\mathcal{T}_1, \ldots, \mathcal{T}_m$ and for $1 \le \ell \le m$ let $A_\ell$ be the largest element of $\mathcal{T}_\ell$. Assume that the sets $A_\ell$ are disjoint, and consider a set $A \supset \bigcup_{\ell \le m} A_\ell$. Then the collection of subsets of $T$ consisting of $A$ and of $\bigcup_{\ell \le m} \mathcal{T}_\ell$ is a tree. The proof is straightforward. This fact allows one to construct iteratively more and more complicated (and larger) trees.

An important structure in a tree is a *branch*. A sequence $A_0, A_1, \ldots, A_k$ is a branch if $A_{\ell+1}$ is a child of $A_\ell$, and if moreover $A_0$ is the largest element of $\mathcal{T}$ while $A_k$ has no child. Then by (4.4) the set $A_k$ is reduced to a single point $t$, and $A_0, \ldots, A_k$ are exactly those elements of $\mathcal{T}$ which contain $t$. So in order to describe the branches of $\mathcal{T}$ it is convenient to introduce the set

$$S_{\mathcal{T}} = \{t \in T\ ;\ \{t\} \in \mathcal{T}\}\ , \tag{4.5}$$

which we call the *support* of $\mathcal{T}$. Thus by considering all the sets $\{A \in \mathcal{T}; t \in A\}$ as $t$ varies in $S_{\mathcal{T}}$ we obtain all the branches of $\mathcal{T}$.

We now quantify our desired property that the children of a given set should be far from each other in an appropriate sense. A *separated* tree is a tree $\mathcal{T}$ such that to each $A$ in $\mathcal{T}$ with $c(A) \ge 1$ is associated an integer $s(A) \in \mathbb{Z}$ with the following properties. First,

$$\text{If } B_1 \text{ and } B_2 \text{ are distinct children of } A, \text{then } d(B_1, B_2) \ge 4^{-s(A)}\,. \tag{4.6}$$

Here $d(B_1, B_2) = \inf\{d(x_1, x_2); x_1 \in B_1, x_2 \in B_2\}$. We observe that in (4.6) we make no restriction on the diameter of the children of $A$. (Such restrictions will however occur in the other notion of tree that we consider later.) Second, we will also make the following purely technical assumption:

$$\text{If } B \text{ is a child of } A\,, \text{ then } s(B) > s(A)\,. \tag{4.7}$$

Although this is not obvious now, the meaning of this condition is that $\mathcal{T}$ contains no sets which are obviously irrelevant for the measure of its size.

To measure the size of a separated tree $T$ we introduce its *depth*, i.e.

$$d(\mathcal{T}) := \inf_{t \in S_{\mathcal{T}}} \sum_{t \in A \in \mathcal{T}} 4^{-s(A)} \sqrt{\log c(A)}\,. \tag{4.8}$$

Here and below we make the convention that the summation does not include the term $A = \{t\}$ (for which $c(A) = 0$). We observe that in (4.8) we have the *infimum* over $t \in S_{\mathcal{T}}$. In words a tree is large if it is large along *every* branch. We can then measure the size of $T$ by

$$\sup\{d(\mathcal{T})\ ;\ \mathcal{T} \text{ separated tree} \subset T\}\ . \tag{4.9}$$

The notion of separated tree we just considered is but one of many possible notions of trees. As it turns out, this notion of separated tree does not seem fundamental. Rather, the quantity (4.9) is used as a convenient intermediate technical step to prove the equivalence of several more important quantities. Let us now consider now another notion of trees, which is more restrictive (and apparently much more important). An *organized* tree is a tree $\mathcal{T}$ such that to each $A \in \mathcal{T}$ with $c(A) \geq 1$ are associated an integer $j = j(A) \in \mathbb{Z}$, and points $t_1, \ldots, t_{c(A)}$ with the properties that

$$1 \leq \ell < \ell' \leq c(A) \Rightarrow 4^{-j-1} \leq d(t_\ell, t_{\ell'}) \leq 4^{-j+2}$$

and that each ball $B(t_\ell, 4^{-j-2})$ contains exactly one child of $A$. Please note that it may happen that $4^{-j}$ is much smaller than $\Delta(A)$.

If $B_1$ and $B_2$ are distinct children of $A$ in an organized tree, then

$$d(B_1, B_2) \geq 4^{-j(A)-2}\,, \tag{4.10}$$

so that an organized tree is also a separated tree, with $s(A) = j(A) + 2$, but the notion of organized tree is more restrictive. (For example we have no control over the diameter of the children of $A$ in a separated tree.)

We define the depth $d'(\mathcal{T})$ of an organized tree by

$$d'(\mathcal{T}) = \inf_{t \in S_{\mathcal{T}}} \sum_{t \in A \in \mathcal{T}} 4^{-j(A)} \sqrt{\log c(A)}\ .$$

Another way to measure the size of $T$ is then

$$\sup\{d'(\mathcal{T})\ ;\ \mathcal{T} \text{ organized tree} \subset T\}\ . \tag{4.11}$$

If we simply view an organized tree $\mathcal{T}$ as a separated tree using (4.10), then $d(\mathcal{T}) = d'(\mathcal{T})/16$ (where $d(\mathcal{T})$ is the depth of $\mathcal{T}$ as a separated tree). Thus we have shown the following.

**Proposition 4.1.1.** *We have*

$$\sup\{d'(\mathcal{T})\ ;\ \mathcal{T} \text{ organized tree}\} \leq 16 \sup\{d(\mathcal{T})\ ;\ \mathcal{T} \text{ separated tree}\}\ . \tag{4.12}$$

The next result provides the fundamental connection between trees and the functional $\gamma_2$.

**Proposition 4.1.2.** *We have*

$$\gamma_2(T, d) \leq L \sup\{d'(\mathcal{T})\ ;\ \mathcal{T} \text{ organized tree}\}\ . \tag{4.13}$$

*Proof.* We consider the functional

$$F(A) = \sup\{d'(\mathcal{T})\ ;\ \mathcal{T} \subset A\,,\ \mathcal{T} \text{ organized tree}\}\,,$$

where we write $\mathcal{T} \subset A$ as a shorthand for "$\forall B \in \mathcal{T}\,,\ B \subset A$".

Next we prove that this functional satisfies the growth condition (2.55) for $r = 16$ whenever $a$ is of the type $16^{-j}$, for $c^* = 1/L$. For this consider $n \geq 1$ and $m = N_n$. Consider $j \in \mathbb{Z}\,,\ t \in T$ and $t_1, \cdots, t_m$ with

$$1 \leq \ell < \ell' \leq m \Rightarrow 16^{-j} \leq d(t_\ell\,,\ t_{\ell'}) \leq 2 \cdot 16^{-j+1}\,. \tag{4.14}$$

Consider sets $H_\ell \subset B(t_\ell, 2\cdot 16^{-j-1})$ and $c < \min_{\ell \leq m} F(H_\ell)$. Consider, for $\ell \leq m$ a tree $\mathcal{T}_\ell \subset H_\ell$ with $d'(\mathcal{T}_\ell) > c$ and denote by $A_\ell$ its largest element. Then it should be obvious that the tree $\mathcal{T}$ consisting of $C = \bigcup_{\ell \leq m} H_\ell$ (its largest element) and the union of the trees $\mathcal{T}_\ell\,,\ \ell \leq m$, is organized (with $j(C) = 2j - 1$, and $A_1, \ldots, A_m$ as children of $C$, and since $4^{-j(C)-1} = 16^{-j}$ and $2 \cdot 16^{-j+1} \leq 4^{-j(C)+2}$). Moreover $S_{\mathcal{T}} = \bigcup_{\ell \leq m} S_{\mathcal{T}_\ell}$.

Consider $t \in S_{\mathcal{T}}$, and let $\ell$ with $t \in S_{\mathcal{T}_\ell}$. Then

$$\begin{aligned}\sum_{t \in A \in \mathcal{T}} 4^{-j(A)}\sqrt{\log c(A)} &= 4 \cdot 16^{-j}\sqrt{\log m} + \sum_{t \in A \in \mathcal{T}_\ell} 4^{-j(A)}\sqrt{\log c(A)} \\ &\geq 4 \cdot 16^{-j}\sqrt{\log m} + d'(\mathcal{T}_\ell) \geq \frac{1}{L}16^{-j}2^{n/2} + c\,.\end{aligned}$$

This completes the proof of the growth condition (2.55).

It should be obvious that Theorem 2.4.1 requires only the growth condition (2.55) to hold true when $a$ is of the type $r^{-j}$, and we have just proved that this is the case (for $r = 16$), so that from (2.59) we have proved that $\gamma_1(T, d) \leq L(F(T) + \Delta(T))$. It remains only to prove that $\Delta(T) \leq LF(T)$. For this we simply note that if $s, t \in T$, and $j$ is the largest integer with $4^{-j} \geq d(s, t)$, then the tree $\mathcal{T}$ consisting of $T, \{t\}, \{s\}$, is organized with $j = j(T)$ and $c(T) = 2$, so $d'(\mathcal{T}) \geq 4^{-j}\sqrt{\log 2}$ and $4^{-j} \leq F(T)$. □

For a probability measure $\mu$ on a metric space $(T, d)$, with countable support, we define for each $t \in T$ the quantity

$$I_\mu(t) := \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon = \int_0^{\Delta(T)} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon\,.$$

The equality follows from the fact that $\mu(B(t,\epsilon)) = 1$ when $B(t,\epsilon) = T$, so that then the integrand is 0.

**Proposition 4.1.3.** *Given a metric space $(T, d)$ we can find on $T$ a probability measure $\mu$, supported by a countable subset of $T$, and such that*

$$\sup_{t \in T} I_\mu(t) = \sup_{t \in T} \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon \leq L\gamma_2(T, d)\,. \tag{4.15}$$

A probability measure $\mu$ on $(T,d)$ is called a majorizing measure.[2] The reason for this somewhat unsatisfactory name is that X. Fernique proved that for a Gaussian process and probability measure $\mu$ on $T$ one has

$$\mathsf{E}\sup_{t\in T} X_t \le L \sup_{t\in T} I_\mu(t)\ , \tag{4.16}$$

so that $\mu$ can be used to "majorize" the process $(X_t)_{t\in T}$. This was a major advance over Dudley's bound. The (in)famous theory of majorizing measures used the quantity

$$\inf_\mu \sup_{t\in T} I_\mu(t) \tag{4.17}$$

as a measure of the size of the metric space $(T,d)$, where the infimum is over all choices of the probability measure $\mu$. This method is technically quite challenging. It has been replaced by the generic chaining. A related idea which is still very useful is explained in Section 4.2.

*Proof.* Consider an admissible sequence $(\mathcal{A}_n)$ with

$$\forall t\in T\ ,\ \sum_{n\ge 0} 2^{n/2}\Delta(A_n(t)) \le 2\gamma_2(T,d)\ .$$

Let us now pick a point $t_{n,A}$ in each set $A\in\mathcal{A}_n$, for each $n\ge 0$. Since card $\mathcal{A}_n \le N_n$, there is a probability measure $\mu$ on $T$, supported by a countable set, and satisfying $\mu(\{t_{n,A}\}) \ge 1/(2^n N_n)$ for each $n\ge 0$ and each $A\in\mathcal{A}_n$. Then,

$$\forall n\ge 1\ ,\ \forall A\in\mathcal{A}_n\ ,\ \mu(A)\ge \mu(\{t_{n,A}\}) \ge \frac{1}{2^n N_n} \ge \frac{1}{N_n^2}$$

so that given $t\in T$ and $n\ge 1$,

$$\begin{aligned}\epsilon > \Delta(A_n(t)) &\Rightarrow \mu(B(t,\epsilon)) \ge \frac{1}{N_n^2}\\ &\Rightarrow \sqrt{\log\frac{1}{\mu(B(t,\epsilon))}} \le 2^{n/2+1}\ .\end{aligned} \tag{4.18}$$

Now, since $\mu$ is a probability, $\mu(B(t,\epsilon)) = 1$ for $\epsilon > \Delta(T)$, and then $\log(1/\mu(B(t,\epsilon))) = 0$. Thus

$$\begin{aligned}I_\mu(t) = \int_0^\infty \sqrt{\log\frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon &= \sum_{n\ge 1}\int_{\Delta(A_n(t))}^{\Delta(A_{n-1}(t))} \sqrt{\log\frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon\\ &\le \sum_{n\ge 1} 2^{n/2+1}\Delta(A_{n-1}(t)) \le L\gamma_2(T,d)\end{aligned}$$

using (4.18). □

[2] One typically uses the name only when such the left-hand side of (4.15) is usefully small

**Proposition 4.1.4.** *If $\mu$ is a probability measure on $T$ (supported by a countable set) and $\mathcal{T}$ is a separated tree on $T$, then*

$$d(\mathcal{T}) \le L \sup_{t \in T} I_\mu(t) \ .$$

This completes the proof that the four "measures of the size of $T$" considered in this section, namely (4.9), (4.11), (4.17) and $\gamma_2(T,d)$ are indeed equivalent.

*Proof.* The basic observation is as follows. The sets

$$B(C, 4^{-s(A)-1}) = \{x \in T \ ; \ d(x,C) < 4^{-s(A)-1}\}$$

are disjoint as $C$ varies over the children of $A$ (as follows from (4.6)), so that one of them has measure $\le c(A)^{-1}$.

We then proceed in the following manner, constructing recursively an appropriate branch of the tree. This is a typical and fundamental way to proceed when working with trees. We start with the largest element $A_0$ of $\mathcal{T}$. We then select a child $A_1$ of $A_0$ with $\mu(B(A_1, 4^{-s(A_0)-1})) \le 1/c(A_0)$, and a child $A_2$ of $A_1$ with $\mu(B(A_2, 4^{-s(A_1)-1})) \le 1/c(A_1)$, etc., and continue this construction as long as we can. It ends only when we reach a set of $\mathcal{T}$ that has no child, and hence by (4.4) is reduced to a single point $t$ which we now fix. For any set $A$ with $t \in A \in \mathcal{T}$, by construction we have

$$\mu(B(t, 4^{-s(A)-1})) \le \frac{1}{c(A)}$$

so that

$$4^{-s(A)-2}\sqrt{\log c(A)} \le \int_{4^{-s(A)-2}}^{4^{-s(A)-1}} \sqrt{\frac{1}{\log \mu(B(t,\epsilon))}} \mathrm{d}\epsilon \ . \tag{4.19}$$

By (4.7) the intervals $]4^{-s(A)-2}, 4^{-s(A)-1}[$ are disjoint for different sets $A$ with $t \in A \in \mathcal{T}$, so summation of the inequalities (4.19) yields

$$\frac{1}{16} d(\mathcal{T}) \le \sum_{t \in A \in \mathcal{T}} 4^{-s(A)-2}\sqrt{\log c(A)} \le \int_0^\infty \sqrt{\frac{1}{\log \mu(B(t,\epsilon))}} \mathrm{d}\epsilon = I_\mu(t) \ . \quad \square$$

In the rest of this chapter, we will implicitly use the previous method of "selecting recursively the branch of the tree we follow" to prove lower bounds without mentioning trees.

## 4.2 Witnessing Measures

We start by a simple useful fact.

**Proposition 4.2.1.** *Consider a probability measure $\mu$ on a metric space $(T,d)$. Then*

$$\int I_\mu(t)\mathrm{d}\mu(t) \le L\gamma_2(T,d) . \tag{4.20}$$

*Proof.* Consider an admissible sequence $\mathcal{A}_n$ of partitions with

$$\forall t \in T \; ; \; \sum_{n\ge 0} 2^{n/2}\Delta(A_n(t)) \le 2\gamma_2(T,d) . \tag{4.21}$$

Then $\mu(B(t,\epsilon)) \ge \mu(A_n(t))$ for $\epsilon \ge \Delta(A_n(t))$, so that

$$I_\mu(t) \le \sum_{n\ge 1} \Delta(A_{n-1}(t))\sqrt{\log\frac{1}{\mu(A_n(t))}} .$$

For $n \ge 1$ and $A \in \mathcal{A}_n$ let us denote by $A'$ the unique element of $\mathcal{A}_{n-1}$ which contains $A$. By integrating the previous inequality we get

$$\begin{aligned}\int I_\mu(t)\mathrm{d}\mu(t) &\le \sum_{n\ge 1}\sum_{A\in\mathcal{A}_n} \Delta(A')\mu(A)\sqrt{\log\frac{1}{\mu(A)}} \\ &\le L\sum_{n\ge 1}\sum_{A\in\mathcal{A}_n} \Delta(A')\mu(A)2^{n/2} + L\Delta(T) . \end{aligned} \tag{4.22}$$

To obtain the second inequality we group in the first term the contributions for which $\mu(A) \ge N_{n+1}^{-1}$ and in the second term the other contributions, remembering that $\operatorname{card} A_n \le N_n$ and observing that the function $x \mapsto x\sqrt{\log(1/x)}$ increases for small $x$. Finally integrating with respect to $\mu$ the inequality (4.21) provides the desired control of the first term in the right-hand side of (4.22). □

**Proposition 4.2.2.** *[[75]] For a metric space $(T,d)$, define*

$$\delta_2(T,d) = \sup_\mu \inf_{t\in T} I_\mu(t) ,$$

*where the supremum is taken over all probability measures $\mu$ on $T$.*[3] *Then*

$$\frac{1}{L}\gamma_2(T,d) \le \delta_2(T,d) \le L\gamma_2(T,d) .$$

A probability measure on $T$ will be called a *witnessing measure*, and its "size" is the quantity $\inf_{t\in T} I_\mu(t)$.[4] The important part of Proposition 4.2.2

[3] Please observe that the infimum and the supremum are not as in (4.17).

[4] Thus a probability measure $\mu$ on $T$ is both a majorizing and a witnessing measure. It bounds above $\gamma_2(T,d)$ by $C\sup_{t\in T} I_\mu(t)$ and it bounds from below $\gamma_2(T,d)$ by $\inf_{t\in T} I_\mu(t)/C$. Furthermore one may find $\mu$ such that these two bounds are of the same order.

is that we can find a witnessing measure of size about $\gamma_2(T,d)$. As the name should indicate, a witnessing measure witnesses that $T$ is large. The magic is simply that witnessing measures are convenient to organize counting arguments.

*Proof.* The right-hand side inequality follows from Proposition 4.2.1 and the trivial fact that $\inf_{t\in T} I_\mu(t) \le \int I_\mu(t)\mathrm{d}\mu(t)$. The reader should review the material of the previous section to follow the proof of the converse. Given an organized tree $\mathcal{T}$ we define a measure $\mu$ on $T$ by $\mu(A)=0$ if $A\cap\mathcal{S}(\mathcal{T})=\emptyset$ and by

$$t\in\mathcal{S}(\mathcal{T})\Rightarrow\mu(\{t\})=\frac{1}{\prod_{t\in A\in\mathcal{T}}c(A)}\ .$$

The intuition is that the mass carried by $A\in\mathcal{T}$ is equally divided between the children of $A$. Then $I_\mu(t)=\infty$ if $t\notin\mathcal{S}(\mathcal{T})$. Consider $t\in A\in\mathcal{T}$ and $j=j(A)$. Then $B(t,4^{-j-2})$ meets only one child of $A$ so that $\mu(B(t,4^{-j-2}))\le 1/c(A)$. This readily implies that $LI_\mu(t)\ge\sum_{t\in A\in\mathcal{T}}4^{-j(A)}\sqrt{\log c(A)}$ from which the result follows by (4.13). □

**Exercise 4.2.3.** For a metric space $(T,d)$ define

$$\chi_2(T,d)=\sup_\mu\inf\int\sum_{n\ge 0}2^{n/2}\Delta(A_n(t))\mathrm{d}\mu(t)\ ,$$

where the infimum is taken over all admissible sequences and the supremum over all probability measures. Prove that this measure of size is equivalent to $\gamma_2(T,d)$. It is obvious that $\chi_2(T,d)\le\gamma_2(T,d)$, but the converse is far from trivial. Hint: The same measure $\mu$ as in the proof of Proposition 4.2.2 works, but you may find proving this quite challenging.