MICHEL TALAGRAND

The Generic Chaining

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

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

Springer

Springer Monographs in Mathematics

Springer Monographs in Mathematics

Michel Talagrand

The Generic Chaining

Upper and Lower Bounds of Stochastic Processes

Springer

Michel Talagrand
Equipe d'Analyse fonctionelle
Institut Mathématiques de Jussieu
U.M.R. 7586 C.N.R.S.
Université Paris VI
Boite 186
4 Place Jussieu 75230 Paris, France

and

The Ohio-State University
281 West 18th Avenue
Columbus, OH 43210, USA
e-mail: mit@ccr.jussieu.fr

Library of Congress Control Number: 2005920702

Mathematics Subject Classification (2000): 60G15, 60G17, 60G52, 60D05, 60G70, 60B11

ISSN 1439-7382
ISBN 3-540-24518-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Typeset by the authors using a Springer LaTeX macro package
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: Erich Kirchner, Heidelberg

Printed on acid-free paper 44/3141/YL - 5 4 3 2 1 0

To my father

Contents

Introduction

What is the maximum level a certain river is likely to reach over the next 25 years? (Having experienced three times a few feet of water in my house, I feel a keen personal interest in this question.) There are many questions of the same nature: what is the likely magnitude of the strongest earthquake to occur during the life of a planned building, or the speed of the strongest wind a suspension bridge will have to stand? All these situations can be modeled in the same manner. The value X_t of the quantity of interest (be it water level or speed of wind) at time t is a random variable. What can be said about the maximum value of X_t over a certain range of t?

A collection of random variables (X_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 to find upper and lower bounds for the quantity

$$\mathsf{E} \sup_{t \in T} X_t \ . \tag{0.1}$$

Since T might be uncountable, some care has to be taken to define this quantity. For any reasonable definition of $\mathsf{E} \sup_{t \in T} X_t$ we have

$$\mathsf{E} \sup_{t \in T} X_t = \sup\{\mathsf{E} \sup_{t \in F} X_t \ ; \ F \subset T \ , \ F \text{ finite}\} \ , \tag{0.2}$$

an equality that we will take as the definition of the quantity $\mathsf{E} \sup_{t \in T} X_t$. Thus, the crucial case for the estimation of the quantity (0.1) is the case where T is finite, an observation that should stress that this book is mostly about inequalities.

The most important random variables (r.v.) are arguably Gaussian r.v. The study of conditions under which Gaussian processes are bounded (i.e. the quantity (0.1) is finite) goes back at least to Kolmogorov. The celebrated Kolmogorov conditions for the boundedness of a stochastic process are still useful today, but they are far from being necessary and sufficient. 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 the distance d is given by

$$d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2} . \tag{0.3}$$

In 1967, R. Dudley obtained a sharp sufficient condition for the boundedness of a Gaussian process, the so-called Dudley entropy condition. It is based on the fact that, for a Gaussian process,

$$\forall u > 0 \, , \; \mathsf{P}(|X_s - X_t| \geq u) \leq 2 \exp\left(-\frac{u^2}{2d(s,t)^2}\right) . \tag{0.4}$$

Dudley's condition is however not necessary. A few years later, X. Fernique (building on earlier ideas of C. Preston) introduced a condition based on the use of a new tool called majorizing measures. Fernique's condition is weaker than Dudley's, and Fernique conjectured that his condition was in fact necessary. Gilles Pisier suggested in 1983 that I should work on this conjecture, and kept goading me until I proved it in 1985, obtaining thus a necessary and sufficient condition for the boundedness of a Gaussian process, or equivalently, upper and lower bounds of the same order for the quantity (0.1) in terms of the structure of the metric space (T, d). A few years of great excitement followed this discovery, during which I proved a number of extensions of this result, or of parts of it, to other classes of processes. I was excited because I liked (and still like) these results. Unfortunately, I was about the only one to get excited. Part of the reason is that Fernique's concept of majorizing measures is very difficult to grasp at the beginning, and was consequently dismissed by the main body of probabilists as a mere curiosity. (I must admit that I did have a terrible time myself to understand it.)

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 a suitable variation on the usual chaining arguments, a variation that is moreover totally natural. That this was not discovered much earlier is a striking illustration of the inefficiency of the human brain (and of mine in particular). This new approach not only removes the psychological obstacle of having to understand the somewhat disturbing idea of majorizing measures, it also removes a number of technicalities, and allows one to give significantly shorter proofs. I thus felt the time had come to make a new exposition of my body of work on lower and upper bounds for stochastic processes. The feeling that, this time, the approach was possibly (and even probably) the correct one gave me the energy to rework all the proofs. For several of the most striking results, such as Shor's matching theorem, the decomposition theorem for infinitely divisible processes, and Bourgain's solution of the Λ_p problem, the proofs given here are at least three times shorter than the previously published proofs.

Beside enjoying myself immensely and giving others a chance to understand the results presented here (and even possibly to get excited about them) a main objective of this book is to point out several problems that remain open. Of course opinions differ as to what constitutes an important problem, but I like those presented here. One of them deals with the geometry of Hilbert space, a topic that can hardly be dismissed as exotic. I stated only the problems that I find really interesting. Possibly they are challenging. At least, I made every effort to make progress on them. A significant part of the material of the book was discovered while trying to solve the "Bernoulli problem" of Chapter 4. I have spent many years thinking to that problem, and will be glad to offer a prize of $ 5000 for a positive solution of it. A smaller prize of $ 1000 is offered for a positive solution of the possibly even more important problem raised at the end of chapter 5. The smaller amount simply reflects the fact that I have spent less time on this question than on the Bernoulli problem. It is of course advisable to claim these prizes before I am too senile to understand the solution, for there will be no guarantee of payment afterwards. (Cash awards will also be given for a negative solution of any of these two problems, the amount depending on the beauty of the solution.)

It is my pleasure to thank the Ohio State University and the National Science Foundation for supporting the typing of this book, and making its publication possible.

I must apologize for the countless inaccuracies and mistakes, small or big, that this book is bound to contain despite all the efforts made to remove them. I was very much helped in this endeavor by a number of colleagues, and in particular by A. Hanen and R. Latala, who read the entire book. Of course, all the remaining mistakes are my sole responsibility.

In conclusion, a bit of wisdom. I think that I finally discovered a foolproof way to ensure that the writing of a book of this size be a delightful and easy experience. Just write a 600 page book first!

1 Overview and Basic Facts

1.1 Overview of the Book

This section will describe the philosophy underlying this book, and some of its highlights. This will be done using words rather than formulas, so that the description is necessarily imprecise, and is only intended to provide some insight in 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 that will typically involve a rather rich and complicated structure. There is a near infinite supply of such models. Fashions come and go, and the importance with which we view any specific model is likely to strongly 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 the study of problems with a kind of minimal and intrinsic structure. Besides the fact that it is much easier to find the crux of the matter in a simple structure than in a complicated one, there are not so many really basic structures, so one can hope that they will remain of interest for a very long time. This book is devoted to the study of a few of these structures.

It is of course very nice to enjoy the feeling, real or imaginary, that one is studying structures that might be of intrinsic importance, but the success of the approach of studying "minimal structures" has ultimately to be judged by the results it obtains. It is a fact of life that general principles are, more often than not, insufficient to answer specific questions. Still, as we will demonstrate, they are able to explain in complete detail a number of fascinating and very deep facts.

The most important question considered in the book is the boundedness of Gaussian processes. As we already noticed, the intrinsic distance (0.3) points to the fact that the relevant object is the metric space (T, d) where T is the index set. This metric space is far from being arbitrary, since it is isometric to a subset of a Hilbert space. (By its very nature, this introduction is going to contain many statements, like the previous one, that might or might not look obvious to the reader, depending on his background. The best way to obtain complete clarification about these statements is to start reading from the next section on.) It turns out, quite surprisingly, that it is much better to forget this specificity of the metric space (T, d) and to just think of it as a general metric space. Since there is only so much one can do with a bare metric space structure, nothing can get really complicated then.

In Section 1.2 we explain the basic idea, how to get a sharp upper bound on a process satisfying the increment condition (0.4) through the "generic chaining". This simple bound involves only the structure of the metric space (T, d). The author feels that this result requires less energy than most books spend e.g. to prove the continuity of Brownian motion by weaker methods. Yet, it turns out to be the very best result possible. A convenient way to use the generic chaining bound is through sequences of partitions of the metric space (T, d), and in Section 1.3 we learn how to construct these partitions. This is the core result in the direction of lower bounds. This construction takes place in a general metric space, and while it is definitely non-trivial, it is not very complicated either. The reader who has never thought about metric spaces might disagree with this latter statement, so the best action to take in that case is simply to skip this proof and to judge the efficiency of the approach by the subsequent results.

In Section 2.1 we give the first application of the general tools of Chapter 1, the characterization of sample-boundedness of Gaussian processes. Gaussian processes are deeply related to the geometry of Hilbert space, and a number of basic questions in this direction remain unanswered. In Section 2.2, we investigate ellipsoids of a Hilbert space, and we explain why their structure as metric spaces (with the distance induced by the entire space) is not trivial from the point of view of Gaussian processes. Ellipsoids will play a basic role in Chapter 3.

It is natural to expect that this understanding of Gaussian processes will yield information on processes that are conditionally Gaussian. It turns out that p-stable processes, an important class of processes, are conditionally Gaussian, and in Section 2.3 we provide lower bounds for such processes. These bounds are the best possible of their type. Essentially more general (but more difficult) results are proved later in Chapter 5 for infinitely divisible processes. Another natural class of processes that are conditionally Gaussian are order 2 Gaussian chaos (these are essentially second degree polynomials of Gaussian r.v.). It seems at present a hopelessly difficult task to give lower and upper bounds of the same order for these processes, but in Section 2.5 we obtain a number of results in the right direction. The results of Section 2.5 are not used in the sequel.

In Section 2.6 we investigate the structure of subsets of the classical Banach space $L^2(\mu)$ from different points of view. Interpolation is the basic idea, in the sense that a generic subset U can be obtained by interpolation between two sets, each of which having in some respect a simpler structure than the set U itself. In particular, for one of the pieces of the decomposition, we achieve control not only in the L^2 norm, but also in the L^∞ norm, and this is very helpful to use Bernstein's inequality. The results of this section are abstract and somewhat technical, but are very useful in the long range. They are however not needed at all for Chapter 3, and it is advised that the casual reader jumps directly to this chapter at this point. Section 2.7 investigates on which classes of functions the empirical process behaves uniformly well, "with the convergence speed of the central limit theorem". The "geometry" of such classes can be described to a large extent in full generality. We then give a sharp version of Ossiander's bracketing theorem, a practical criteria to control the empirical process uniformly on a class of functions.

Chapter 3 is completely independent of the material of the last four sections of Chapter 2. It is devoted to matchings, or, equivalently, to the problem of understanding precisely how far N points independently and uniformly distributed in the unit square are from being "evenly spread". This is measured by the "cost" of pairing (=matching) these points with N fixed points that are very uniformly spread, for various notions of cost. The results of this chapter illustrate particularly well the benefits of an abstract point of view: we are able to trace some deep results about a simple concrete structure back to the geometry of ellipsoids. In the first section of the chapter, we investigate further the structure of ellipsoids as metric spaces. The philosophy of the main result, the Ellipsoid Theorem, is that an ellipsoid is in some sense somewhat smaller that what one might think at first. This is due to the fact that an ellipsoid is sufficiently convex, and that, somehow, it gets "thinner" when one gets away from its center. For the reader willing to accept this result without proof, only Section 1.2 is required reading for this chapter. The Ellipsoid Theorem is a special case of a more general result (with the same proof) about the structure of sufficiently convex bodies, that will have important applications in Chapter 6. In Section 3.3 we investigate the situation 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. In Section 3.4 we investigate the situation 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 $L(\log N)^{3/4}/\sqrt{N}$. The Ellipsoid Theorem explains the occurrence of these fractional powers of log in a transparent way. In both situations, the link with ellipsoids is obtained by parameterizing a suitable class of function by an ellipsoid using Fourier transforms. In Section 3.5 we prove (an extension of) a deep improvement of the Ajtai, Komlós, Tusnády theorem due to P. Shor. To prove this result, the Ellipsoid Theorem is no longer sufficient. The arguments we use instead are not fully satisfactory, and the best conceivable matching theorem, that would encompass all the results of this chapter, and much more, remains as a challenging problem, "the ultimate matching conjecture". With the exception of Section 3.1, the results of Chapter 3 are not connected to any subsequent result of the book.

In Chapter 4 we investigate Bernoulli processes, where the individual random variables X_t are linear combination of independent random signs. Random signs are obviously important r.v., 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. The Bernoulli process has better tails than the corresponding Gaussian process (it is "subgaussian") and is bounded whenever the 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 reasons why a Bernoulli process can be bounded, in the sense that a suitable "mixture" of them occurs in every bounded Bernoulli process. This is the yet unsolved "Bernoulli Conjecture". In Chapter 4 we develop tools to study Bernoulli processes, and give partial positive results in the direction of the Bernoulli Conjecture. These partial results fall far short of solving this conjecture, but are sufficient to obtain striking applications to infinitely divisible processes in Chapter 5 and to Banach spaces in Chapter 6.

Up to this point we have studied special processes: Gaussian, p-stable, Gaussian chaos, Bernoulli. These share the property that they are built on r.v. that have tails which can be well described with one or two parameters, such as in (0.4). More often a good description of the tail of a r.v. requires an entire sequence of parameters. Chapter 5 studies certain processes based on such r.v. In this case, the natural underlying structure is not a metric space, but a space equipped with a suitable family of distances. Once one has survived the initial surprise of this new idea, it is very pleasant to realize that the tools of Section 1.3 can be extended to this setting. This is the purpose of Section 5.1. In Section 5.2 we apply these tools to the situation of "canonical process" where the r.v. X_t are linear combinations of independent copies of symmetric r.v. with density proportional to $\exp(-|x|^\alpha)$ where $\alpha \geq 1$. The material of this section is independent of the rest of Chapter 5, which is devoted to infinitely divisible processes. These processes are studied in a much more general setting than what mainstream probability theory has yet investigated. (There is no assumption of stationarity of increments of any kind; the processes are actually indexed by an abstract set.) The main tool there is the Rosinski representation, that makes infinitely divisible processes appear as conditionally Bernoulli processes. (Unfortunately they do not seem to be conditionally Gaussian.) By using the tools of Chapter 4, for a large class of these processes, we are able to prove lower bounds that extend those given in Section 2.3 for p-stable process, and perhaps more importantly, to prove a general decomposition theorem showing that each bounded process in this class naturally decomposes into two parts, each of which is bounded for a rather obvious reason. We then give a sharp version of a "bracketing theorem", in the spirit of Ossiander's theorem of Chapter 2, a practical method to control infinitely divisible processes in great generality.

Chapter 6 gives applications to Banach space theory. The sections of this Chapter are largely independent of each other, and the link between them is mostly that they all reflect past interests of the author. The results of this chapter do not use those of Chapter 5. In Section 6.1, we study the cotype of operators from ℓ_N^∞ into a Banach space. In Section 6.2, we prove a new 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 6.3 we study the norm of the restriction of an operator from ℓ_N^q to the subspace generated by a randomly chosen small proportion of the coordinate vectors, and in Section 6.4 we use these results to obtain a sharpened version of the celebrated results of J. Bourgain on the Λ_p problem. A pretty recent theorem of G. Schechtman concludes this chapter in Section 6.5.

1.2 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 (0.4). We want to find bounds for $\mathsf{E}\sup_{t \in T} X_t$ depending on the structure of the metric space (T, d). We will *always* assume that

$$\forall t \in T,\ \mathsf{E} X_t = 0 \ . \tag{1.1}$$

Thus, given any t_0 in T, we have

$$\mathsf{E}\sup_{t \in T} X_t = \mathsf{E}\sup_{t \in T}(X_t - X_{t_0}) \ . \tag{1.2}$$

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

$$\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y > u)\, \mathrm{d}u \ . \tag{1.3}$$

Thus we look for bounds of

$$\mathsf{P}\Big(\sup_{t \in T}(X_t - X_{t_0}) \ge u\Big) \ . \tag{1.4}$$

We will assume that T is finite, which, as explained, does not decrease generality. The first bound that comes to mind is

$$\mathsf{P}\big(\sup_{t\in T}(X_t - X_{t_0}) \ge u\big) \le \sum_{t\in T} \mathsf{P}(X_t - X_{t_0} \ge u)\,. \tag{1.5}$$

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

$$X_t - X_{t_0} = X_t - X_{\pi_1(t)} + X_{\pi_1(t)} - X_{t_0}\,. \tag{1.6}$$

The idea is that it will be effective to use (1.5) on the variables $X_{\pi_1(t)} - X_{t_0}$, because there are not too many of them, and they are rather different. 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 $n \ge 0$, we consider a subset T_n of T, and for $t \in T$ we consider $\pi_n(t)$ in T_n. (The idea is of course that the points $\pi_n(t)$ are successive approximations of t.) We assume that T_0 consists of a single element t_0, so that $\pi_0(t) = t_0$ for each t in T. The fundamental relation is

$$X_t - X_{t_0} = \sum_{n\ge 1}\big(X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\big)\,, \tag{1.7}$$

that holds provided we arrange that $\pi_n(t) = t$ for n large enough, in which case the series is actually a finite sum. Relation (1.7) decomposes the increments of the process $X_t - X_{t_0}$ along the "chain" $(\pi_n(t))_{n\ge 0}$.

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

$$\operatorname{card} T_n \le N_n \tag{1.8}$$

where

$$N_0 = 1\,;\ N_n = 2^{2^n} \ \text{ if } \ n \ge 1\,. \tag{1.9}$$

The notation (1.9) will be used throughout the book.

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

$$d(t, \pi_n(t)) = d(t, T_n) = \inf_{s\in T_n} d(t,s)\,. \tag{1.10}$$

Using (0.4) we get that for $u > 0$ we have

$$\mathsf{P}\big(|X_{\pi_n(t)} - X_{\pi_{n-1}}(t)| \ge u2^{n/2} d(\pi_n(t), \pi_{n-1}(t)\big) \le 2\exp(-u^2 2^n)\,.$$

The number of possible pairs $(\pi_n(t), \pi_{n-1}(t))$ is bounded by $\operatorname{card} T_n \cdot \operatorname{card} T_{n-1} \le N_n N_{n-1} \le N_{n+1} = 2^{2^{n+1}}$. Thus, if we denote by Ω_u the event defined by

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

we see that

$$\mathsf{P}(\Omega_u^c) \leq p(u) := \sum_{n \geq 1} 2 \cdot 2^{2^{n+1}} \exp(-u^2 2^n)\,. \tag{1.11}$$

When Ω_u occurs, we see from (1.7) that

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

so that we have

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

where

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

and thus we have

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

Writing $u^2 2^n \geq u^2/2 + u^2 2^{n-1} \geq u^2/2 + 2^{n+1}$ for $u \geq 2$, we see that for $u \geq 2$ we have $p(u) \leq L \exp(-u^2/2)$. Here, as well as in the entire book, L denotes a universal constant, not necessarily the same at each occurrence. Thus, using (1.3) and keeping in mind that the integrand is ≤ 1 we get

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

Using the triangle inequality and (1.3) we see that

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

so that $S \leq L \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} d(t, T_n)$, and we have proved that

$$\mathsf{E} \sup_{t \in T} X_t \leq L \sup_t \sum_{n \geq 0} 2^{n/2} d(t, T_n)\,. \tag{1.12}$$

Now, how do we construct the sets T_n? The traditional method chooses them so that

$$\sup_t d(t, T_n)$$

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

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

Thus we define

$$e_n(T) = \inf \sup_t d(t, T_n) \,, \tag{1.13}$$

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.) This definition is convenient for our purposes. It is unfortunately not consistent with the conventions of Operator Theory, that denotes by e_{2^n} what we denote by e_n.

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

$$\frac{\Delta(T)}{2} \le e_0(T) \le \Delta(T) \,. \tag{1.14}$$

Here and in the sequel, $\Delta(T)$ denotes the diameter of T,

$$\Delta(T) = \sup_{t_1, t_2 \in T} d(t_1, t_2) \,. \tag{1.15}$$

When there is need to make clear which distance we use in the definition of the diameter, we will write $\Delta(T, d)$ rather than $\Delta(T)$.

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, we see that from (1.12) we have proved the following.

Proposition 1.2.1. *(Dudley's entropy bound [7]) Under the increment condition (0.4), we have*

$$\mathsf{E} \sup_{t \in T} X_t \le L \sum_{n \ge 0} 2^{n/2} e_n(T) \,. \tag{1.16}$$

This bound was proved only when T is finite, but using (0.2) it also extends to the case where T is infinite, as is shown by the following easy fact.

Lemma 1.2.2. *If U is a subset of T, we have $e_n(U) \le 2e_n(T)$.*

Proof. Indeed, if $a > e_n(T)$, one can cover T by N_n balls for d of 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 of radius $2a$. $\square$

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

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

Thus

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

and

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

So we have

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

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

$$\sqrt{\log 2}\sum_{n\ge 0} 2^{n/2}(e_n(T)-e_{n+1}(T)) \le \int_0^{e_0(T)} \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\ . \tag{1.17}$$

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 (1.17) yields

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

and hence Dudley's bound in the familiar form

$$\mathsf{E}\sup_{t\in T} X_t \le L\int_0^\infty \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\ . \tag{1.18}$$

Of course, since $\log 1 = 0$, the integral is in fact over $1 \le \epsilon \le \Delta(T)$.

We leave as an exercise the proof of the fact 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 (1.16) is not an improvement over (1.18).

We can however notice that the bound (1.12) seems genuinely better than the bound (1.16) because when going from (1.12) to (1.16) we have used the 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 bound (1.12) is the central idea of this work. Of course the fact that it appears now so naturally does not reflect the history of the subject, but rather that the proper approach is being used. When using this bound, we will choose the sets T_n in order to minimize the right-hand side of (1.12) instead of choosing them as in (1.13).

While at first one might think that (1.12) is not much of an improvement over (1.16), its importance arises from the fact that, as will be demonstrated later, in many cases it is essentially the best possible bound for $\mathsf{E}\sup_{t\in T} X_t$.

It turns out that the idea behind the bound (1.12) admits a more convenient formulation.

Definition 1.2.3. *Given a set T an* admissible sequence *is an increasing sequence $(\mathcal{A}_n)$ of partitions of T such that* $\operatorname{card}\mathcal{A}_n \le N_n$.

By increasing sequence of partitions we mean that every set of $\mathcal{A}_{n+1}$ is contained in a set of $\mathcal{A}_n$. Throughout the book we denote by $A_n(t)$ the unique element of $\mathcal{A}_n$ that contains t.

Theorem 1.2.4. *(The generic chaining bound). Under the increment condition (0.4) (and if $\mathsf{E}X_t = 0$) for each admissible sequence we have*

$$\mathsf{E}\sup_{t\in T} X_t \le L \sup_{t\in T} \sum_{n\ge 0} 2^{n/2} \Delta(A_n(t)) \ . \tag{1.19}$$

Here of course, as always, $\Delta(A_n(t))$ is the diameter of $A_n(t)$.

Proof. We can assume T finite. We construct a subset T_n of T by taking exactly one point in each set A of $\mathcal{A}_n$. We define $\pi_n(t)$ by

$$T_n \cap A_n(t) = \{\pi_n(t)\} \ .$$

Then, since $t\,,\ \pi_n(t) \in A_n(t)$ for $n \ge 0$, we have $d(t, \pi_n(t)) \le \Delta(A_n(t))$ and the result follows from (1.12). □

Definition 1.2.5. *Given $\alpha > 0$, and a metric space (T, d) (that need not be finite) we define*

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

where the infimum is taken over all admissible sequences.

It is good to observe that since $A_0(t) = T$ we have $\gamma_\alpha(T, d) \ge \Delta(T)$. An immediate consequence of Theorem 1.2.4 is as follows.

Theorem 1.2.6. *Under (0.4) and (1.1) we have*

$$\mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T, d) \ . \tag{1.20}$$

Of course to make this of interest we must learn how to control $\gamma_2(T, d)$, i.e. we must learn how to construct admissible sequences, a topic that we will first address in Section 1.3.

The following theorem applies to processes that satisfy a weaker bound than (0.4). It will be used many times.

Theorem 1.2.7. *Consider a set T provided with two distances d_1 and d_2. Consider a process $(X_t)_{t\in T}$ that satisfies $\mathsf{E}X_t = 0$ and*

$$\forall s,t \in T\ ,\ \forall u > 0\,, \tag{1.21}$$
$$\mathsf{P}(|X_s - X_t| \ge u) \le 2\exp\Big(-\min\Big(\tfrac{u^2}{d_2(s,t)^2}, \tfrac{u}{d_1(s,t)}\Big)\Big)\,.$$

Then

$$\mathsf{E}\sup_{s,t\in T} |X_s - X_t| \le L(\gamma_1(T,d_1) + \gamma_2(T,d_2))\ . \tag{1.22}$$

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

$$\forall t \in T\,,\ \sum_{n\ge 0} 2^n \Delta_1(B_n(t)) \le 2\gamma_1(T,d_1) \tag{1.23}$$

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

$$\forall t \in T\,,\ \sum_{n\ge 0} 2^{n/2} \Delta_2(C_n(t)) \le 2\gamma_2(T,d_2)\ . \tag{1.24}$$

Of course here $B_n(t)$ is the unique element of $\mathcal{B}_n$ that contains t (etc.). We define partitions $\mathcal{A}_n$ of T as follows. We set $\mathcal{A}_0 = \{T\}$, and, for $n \ge 1$, we define $\mathcal{A}_n$ as the partition generated by $\mathcal{B}_{n-1}$ and $\mathcal{C}_{n-1}$, that is 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. Let us define $\pi_n(t)$ as in the proof of Theorem 1.2.4. From (1.21) we see that, given $u \ge 1$, we have

$$\begin{aligned} \mathsf{P}\Big(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| &\ge u\big(2^n d_1(\pi_n(t),\pi_{n-1}(t)) + 2^{n/2} d_2(\pi_n(t),\pi_{n-1}(t))\big)\Big) \\ &\le 2\exp(-u2^n)\,, \end{aligned} \tag{1.25}$$

so that, proceeding as in (1.11), with probability $\ge 1 - L\exp(-u)$ we have

$$\begin{aligned} \forall n\ ,\ \forall t\ ,\ |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \le u\big(&2^n d_1(\pi_n(t),\pi_{n-1}(t)) \\ &+ 2^{n/2} d_2(\pi_n(t),\pi_{n-1}(t))\big)\,, \end{aligned}$$

and thus

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

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

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

and hence, since $d_1(\pi_1(t), \pi_0(t)) \le \Delta_1(B_0(t)) = \Delta_1(T)$,

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

Proceeding similarly for d_2 gives that

$$\mathsf{P}\Big(\sup_{t\in T} |X_t - X_{t_0}| \ge Lu(\gamma_1(T,d_1) + \gamma_2(T,d_2))\Big) \le L\exp(-u)$$

and (1.3) finishes the proof, using that

$$|X_s - X_t| \le |X_s - X_{t_0}| + |X_t - X_{t_0}| \;.$$

□

The reader has certainly noted that the left-hand sides of (1.20) and (1.22) are different. He might also have noticed that our proof of (1.20) gives in fact the apparently stronger result that

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \le L\gamma_2(t,d) \;. \tag{1.26}$$

This inequality is in fact not much stronger than (1.20). 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 will consider will be symmetric.

Lemma 1.2.8. *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 expectation. □

In this book, we 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 will at times need the following more precise version of Theorem 1.2.7. This more specialized version could be skipped at first reading.

Theorem 1.2.9. *Under the conditions of Theorem 1.2.7, for all values $u_1, u_2 > 0$ we have*

$$\begin{aligned} \mathsf{P}\Big(\sup_{t\in T} |X_t - X_{t_0}| &\ge L(\gamma_1(T,d_1) + \gamma_2(T,d_2)) + u_1 D_1 + u_2 D_2\Big) \\ &\le L\exp(-\min(u_2^2, u_1)) \;, \end{aligned} \tag{1.27}$$

where $D_j = 2\sum_{n\ge 0} e_n(T, d_j)$.

This is better than Theorem 1.2.7 because $D_j \le L\gamma_j(T, d_j)$.

Proof. There exists a partition of T into N_n sets, each of which having a diameter $\le 2e_n(T, d_1)$ for d_1. Thus there is an admissible sequence $(\mathcal{B}'_n)$ such that

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

and an admissible sequence $(\mathcal{C}'_n)$ that has the same property for d_2. We define $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$, and for $n \ge 2$ we define $\mathcal{A}_n$ as being the partition generated by $\mathcal{B}_{n-2}$, $\mathcal{B}'_{n-2}$, $\mathcal{C}_{n-2}$ and $\mathcal{C}'_{n-2}$, where $\mathcal{B}_n$ and $\mathcal{C}_n$ are as in (1.23) and (1.24) respectively.

Instead of (1.25) we use that for

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

we have

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

so that, with probability at least $1 - L\exp(-\min(u_2^2, u_1))$ we have

$$\begin{aligned}\forall n \ge 3\,,\ \forall t \in T\,,\ |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \le{}& 2^n\Delta_1(B_{n-3}(t)) + 2^{n/2}\Delta_2(C_{n-3}(t)) \\ &+ 2u_1e_{n-3}(T, d_1) + 2u_2e_{n-3}(T, d_2)\ .\end{aligned}$$

This inequality remains true for $n = 1, 2$ if in the right-hand side one replaces $n - 3$ by 0. □

1.3 A Partitioning Scheme

To make Theorem 1.2.4 useful, we must be able to construct good admissible sequences. In this section we explain our basic method. This method, and its variations, are at the core of the book.

We will say that a map F is a *functional* on a set T if, to each subset A of T it associates a number $F(A) \ge 0$, and if it is increasing, i.e.

$$A \subset A' \subset T \Rightarrow F(A) \le F(A')\,. \tag{1.28}$$

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

Consider a metric space (T, d) (that need *not* be finite), and a decreasing sequence $(F_n)_{n\ge0}$ of functionals on T, that is

$$\forall A \subset T\,,\ F_{n+1}(A) \le F_n(A)\ . \tag{1.29}$$

The basic property of these functionals is (somewhat imprecisely) that if we consider a set that is the union of many small pieces well separated from each other, then this set is significantly larger (as measured by the functionals) than the *smallest* of its pieces. "Significantly larger" depends on the scale of the pieces, and on their number through a function

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

The condition we are about to state involves two parameters of secondary importance, β and τ. At first reading one should assume $\beta = 1$ and $\tau = 1$.

Definition 1.3.1. *We say that the functionals F_n satisfy the* growth condition *if for a certain integer $\tau \geq 1$, and for certain numbers $r \geq 4$ and $\beta > 0$, the following holds true. Consider any integer $n \geq 0$, and set $m = N_{n+\tau}$. Then for any $s \in T$, any $a > 0$, any $t_1, \ldots, t_m$ such that*

$$\forall \ell \leq m\,,\ t_\ell \in B(s, ar)\ ;\ \forall \ell\,,\ \ell' \leq m\,,\ \ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a\,, \tag{1.30}$$

and any sets $H_1, \ldots, H_m \subset T$, we have

$$\begin{aligned} &\forall\ \ell \leq m\,,\ H_\ell \subset B(t_\ell, a/r) \\ &\Rightarrow F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq a^\beta \theta(n+1) + \min_{\ell \leq m} F_{n+1}(H_\ell)\ . \end{aligned} \tag{1.31}$$

Of course here $B(s, a)$ denotes the ball with center s and radius a in the metric space (T, d). A crucial fact in Condition (1.31) is that $H_\ell \subset B(t_\ell, a/r)$, while the points t_ℓ are at distance a from each other. The sets (H_ℓ) are "well separated". Only for such families of sets do we need to have some control of the functionals F_n. The role of the parameter r is to control how well these sets are separated. (The separation is better for larger r.) In the right-hand side of (1.31), the term $a^\beta \theta(n+1)$ is made up of the part a^β that account for the scale at which the sets H_ℓ are separated, and of the term $\theta(n+1)$ that accounts for the number of these sets. The "linear case" $\beta = 1$ is by far the most important. The role of the parameter τ is to give us some room. When τ is large, there are more sets and it should be easier to prove (1.31).

The first concrete example of the growth property occurs in this book is the (fundamental) case of Gaussian processes, where the functionals F_n do not depend on n and are given by $F_n(A) = \mathsf{E} \sup_{t \in A} X_t$, and the growth property for these functionals is proved in Proposition 2.1.4.

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

$$\xi \theta(n) \leq \theta(n+1) \leq \frac{r^\beta}{2} \theta(n)\ . \tag{1.32}$$

The most important example is $\theta(n) = 2^{n/2}$, $\beta = 1$, in which case (1.32) holds for $\xi = \sqrt{2}$.

It should be obvious that condition (1.31) imposes strong restrictions on the metric space (T,d). For example, when $\beta = 1, \tau = 1$ and $\theta(n) = 2^{n/2}$, taking $H_\ell = \{t_\ell\}$, and since $F_{n+1} \geq 0$, we get $F_0(T) \geq F_n(T) \geq a2^{(n+1)/2}$. Consider points $t_1, \cdots, t_k$ in $B(s, ar)$ such that $d(t_\ell, t_{\ell'}) \geq a$ whenever $\ell \neq \ell'$. If k is as large as possible, then $B(s, ar)$ is covered by the balls $B(t_\ell, a)$ so that whenever $a2^{(n+1)/2} > F_0(T)$, the ball $B(s, ar)$ can be covered by N_{n+1} balls $B(t,a)$. If $r = 4$, as will often be the case, it is then a simple matter to show that $2^{n/2} e_n(T) \leq L F_0(T)$. There is however a slight gap between this lower bound and the upper bound given by (1.16). This crude argument will be improved in the following theorem, and the resulting information provides a lower bound that is exactly of the order of the upper bound of (1.19).

Theorem 1.3.2. *Under the preceeding conditions we can find an increasing sequence $(\mathcal{A}_n)$ of partitions of T with* $\operatorname{card} \mathcal{A}_n \leq N_{n+\tau}$ *such that*

$$\sup_{t \in T} \sum_{n \geq 0} \theta(n) \Delta^\beta(A_n(t)) \leq L(2r)^\beta \Big(\frac{F_0(T)}{\xi - 1} + \theta(0) \Delta^\beta(T) \Big) . \tag{1.33}$$

In all the situations we will consider, it will be true that $F_0(t_1, t_2) \geq \theta(0) d^\beta(t_1, t_2)$ for any points t_1 and t_2 of T. (Since $F_1(H) \geq 0$ for any set H, this condition is essentially weaker in spirit than (1.31) for $n = 0$.) Then $\theta(0)\Delta^\beta(T) \leq F_0(T)$.

Theorem 1.3.2 constructs partitions given the functionals F_n, but it does not say how to find these functionals. One must understand that there is no magic. Admissible sequences are not going to come out of thin air, but rather reflect the geometry of the space (T,d). Once this geometry is understood, it is usually possible to guess a good choice for the functionals F_n. Many examples will be given in subsequent chapters. It seems, at least to the author, that it is much easier to guess the functionals F_n rather than the partitions of Theorem 1.3.2. Besides, as Theorem 1.3.4 below shows, we really have no choice. Functionals with the growth property are intimately connected with admissible sequences of partitions.

The sequence $(\mathcal{A}_n)$ of Theorem 1.3.2 is not admissible because $\operatorname{card} \mathcal{A}_n$ is too large. To construct good admissible sequences we will combine Theorem 1.3.2 with the following lemma.

Lemma 1.3.3. *Consider $\alpha > 0$, an integer τ and an increasing sequence of partitions $(\mathcal{B}_n)_{n \geq 0}$ with* $\operatorname{card} \mathcal{B}_n \leq N_{n+\tau}$. *Let*

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

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

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

Of course here $K(\alpha)$ denotes a number depending on α only.

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

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

and, using the bound $\Delta(A_n(t)) \le \Delta(T)$, we have

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

□

Our next result makes the point in the most important case that increasing sequences of functionals satisfying a growth conditions are canonical objects. It can be easily extended to the generality of Theorem 1.3.2.

Theorem 1.3.4. (a) *Assume that on the metric space (T, d) there exists an increasing sequence of functionals $(F_n)_{n \ge 0}$ that satisfies the growth condition for a certain $r \ge 4$, for $\beta = 1$, $\tau = 1$ and $\theta(n) = c 2^{n/2}$, where $c > 0$. Then*

$$\gamma_2(T, d) \le \frac{Lr}{c}(F_0(T) + \Delta(T)) .$$

(b) *In any metric space (T, d) there exists an increasing sequence of functionals $(F_n)_{n \ge 0}$ with $F_0(T) = \gamma_2(T, d)$ that satisfies the growth condition for $r = 4$, $\beta = 1$, $\tau = 1$ and $\theta(n) = 2^{n/2-1}$.*

In this theorem, (b) is a kind of converse of (a), showing that sequences of functionals satisfying the growth condition are a canonical method to control $\gamma_2(T, d)$ from above.

Proof. We first observe that (a) is a straightforward consequence of Theorem 1.3.2 and Lemma 1.3.3.

To prove (b) we define

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

where the infimum is taken over all admissible sequences of partitions of A. Thus $F_0(T) = \gamma_2(T, d)$. To prove the growth condition (1.31), consider $m = N_{n+1}$ and consider points $(t_\ell)_{\ell \le m}$ of T, with $d(t_\ell, t_{\ell'}) \ge a$ if $\ell \ne \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/4)$, $H = \bigcup_{\ell \le m} H_\ell$ and $c < \min_{\ell \le m} F_{n+1}(H_\ell)$. Consider an admissible sequence $(\mathcal{A}_n)$ of H, and

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

so that, since the sets H_ℓ for $\ell \le m$ are disjoint, we have $\operatorname{card} I \le N_n$, and thus there exists $\ell \le m$ with $\ell \notin I$. Then for $t \in H_\ell$, we have $A_n(t) \not\subset H_\ell$, so that since $A_n(t) \subset H$, the set $A_n(t)$ must meet a ball $B(t_{\ell'}, a/4)$ for $\ell \neq \ell'$, and hence $\Delta(A_n(t)) \geq a/2$, so that

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

and hence

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

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

$$F_n(H) \geq a 2^{n/2-1} + c ,$$

which is (1.31). □

The proof of Theorem 1.3.2 is not really difficult, but it requires some tedious bookkeeping. Probably the reader should first understand Section 2.1 to get convinced that the power of Theorem 1.3.2 is worth the effort of proving it.

Proof of Theorem 1.3.2. The beauty of this proof is that (almost) the most obvious ("greedy") construction works, but that it requires some skill to prove that this is the case.

All the balls we will consider will have a radius of the type r^{-j} for j in $\mathbb{Z}$, and before going into the proof, we rewrite (1.30) and (1.31) in the case where $a = r^{-j-1}$ respectively as

$$\forall \ell \leq m \, , \, t_\ell \in B(s, r^{-j}) \; ; \; \forall \ell \, , \, \ell' \leq m \, , \, \ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq r^{-j-1} ,$$

and

$$\begin{aligned} &\forall \;\; \ell \leq m \, , \, H_\ell \subset B(t_\ell, r^{-j-2}) \\ &\Rightarrow F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq r^{-\beta(j+1)} \theta(n+1) + \min_{\ell \leq m} F_{n+1}(H_\ell) . \end{aligned}$$

We are going to construct the increasing sequence $(\mathcal{A}_n)$ of partitions by induction. Together with $C \in \mathcal{A}_n$, we will construct a point t_C of T, an integer $j(C)$ in $\mathbb{Z}$ and three numbers $b_i(C)$ for $i = 0 \, , \, 1 \, , \, 2$. We assume

$$C \subset B(t_C, r^{-j(C)}) \tag{1.34}$$

so that in particular $\Delta(C) \leq 2r^{-j(C)}$. We assume

$$F_n(C) \leq b_0(C) \tag{1.35}$$

$$\forall t \in C \, , \, F_n(C \cap B(t, r^{-j(C)-1})) \leq b_1(C) \tag{1.36}$$

$$\forall t \in C, F_n(C \cap B(t, r^{-j(C)-2})) \leq b_2(C) . \tag{1.37}$$

The idea here is that

$$a_1(C) = \sup_{t \in C} F_n(C \cap B(t, r^{-j(C)-1}))$$

can be significantly smaller than $a_0(C) = F_n(C)$, and that we need to keep track of this; but, for technical reasons, $a_0(C)$ and $a_1(C)$ are not convenient and $b_0(C), b_1(C)$ are a "regularized version" of these. The (technical) regularity conditions we assume are

$$b_1(C) \le b_0(C) \tag{1.38}$$

and

$$b_0(C) - r^{-\beta(j(C)+1)}\theta(n) \le b_2(C) \le b_0(C) + \epsilon_n \ , \tag{1.39}$$

where

$$\epsilon_n = 2^{-n} F_0(T) \ . \tag{1.40}$$

The all important relation will be as follows.

If $n \ge 0$, $A \in \mathcal{A}_{n+1}$, $C \in \mathcal{A}_n$, $A \subset C$, then

$$\begin{aligned} &\sum_{0 \le i \le 2} b_i(A) + (1 - \frac{1}{\xi}) r^{-\beta(j(A)+1)} \theta(n+1) \\ &\le \sum_{0 \le i \le 2} b_i(C) + \frac{1}{2}(1 - \frac{1}{\xi}) r^{-\beta(j(C)+1)} \theta(n) + \epsilon_{n+1} \ . \end{aligned} \tag{1.41}$$

As we will show below, summation of these relations over $n \ge 0$ implies (1.33).

To start the construction, we set

$$\mathcal{A}_0 = \{T\} \, , \, b_0(T) = b_1(T) = b_2(T) = F_0(T) \, ,$$

and we choose any point $t_T \in T$. We then take $j(T)$ the largest possible such that $T \subset B(t_T, r^{-j(T)})$.

Let us now assume that for a certain $n \ge 0$ we have already constructed $\mathcal{A}_n$ with card $\mathcal{A}_n \le N_{n+\tau}$. To construct $\mathcal{A}_{n+1}$ we will split each set of $\mathcal{A}_n$ in at most $N_{n+\tau}$ pieces (so that, since $N_{n+\tau}^2 \le N_{n+\tau+1}$, we will have card $\mathcal{A}_{n+1} \le N_{n+\tau+1}$). So, let us fix $C \in \mathcal{A}_n$, and let $j = j(C)$.

By induction over $1 \le \ell \le m = N_{n+\tau}$ we construct points $t_\ell \in C$ and sets $A_\ell \subset C$ as follows.

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

$$F_{n+1}(C \cap B(t_1, r^{-j-2})) \ge \sup_{t \in C} F_{n+1}(C \cap B(t, r^{-j-2})) - \epsilon_{n+1} \ . \tag{1.42}$$

We then set $A_1 = C \cap B(t_1, r^{-j-1})$. The idea is simply that "we almost take the largest possible piece of C". The reader notices that the radius of the balls in (1.42) is r^{-j-2} while it is r^{-j-1} in the definition of A_1. This is the one trick of the proof. A "large piece" of C is a piece of the type $A_1 = C \cap B(t_1, r^{-j-1})$ for which $F_{n+1}(C \cap B(t_1, r^{-j-2}))$ (rather than $F_{n+1}(A_1)$) is large. The rest of the proof simply consists in iterating this construction and carefully checking (1.41). It is of course unfortunate that this somewhat tedious argument arises so early in this book. But the reader should take heart. The investment we are about to make will pay handsome dividends.

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

$$F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) \ge \sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, r^{-j-2})) - \epsilon_{n+1} \; . \quad (1.43)$$

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

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

If D_{m-1} is empty, the construction is finished. Otherwise we set $A_m = D_{m-1}$, so that $A_1, \dots, A_m$ form a partition of C.

In this manner we have partitioned C in at most m pieces. Let A be one of these.

If $A = A_m$, we define $j(A) = j(= j(C))$, $t_A = t_C$,

$$\begin{aligned} b_0(A) &= b_0(C)\,,\; b_1(A) = b_1(C) \\ b_2(A) &= b_0(C) - r^{-\beta(j+1)}\theta(n+1) + \epsilon_{n+1} \; . \end{aligned}$$

It is obvious that A and $n+1$ in place of C and n satisfy the relations (1.34), (1.38) and (1.39). The relations (1.35) and (1.36) for A follow from the fact that similar relations holds for C rather than A, that $F_{n+1} \le F_n$, and that the functional F_{n+1} is increasing.

We prove (1.37) for A. Consider any point $t = t_m \in A_m$. By construction, for $1 \le l \le m$, we have $t_\ell \in D_{\ell-1}$, and thus if $\ell' < \ell$ we have $d(t_\ell, t_{\ell'}) \ge r^{-j-1}$. Hence by (1.31), used for $a = r^{-j-1}$ and $H_\ell = D_\ell \cap B(t_{\ell+1}, r^{-j-2})$ we have (since F_n is increasing),

$$F_n(C) \ge r^{-\beta(j+1)}\theta(n+1) + \min_{0 \le \ell \le m-1} \big(F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) \big) \; . \quad (1.44)$$

Now, by (1.43), since $t_m \in D_\ell$, we have

$$\begin{aligned} F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) &\ge F_{n+1}(D_\ell \cap B(t_m, r^{-j-2})) - \epsilon_{n+1} \\ &\ge F_{n+1}(A \cap B(t_m, r^{-j-2})) - \epsilon_{n+1} \end{aligned}$$

because $A \subset D_\ell$.

Since $F_n(C) \le b_0(C)$, (1.44) yields

$$b_0(C) \ge r^{-\beta(j+1)}\theta(n+1) - \epsilon_{n+1} + F_{n+1}(A \cap B(t_m, r^{-j-2}))$$

and this proves (1.37) by definition of $b_2(A)$.

To prove (1.41), we observe that by definition

$$\begin{aligned}
&\sum_{0\le i\le 2} b_i(A) + (1-\frac{1}{\xi})r^{-\beta(j+1)}\theta(n+1) \qquad (1.45)\\
&= 2b_0(C) + b_1(C) - \frac{1}{\xi}r^{-\beta(j+1)}\theta(n+1) + \epsilon_{n+1}\\
&\le 2b_0(C) + b_1(C) - r^{-\beta(j+1)}\theta(n) + \epsilon_{n+1}
\end{aligned}$$

using the regularity condition on $\theta(n)$ (1.32) in the last inequality. But by (1.39) we have

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

so that (1.45) implies (1.41).

We are finished with the case $A = A_m$ and we suppose now that $A = A_\ell$, $\ell < m$. We define $j(A) = j+1$ and $t_A = t_\ell$, so that

$$A = A_\ell \subset B(t_\ell, r^{-j-1}) = B(t_A, r^{-j(A)}) \;,$$

and we define

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

Relations (1.38) and (1.39) for A are obvious. To prove (1.35) for A, we write

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

using (1.36) for C. In a similar manner, we have, if $t\in A$,

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

and this proves (1.36) for A. Also, (1.37) for A follows from (1.35) for A since $b_2(A) = b_0(A)$.

To prove (1.41), we observe that

$$\sum_{0\le i\le 2} b_i(A) \le 2b_1(C) + b_2(C) \le \sum_{0\le i\le 2} b_i(C) \qquad (1.46)$$

since $b_1(C)\le b_0(C)$ by (1.38). We observe that, since $j(A) = j(C)+1$, and since $r^{-\beta}\theta(n+1)\le\theta(n)/2$ by (1.32), we have

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

and combining with (1.46) this proves (1.41).

We have completed the construction, and we turn to the proof of (1.33). By (1.41), for any t in T, any $n \geq 0$, we have, setting $j_n(t) = j(A_n(t))$

$$\sum_{0 \leq i \leq 2} b_i(A_{n+1}(t)) + (1 - \frac{1}{\xi}) r^{-\beta(j_{n+1}(t)+1)} \theta(n+1)$$
$$\leq \sum_{0 \leq i \leq 2} b_i(A_n(t)) + \frac{1}{2}(1 - \frac{1}{\xi}) r^{-\beta(j_n(t)+1)} \theta(n) + \epsilon_{n+1} \ .$$

Since $b_i(T) = F_0(T)$ and since $b_i(A) \geq 0$ by (1.35) to (1.37), summation of these relations for $0 \leq n \leq q$ implies

$$(1-\frac{1}{\xi}) \sum_{0 \leq n \leq q} r^{-\beta(j_{n+1}(t)+1)} \theta(n+1) \leq 4F_0(T) + \frac{1}{2}(1-\frac{1}{\xi}) \sum_{0 \leq n \leq q} r^{-\beta(j_n(t)+1)} \theta(n) \tag{1.47}$$

and thus

$$\frac{1}{2}(1 - \frac{1}{\xi}) \sum_{0 \leq n \leq q} r^{-\beta(j_n(t)+1)} \theta(n) \leq 4F_0(T) + (1 - \frac{1}{\xi}) r^{-\beta(j(T)+1)} \theta(0) \ .$$

By (1.34), we have $\Delta(A_n(t)) \leq 2r^{-j_n(t)}$, and by the choice of $j(T)$ we have $r^{-j(T)-1} \leq \Delta(T)$ so that, since $\xi \leq 2$

$$\sum_{n \geq 0} \theta(n) \Delta^\beta(A_n(t)) \leq \frac{L(2r)^\beta}{\xi - 1}(F_0(T) + \Delta^\beta(T)\theta(0)) \ . \tag{1.48}$$

□

Theorem 1.3.5. *Consider a metric space (T, d), an integer $\tau' \geq 0$ and for $n \geq 0$, consider subsets T_n of T with* card $T_0 = 1$ *and* card $T_n \leq N_{n+\tau'} = 2^{2^{n+\tau'}}$ *for $n \geq 1$. Consider numbers $\alpha > 0$, $S > 0$, and let*

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

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

It is good to observe that this allows one to control $\gamma_\alpha(U)$ using sets T_n that need not be subsets of U. When $U = T$, we have in particular that

$$\gamma_\alpha(T, d) \leq K(\alpha) \sup_{t \in T} \sum_{n \geq 0} 2^{n/\alpha} d(t, T_n) \, , \tag{1.49}$$

which shows that the bound (1.19) is as good as the bound (1.12), if one does not mind the possible loss of a constant factor. As the proof of Theorem 1.2.7 should indicate, the bound (1.19) is usually more convenient.

One way to interpret (1.49) is as follows. Consider the quantity

$$\gamma'_\alpha(T,d) = \inf \sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} d(t,T_n)\,,$$

where the infimum is over all choices of the sets T_n with $\operatorname{card} T_n \leq N_n$. It is shown in the proof of Theorem 1.2.4 that $\gamma'_\alpha(T,d) \leq \gamma_\alpha(T,d)$, and (1.49) implies that $\gamma_\alpha(T,d) \leq K(\alpha)\gamma'_\alpha(T,d)$.

We will give two proofs of Theorem 1.3.5. The first proof relies on Theorem 1.3.2. It should help the reader to penetrate further applications of this theorem in the same spirit. The second proof is a simple direct argument.

First proof of Theorem 1.3.5. We will use Theorem 1.3.2 with $r=4$, $\beta = 1$ and $\tau = \tau' + 1$. For $n \geq 0$ and a subset A of U we define

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

Consider $m = N_{n+\tau'+1}$, points $t_1, \dots, t_m$ of U such that

$$1 \leq \ell < \ell' \leq m \Rightarrow d(t_\ell, t_{\ell'}) \geq a\,,$$

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

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

Since $d(t_\ell, t_{\ell'}) \geq a$ for $\ell \neq \ell'$, the open balls $B(t_\ell, a/2)$ are disjoint. Since there are $N_{n+\tau'+1}$ of them, whereas $\operatorname{card} T_n \leq N_{n+\tau'}$, one of these balls does not meet T_n. Thus there is $\ell \leq m$ with $d(t_\ell, T_n) \geq a/2$. Since we have $u_\ell \in H_\ell \subset B(t_\ell, a/4)$, we have $d(u_\ell, T_n) \geq a/4$ and

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

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

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

and since ϵ is arbitrary, this proves that (1.31) holds with $\theta(n+1) = 2^{n/\alpha - 2}$. (Condition (1.32) holds only when $\alpha \geq 1$, which is the most interesting case. We leave to the reader to take care of the case $\alpha < 1$ by using a different value of r.) We have $F_0(U) \leq S$, and since $d(t,T_0) \leq S$ for $t \in U$, and $\operatorname{card} T_0 = 1$, we have $\Delta(U) \leq 2S$. To finish the proof one simply applies Theorem 1.3.2 and Lemma 1.3.3. $\square$

Second proof of Theorem 1.3.5. For simplicity we assume $\tau' = 1$. For $u \in T_n$, let

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

We have $U = \bigcup_{u \in T_n} V(u)$, so we can find a partition $\mathcal{C}_n$ of U, with $\operatorname{card} \mathcal{C}_n \le N_n$, and the property that

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

Consider C as above, the smallest integer $b > 1/\alpha + 1$, the set

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

and, for $0 \le k < bn$, the set

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

Thus $\Delta(C_k) \le 2^{-k+1} \Delta(U)$, and, if $k < bn$,

$$\forall t \in C_k\,,\ \Delta(C_k) \le 4d(t, T_n)$$

and hence

$$\forall k \le bn\,,\ \forall t \in C_k\,,\ \Delta(C_k) \le 4d(t, T_n) + 2^{-bn+1} \Delta(U)\ . \tag{1.50}$$

Consider the partition $\mathcal{B}_n$ consisting of the sets C_k for $C \in \mathcal{C}_n\,,\ 0 \le k \le bn$, so that $\operatorname{card} \mathcal{B}_n \le (bn + 1) N_n$. Consider the partition $\mathcal{A}_n$ generated by $\mathcal{B}_0\,, \ldots, \mathcal{B}_n$, so that the sequence $(\mathcal{A}_n)$ increases, and $\operatorname{card} \mathcal{A}_n \le N_{n+\tau}$, where τ depends on α only. From (1.50) we get that

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

and thus

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

Since $\Delta(U) \le 2S$, the conclusion follows from Lemma 1.3.3. □

It is good to observe the following simple facts.

Theorem 1.3.6. *(a) If U is a subset of T, then*

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

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

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

then

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

(c) We have

$$\gamma_\alpha(T, d) \le K(\alpha) \sup \gamma_\alpha(F, d)\ , \tag{1.51}$$

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

Proof. Part (a) is obvious. To prove (b) we consider sets $T_n \subset T$ with $\operatorname{card} T_n \leq N_n$ and $\sup_{t\in T}\sum_{n\geq 0} 2^{n/\alpha} d(t,T_n) \leq 2\gamma_\alpha(T,d)$, we observe that $\sup_{s\in U}\sum_{n\geq 0} 2^{n/\alpha} d'(s,f(T_n)) \leq 2A\gamma_\alpha(T,d)$, and we apply Theorem 1.3.5.

To prove (c) we essentially repeat the argument of Theorem 1.3.4. We define

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

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

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

where the supremum is over $G \subset A$ and G finite. We will use Theorem 1.3.2 with $\beta = 1$, $\theta(n+1) = 2^{n/\alpha - 1}$, $\tau = 1$, and $r = 4$. (As in Theorem 1.3.5 this works only for $\alpha \geq 1$, and the case $\alpha < 1$ requires a different choice of r.) To prove (1.31), consider $m = N_{n+1}$ and consider points $(t_\ell)_{\ell\leq m}$ of T, with $d(t_\ell, t_{\ell'}) \geq a$ if $\ell \neq \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/4)$ and $c < \min_{\ell\leq m} F_{n+1}(H_\ell)$. For $\ell \leq m$, consider finite sets $G_\ell \subset H_\ell$ with $\gamma_{\alpha,n+1}(G_\ell, d) > c$, and $G = \bigcup_{\ell\leq m} G_\ell$. Consider an admissible sequence $(\mathcal{A}_n)$ of G, and

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

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

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

and hence

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

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

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

and thus

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

which is (1.31). Finally, we have $F_0(T) = \sup \gamma_\alpha(G,d)$, where the supremum is over $G \subset T$, G finite, and since $\Delta(G) \leq \gamma_\alpha(G,d)$, we have that $\Delta(T) \leq F_0(T)$ and we conclude by Lemma 1.3.3 and Theorem 1.3.2.

□

There are many possible variations for the scheme of proof of Theorem 1.3.2. We end this section by such a version. The proof of this specialized result could be omitted at first reading.

There are natural situations, where, in order to be able to obtain inequality on the right-hand side of (1.31), we need to know that $H_\ell \subset B(t_\ell, \eta a)$ where η is very small. In order to apply Theorem 1.3.2, we have to take $r \geq 1/\eta$, which (when $\beta = 1$) produces a loss of a factor $1/\eta$. We will give a simple modification of Theorem 1.3.2 that produces only a loss of a factor $\log(1/\eta)$.

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

Theorem 1.3.7. *Assume that the hypothesis of Theorem 1.3.2 are modified as follows. Whenever $t_1, \ldots, t_m$ are as in (1.30), and whenever $H_\ell \subset B(t_\ell, a4^{-q})$, we have*

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq a2^{n/2} + \min_{\ell \leq m} F_{n+1}(H_\ell) .$$

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

$$\sup_{t \in T} \sum_{n \geq 0} 2^{n/2} \Delta(A_n(t)) \leq Lq(F_0(T) + \Delta(T)) .$$

Proof. We closely follow the proof of Theorem 1.3.2. Together with each set C in $\mathcal{A}_n$, we construct numbers $b_i(C) \geq 0$ for $0 \leq i \leq q$, such that if $\epsilon_n = 2^{-n} F_0(T)$, we have

$$\forall i\,,\ 1 \leq i \leq q\,,\ b_i(C) \leq b_0(C)$$

$$\epsilon_n + b_0(C) \geq b_q(C) \geq b_0(C) - 4^{-j(A)-1} 2^{n/2}$$

$$F_n(C) \leq b_0(C)$$

$$\forall i\,,\ 1 \leq i \leq q\,,\ \forall t \in C\,,\ F_n(C \cap B(t, 4^{-j(C)-i})) \leq b_i(C) .$$

We set $b_i(T) = F_0(T)$ for $0 \leq i \leq q$. In the construction by induction, we take $m = N_{n+1}$ and we replace (1.43) by

$$F_{n+1}(D_\ell \cap B(t_{\ell+1}, 4^{-j-q})) \geq \sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, 4^{-j-q})) - \epsilon_{n+1}$$

Consider one of the pieces A of the partition of C. If $A = A_m$, we set

$$\forall i\,,\ 1 \leq i < q\,,\ b_i(A) = b_i(C)$$

$$b_q(A) = b_0(A) - 4^{-j-1} 2^{(n+1)/2} .$$

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

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

Exactly as previously we show in both cases that

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

and we finish the proof in the same manner. □

1.4 Notes and Comments

Consider an Orlicz function, that is a convex function Ψ on $\mathbb{R}^+$ such that $\Psi(0) = 0$. For a r.v. X, the Orlicz norm $\|X\|_\Psi$ is defined by

$$\|X\|_\Psi = \inf\{c > 0\,;\ \mathsf{E}\Psi\Big(\frac{|X|}{c}\Big) \le 1\}\,.$$

A general problem is to control $\mathsf{E}\sup_{t\in T} X_t$ under the increment condition

$$\forall s,t\in T\ \|X_s - X_t\|_\Psi \le d(s,t)\,. \tag{1.52}$$

The case of condition (0.4) is essentially the same as the case where $\Psi(x) = \exp x^2 - 1$. The most important result on this problem is a generalization of Dudley's entropy bound due to Pisier [33] and Kono [15].

$$\mathsf{E}\sup_{t\in T} X_t \le L\int_0^{\Delta(T,d)} \Psi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon\,. \tag{1.53}$$

In the study of the general problem of controlling the supremum of a process under (1.52) it is useful to distinguish the "polynomial growth case" e.g. $\Psi(x) = x^p$ where $p \ge 1$ from the "exponential growth" case, e.g. $\Psi(x) = \exp x^\alpha - 1$ where $\alpha \ge 1$. The polynomial growth case is the most difficult, and in that case it is hard to go beyond (1.53). A serious attempt in that direction took place in [46], using the tool of majorizing measures. (What majorizing measures are will be briefly discussed in the Appendix.) It unfortunately seems that the idea of the generic chaining *fails* to work in that case. For example, when $\Psi(x) = x^p$, the natural generalization of (1.12) seems to be that if the sets T_n satisfy $\operatorname{card} T_n \le 2^n$ then

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

Unfortunately, this inequality does not hold true. Majorizing measures remain of potential interest in that case, even though they are typically very difficult to use. See [78] for an example of use.

On the other hand, the idea of the generic chaining is well adapted to the exponential growth case, that is far more important than the polynomial growth case. In this case generic chaining makes the use of majorizing measures *totally obsolete*. As is shown is the Appendix, it yields the same results, but a look at the existing literature will convince the reader that it does so in a technically simpler fashion. Of course, entropy bounds in the spirit of (1.16) are sufficient for many problems and remain useful.

Probabilists are often more interested in the continuity of processes than in their boundedness. From the theoretical point of view that we adopt here, boundedness is however the central problem. When it is understood, dealing with continuity becomes a much easier question. Anticipating on the results of Chapter 2 we state a typical result about continuity.

Theorem 1.4.1. *Consider a Gaussian process $(X_t)_{t\in T}$, where T is countable, and the distance (0.3). Then the following are equivalent:*

1) The map $t \mapsto X_t(\omega)$ is uniformly continuous on (T, d) with probability 1.

2) We have

$$\lim_{\epsilon\to 0} \mathsf{E} \sup_{d(s,t)\le\epsilon} |X_s - X_t| = 0\,.$$

3) There exists an admissible sequence of partitions of T such that

$$\lim_{k\to\infty} \sup_{t\in T} \sum_{n\ge k} 2^{n/2}\Delta(A_n(t)) = 0.$$

Of course, the assumption that T is countable can be removed when one knows what a separable process is.

We will not develop results about continuity in order to keep this volume attractively thin.

2 Gaussian Processes and Related Structures

2.1 Gaussian Processes and the Mysteries of Hilbert Space

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

$$d(s,t) = \big(\mathsf{E}(X_s - X_t)^2\big)^{1/2} \; . \tag{2.1}$$

Theorem 2.1.1. (The majorizing measure theorem) *For some universal constant L we have*

$$\frac{1}{L}\gamma_2(T,d) \le \mathsf{E} \sup_{t \in T} X_t \le L\gamma_2(T,d) \; . \tag{2.2}$$

The reason for the name is explained in the Appendix. We observe that the right-hand side inequality follows from Theorem 1.2.6. To prove the lower bound, using Theorem 1.3.6 (c) we will assume without loss of generality that T is finite and we will then use Theorem 1.3.2. We will use the functionals

$$F_n(A) = F(A) = \mathsf{E} \sup_{t \in A} X_t \; ,$$

so that F_n does not depend on n. On purpose we give a proof that relies on general principles, and lends itself to generalizations.

Lemma 2.1.2. (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

$$\mathsf{E} \sup_{p \le m} X_{t_p} \ge \frac{a}{L_1}\sqrt{\log m} \; . \tag{2.3}$$

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

A proof of Sudakov minoration can be found in [18], p. 83.

Lemma 2.1.3. *Consider a Gaussian process $(Z_t)_{t\in U}$, where U is finite and a number σ such that $\sigma \geq \sup_{t\in U}(\mathsf{E}Z_t^2)^{1/2}$. Then for $u > 0$ we have*

$$\mathsf{P}\Big(\Big|\sup_{t\in U} Z_t - \mathsf{E}\sup_{t\in U} Z_t\Big| \geq u\Big) \leq 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big) . \tag{2.4}$$

This is a very important property of Gaussian processes. It is a facet of the theory of concentration of measure, a leading idea of modern probability theory. The reader is referred to the (very nice) book [17] to learn about this.

Proposition 2.1.4. *Consider points $(t_\ell)_{\ell\leq m}$ of T. Assume that $d(t_\ell, t_{\ell'}) \geq a$ if $\ell \neq \ell'$. Consider $\sigma > 0$, and for $\ell \leq m$ a finite set $H_\ell \subset B(t_\ell, \sigma)$. Then if $H = \bigcup_{\ell\leq m} H_\ell$ we have*

$$\mathsf{E}\sup_{t\in H} X_t \geq \frac{a}{L_1}\sqrt{\log m} - L_2\sigma\sqrt{\log m} + \min_{\ell\leq m}\mathsf{E}\sup_{t\in H_\ell} X_t . \tag{2.5}$$

If $\sigma \leq a/(2L_1L_2)$, (2.5) implies

$$\mathsf{E}\sup_{t\in H} X_t \geq \frac{a}{2L_1}\sqrt{\log m} + \min_{\ell\leq m}\mathsf{E}\sup_{t\in H_\ell} X_t , \tag{2.6}$$

which can be seen as a generalization of (2.3).

Proof. We can and do assume $m \geq 2$. For $\ell \leq m$, 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 $Z_t = X_t - X_{t_\ell}$ so that for $t \in U$ we have $\mathsf{E}Z_t^2 \leq \sigma$ and, for $u \geq 0$, by (2.4), we have

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

Thus if $V = \max_{\ell\leq m}|Y_\ell - \mathsf{E}Y_\ell|$ we have

$$\mathsf{P}(V \geq u) \leq 2m\exp\Big(-\frac{u^2}{2\sigma^2}\Big) . \tag{2.7}$$

For any non-negative r.v. V we have $\mathsf{E}V = \int_0^\infty \mathsf{P}(V \geq v)\,\mathrm{d}v$, and a simple calculation using (2.7) gives that, since $m \geq 2$,

$$\mathsf{E}V \leq \int_0^\infty \min\Big(1, 2m\exp\Big(-\frac{u^2}{2\sigma^2}\Big)\Big)\mathrm{d}u \leq L_2\sigma\sqrt{\log m} .$$

Now, for each $\ell \leq m$,

$$Y_\ell \geq \mathsf{E}Y_\ell - V \geq \min_{\ell\leq m}\mathsf{E}Y_\ell - V ,$$

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 expectation and use (2.3). □

Proof of Theorem 2.1.1. We fix $r \geq 2L_1L_2$, and we take $\beta = 1, \tau = 1$. To prove the growth condition for the functionals F_n we simply observe that (2.6) implies that (1.31) holds for $\theta(n) = 2^{n/2}/L$. Using Theorem 1.3.2 and Lemma 1.3.3, it remains only to control the term $\Delta(T)$. But we have

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

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

There is a rather interesting feature in the proof of Theorem 2.1.1. The objective of this theorem is to understand $\mathsf{E} \sup_{t \in T} X_t$, and for this we use functionals that are based precisely on this quantity we try to understand. One can then expect that Theorem 2.1.1 in itself has little practical value when we are faced with a concrete situation. Its content can be interpreted as meaning that there is really no other way to bound a Gaussian process than to control the quantity $\gamma_2(T, d)$. But of course, to control this quantity in a specific situation, we must in some way gain understanding of the underlying geometry of this situation.

The following is a noteworthy consequence of Theorem 2.1.1.

Theorem 2.1.5. *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 (2.1) 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 (1.26) and the left-hand side of (2.2). □

Let us now turn to a simple (and classical) example that illustrates well the difference between (1.18) and (1.12). Consider an independent sequence $(g_i)_{i \geq 1}$ of standard Gaussian r.v. and for $i \geq 2$ set

$$X_i = \frac{g_i}{\sqrt{\log i}} \ . \tag{2.8}$$

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

$$\frac{1}{\sqrt{\log(\min(p,q))}} \leq d(p,q) \leq \frac{2}{\sqrt{\log(\min(p,q))}} \,. \tag{2.9}$$

If $T_n \subset T$ and $\operatorname{card} T_n = N_n$ for $1 \leq n < s$, there exists $p \leq N_n + 1$ with $p \notin T_n$, so that by (2.9) we have $d(p, T_n) \geq 2^{-n/2}/L$, and thus $e_n(T) \geq 2^{-n/2}/L$. Thus

$$\sum_n 2^{n/2} e_n(T) \geq \frac{s-1}{L} \,. \tag{2.10}$$

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

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

by (2.9) and since $p, N_s \geq N_m$. Hence we have

$$\sum_n 2^{n/2} d(p, T_n) \leq \sum_{n \leq m} L 2^{n/2} 2^{-m/2} \leq L \,. \tag{2.11}$$

Comparing (2.10) and (2.11) we see that the bound (1.18) is worse than the bound (1.12) by a factor about s. This example is in a sense extremal. It is simple to see that, when T is finite, the bound (1.18) cannot be worse than (1.12) by a factor more than about $\log\log \operatorname{card} T$.

It follows from (2.11) and (1.12) that $\mathsf{E} \sup_{i \geq 1} X_i < \infty$. A simpler proof of this fact is given in Proposition 2.1.7 below.

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 $\|t\| = \|t\|_2 = (\sum_{i \geq 1} t_i^2)^{1/2}$. To each t in ℓ^2 we associate a Gaussian r.v.

$$X_t = \sum_{i \geq 1} t_i g_i \tag{2.12}$$

(the series converges in ℓ^2). In this manner, for each subset T of ℓ^2 we can consider the Gaussian process $(X_t)_{t \in T}$. The distance induced on T by the process coincides with the distance of ℓ^2 since from (2.12) 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 this way. (At least when there is a countable subset T' of T that is dense in the space (T, d), which is the only case of importance for us. Indeed, it suffices to think of the r.v. Y_t of a Gaussian process as a point in $L^2(\Omega)$, where Ω is the underlying probability space, and to identify $L^2(\Omega)$, which is then separable, and ℓ^2 by choosing an orthonormal basis of $L^2(\Omega)$.)

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

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

Theorem 2.1.6. *For a subset T of ℓ^2, we have*

$$\gamma_2(\mathrm{conv}T) \le L\gamma_2(T) \ . \tag{2.13}$$

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

$$\gamma_2(T_1 + T_2) \le L\big(\gamma_2(T_1) + \gamma_2(T_2)\big) \ . \tag{2.14}$$

Proof. To prove (2.13) we use (2.2) and the fact that

$$\sup_{t \in \mathrm{conv}T} X_t = \sup_{t \in T} X_t \tag{2.15}$$

since $X_{a_1 t_1 + a_2 t_2} = a_1 X_{t_1} + a_2 X_{t_2}$. The proof of (2.14) is similar. $\square$

Here is a simple fact.

Proposition 2.1.7. *Consider a set $T = \{t_k \ ; \ k \ge 1\}$ where $\|t_k\| \le 1/\sqrt{\log(k+1)}$. Then $\mathsf{E}\sup_{t\in T} X_t \le L$.*

Proof. We could use (1.12), but it is easier to write

$$\begin{aligned}\mathsf{P}\big(\sup_k |X_{t_k}| \ge u\big) &\le \sum_k \mathsf{P}(|X_{t_k}| \ge u) \\ &\le \sum_k 2\exp\Big(-\frac{u^2}{2}\log(k+1)\Big)\end{aligned} \tag{2.16}$$

since X_{t_k} is Gaussian with $\mathsf{E}X_{t_k}^2 \le 1/\log(k+1)$. Now for $u \ge 2$, the right-hand side of (2.16) is at most $L\exp(-u^2/2)$. $\square$

Combining with (2.15), this shows that $\mathsf{E}\sup_{t\in T} X_t \le L$, where $T = \mathrm{conv}\{t_k, k \ge 1\}$. The following means that this construction is generic.

Theorem 2.1.8. *Consider a countable set $T \subset \ell^2$, with $0 \in T$. Then we can find a sequence (t_k) in ℓ^2, with $\|t_k\|\sqrt{\log(k+1)} \le L\mathsf{E}\sup_{t\in T} X_t$ and*

$$T \subset \mathrm{conv}(\{t_k : k \ge 1\} \cup \{0\}) \ .$$

Proof. By Theorem 2.1.1 we can find an admissible sequence $\mathcal{A}_n$ of T with

$$\forall t \in T \, , \ \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \le L\mathsf{E}\sup_{t \in T} X_t = S \ . \tag{2.17}$$

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 $(u_n)_{n\geq 1}$ with $u_0 = 0$ and we put u_n in T_n.) For $n \geq 1$ consider the set U_n that consists of all the points

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

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 $U = \bigcup_{k\geq 1} U_k$. We observe that U contains at most N_{n+2} elements of norm $\geq 2^{-n/2}$. If we enumerate $U = \{t_k, k = 1, \ldots\}$ where the sequence $\|t_k\|$ is non-increasing, then if $\|t_k\| \geq 2^{-n/2}$ we have $k \leq N_{n+2}$ and this implies that $\|t_k\| \leq L/\sqrt{\log(k+1)}$.

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

$$t = \sum_{1\leq n\leq m} \pi_n(t) - \pi_{n-1}(t) = \sum_{1\leq n\leq m} a_n(t)u_n(t)\ , \tag{2.18}$$

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

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

we see from (2.18) that

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

This concludes the proof. $\square$

The simple proof of Theorem 2.1.6 hides the fact that (2.13) is a near miraculous result. It does not provide any real understanding of what is going on. Here is a simple question.

Research problem 2.1.9. Given a subset T of the unit ball of ℓ^2, give a geometrical proof that $\gamma_2(\mathrm{conv}T) \leq L\sqrt{\log \mathrm{card}\, T}$.

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

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

Research problem 2.1.10. Consider $1 < p < \infty$ and q with $1/p + 1/q = 1$. Let $\alpha = \min(2, q)$. Is it true that for any subset T of ℓ^p one has $\gamma_\alpha(\mathrm{conv}T, d) \leq K(p)\gamma_\alpha(T, d)$ where d denotes the distance of ℓ^p and $K(p)$ depends on p only?

2.2 A First Look at Ellipsoids

We have illustrated the gap between Dudley's bound (1.18) and the sharper bound (1.12), using the example (2.8), but perhaps the reader deems this example artificial. In this section we will illustrate this gap again using ellipsoids in Hilbert space. It is hard to argue that ellipsoids are unnatural or unimportant.

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

$$\mathcal{E} = \Big\{ t \in \ell^2\,; \sum_{i\geq 1} \frac{t_i^2}{a_i^2} \leq 1 \Big\}\,. \tag{2.19}$$

Proposition 2.2.1. *We have*

$$\frac{1}{L}\Big(\sum_{i\geq 1} a_i^2\Big)^{1/2} \leq \mathsf{E} \sup_{t\in\mathcal{E}} X_t \leq \Big(\sum_{i\geq 1} a_i^2\Big)^{1/2}\,. \tag{2.20}$$

Proof. We observe that, by the Cauchy-Schwarz inequality we have

$$Y := \sup_{t\in T} X_t = \sup_{t\in\mathcal{E}} \sum_{i\geq 1} t_i g_i \leq \Big(\sum_{i\geq 1} a_i^2 g_i^2\Big)^{1/2}\,. \tag{2.21}$$

Taking $t_i = a_i^2 g_i / \sum_{j\geq 1} a_j^2 g_j^2$, we see 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 (2.20) follows from the Cauchy-Schwarz inequality. For the left-hand side, let $\sigma = \max_{i\geq 1} |a_i|$. Since $Y = \sup_{t\in\mathcal{E}} X_t \geq |a_i||g_i|$, we have $\sigma \leq L\mathsf{E}Y$. It follows from (2.4) that

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

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

It follows from Theorem 2.1.1 that

$$\gamma_2(\mathcal{E}) \leq L\Big(\sum_{i\geq 1} a_i^2\Big)^{1/2}\,. \tag{2.22}$$

This is a statement about geometry of ellipsoids. Its proof was rather indirect. We will later on 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 can rewrite (2.20) as

$$\frac{1}{L}\Bigl(\sum_{n\geq 0} 2^n a_{2^n}^2\Bigr)^{1/2} \leq \mathsf{E}\sup_{t\in\mathcal{E}} X_t \leq \Bigl(\sum_{n\geq 0} 2^n a_{2^n}^2\Bigr)^{1/2} . \tag{2.23}$$

Proposition 2.2.2. *We have*

$$\frac{1}{L}\sum_{n\geq 0} 2^{n/2} a_{2^n} \leq \sum_{n\geq 0} 2^{n/2} e_n(\mathcal{E}) \leq L\sum_{n\geq 0} 2^{n/2} a_{2^n} . \tag{2.24}$$

The bounds in (2.23) and (2.24) are distinctively different. Dudley's bound (1.16) 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 3.1.

The proof of (2.24) is based on ideas that are at least 50 years old. The left-hand side is the easier part (and also the most important for us). It follows from the next Lemma.

Lemma 2.2.3. *We have* $e_n(\mathcal{E}) \geq \frac{1}{2}a_{2^n}$.

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

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

It should be obvious (using "projection on the first 2^n coordinates") that $e_n(\mathcal{E}_n) \leq e_n(\mathcal{E})$.

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

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

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 if $2\epsilon < a_{2^n}$, we cannot have $\mathcal{E}_n \subset \bigcup_{t\in T}(\epsilon B + t)$. Thus $e_n(\mathcal{E}_n) \geq \epsilon$. □

We now turn to the upper bound.

Lemma 2.2.4. *We have*

$$e_{n+3}(\mathcal{E}) \leq 3\max_{k\leq n} a_{2^k} 2^{k-n} . \tag{2.25}$$

Proof. We keep the notations of Lemma 2.2.3. First we show that

$$e_{n+3}(\mathcal{E}) \le e_{n+3}(\mathcal{E}_n) + a_{2^n} \ . \tag{2.26}$$

To see this, we observe that, if $t \in \mathcal{E}$, 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 $d(t, \mathcal{E}_n) \le a_{2^n}$. This proves (2.26).

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

$$\text{Any two points of } Z \text{ are at mutual distance} \ge 2\epsilon \tag{2.27}$$

$$\operatorname{card} Z \text{ is as large as possible under (2.27).} \tag{2.28}$$

Then by (2.28) the balls centered at points of Z and of radius $\le 2\epsilon$ cover $\mathcal{E}_n$. Thus

$$\operatorname{card} Z \le N_{n+3} \Rightarrow e_{n+3}(\mathcal{E}_n) \le 2\epsilon \ . \tag{2.29}$$

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

$$\operatorname{card} Z \ \mathrm{Vol}(\epsilon B) \le \mathrm{Vol}(\mathcal{E}_n + \epsilon B) \ . \tag{2.30}$$

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 ϵB, we have $\sum_{i\le 2^n} t_i'^2/\epsilon^2 \le 1$. Since $(t_i + t_i')^2 \le 2t_i^2 + 2t_i'^2$, we have

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

where $c_i = 2\max(\epsilon, a_i)$, so that

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

and comparing with (2.30) we have

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

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

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

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

In summary, 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 (2.26). □

Proof of (2.24). We have, using (2.25)

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

Since $e_n(\mathcal{E}) \leq a_1$, the result follows. □

2.3 *p*-stable Processes

Consider a number $0 < p \leq 2$. A r.v. X is called (real, symmetric) p-stable if for each $\lambda \in \mathbb{R}$ we have

$$\mathsf{E} \exp i\lambda X = \exp\Big(-\frac{\sigma^p |\lambda|^p}{2}\Big) , \tag{2.31}$$

where $\sigma = \sigma_p(X)$ is called the parameter of X. The name "p-stable" comes from the fact that if $X_1 , \ldots , X_m$ are independent and p-stable, then $\sum_{j\leq m} a_j X_j$ is p-stable, and

$$\sigma_p\Big(\sum_{j\leq m} a_j X_j\Big) = \Big(\sum_{j\leq m} |a_j|^p \sigma_p(X_j)^p\Big)^{1/p} \ .$$

This is obvious from (2.31).

The reason for the restriction $p \leq 2$ is that for $p > 2$ a r.v. as in (2.31) does not exist. The case $p = 2$ is the Gaussian case. Despite the formal similarity, the case $p < 2$ is very different. It can be shown that

$$\lim_{s\to\infty} s^p \mathsf{P}(|X| \geq s) = c_p \sigma^p \tag{2.32}$$

where $c_p > 0$ depends on p only. Thus X does not have moments of order p, but it has moments of order q for $q < p$.

A process $(X_t)_{t\in T}$ is called p-stable if for every family $(\alpha_t)_{t\in T}$ for which only finitely many of the numbers α_t are not 0 the r.v. $\sum_t \alpha_t X_t$ is p-stable. We can then define a (quasi) distance d on T by

$$d(s,t) = \sigma(X_s - X_t)\,. \tag{2.33}$$

One can also define an equivalent distance by $d(s,t) = (\mathsf{E}|X_s - X_t|^q)^{1/q}$, where $q < p$. In contrast with the Gaussian case, it seems unrealistic to hope to compute $\mathsf{E}\sup_{t\in T} X_t$ as a function of the geometry of (T,d). Yet, as a consequence of Theorem 2.1.1 one can extend the lower bound of Theorem 2.1.1 as follows.

Theorem 2.3.1. *For $1 < p < 2$, there is a number $K(p)$ such that for any p-stable process $(X_t)_{t\in T}$ we have*

$$\gamma_q(T,d) \le K(p)\mathsf{E}\sup_{t\in T} X_t\,,$$

where q is the conjugate exponent of p, i.e. $1/q + 1/p = 1$, and where d is as in (2.33).

At the heart of Theorem 2.3.1 is the fact that for $1 \le p < 2$ the process (X_t) can be represented as a conditionally Gaussian process. That is, we can find two probability spaces $(\Omega\,, \mathsf{P})\,, (\Omega'\,, \mathsf{P}')$ and a family $(Y_t)_{t\in T}$ of r.v. on $\Omega \times \Omega'$ (provided with the product probability), such that

$$\text{Given any finite subset } U \text{ of } T, \text{ the joint laws of } (Y_t)_{t\in U} \text{ and } (X_t)_{t\in U} \text{ are identical.} \tag{2.34}$$

$$\text{Given } \omega \in \Omega\,, \text{ the process } \omega' \mapsto Y_t(\omega,\omega') \text{ is a centered Gaussian process.} \tag{2.35}$$

We refer the reader to [18], Theorem 5.1 (with η Gaussian) for a proof of this and for general background on p-stable processes. A remarkable fact is that we do not need to know precisely how this representation arises.

We denote by E' integration in P' only. Given ω, we consider the random distance d_ω on T given by

$$d_\omega(s,t) = \left(\mathsf{E}'(Y_s(\omega,\omega') - Y_t(\omega,\omega'))^2\right)^{1/2}\,. \tag{2.36}$$

We consider $1 \le p < 2$ and we define α by

$$\frac{1}{\alpha} = \frac{1}{p} - \frac{1}{2}\,. \tag{2.37}$$

Lemma 2.3.2. *If $1 \le p < 2$, for all $s,t \in T$ and $\epsilon > 0$, we have*

$$\mathsf{P}(d_\omega(s,t) \le \epsilon d(s,t)) \le \exp\Bigl(-\frac{b_p}{\epsilon^\alpha}\Bigr) \tag{2.38}$$

where $b_p > 0$ depends on p only.

Proof. Since the process $Y_t(\omega, \cdot)$ is Gaussian, we have

$$\mathsf{E}' \exp i\lambda(Y_s - Y_t) = \exp\Big(-\frac{\lambda^2}{2} d_\omega^2(s,t)\Big) .$$

Taking expectation, using (2.31), and the fact that the pair (Y_s, Y_t) has the same law as the pair (X_s, X_t), we get

$$\exp\Big(-\frac{|\lambda|^p}{2} d^p(s,t)\Big) = \mathsf{E} \exp\Big(-\frac{\lambda^2}{2} d_\omega^2(s,t)\Big) . \tag{2.39}$$

For a r.v. Z we have

$$\mathsf{P}(Z \le u) \le \exp\Big(\frac{\lambda^2 u}{2}\Big) \mathsf{E} \exp\Big(-\frac{\lambda^2}{2} Z\Big) .$$

Using this for $Z = d_\omega^2(s,t)$ and $u = \epsilon^2 d^2(s,t)$, we get, using (2.39),

$$\mathsf{P}(d_\omega(s,t) \le \epsilon d(s,t)) \le \exp\Big(\frac{1}{2}\big(\lambda^2 \epsilon^2 d^2(s,t) - |\lambda|^p d^p(s,t)\big)\Big) ,$$

and the result by optimization over λ. □

This lemma shows the relevance of the following.

Theorem 2.3.3. *Consider a (finite) metric space (T, d) and a random distance d_ω on T. Assume that for some $b > 0$ we have*

$$\forall s\,,\, t \in T\,,\, \forall \epsilon > 0\,,\, \mathsf{P}(d_\omega(s,t) \le \epsilon d(s,t)) \le \exp\Big(-\frac{b}{\epsilon^\alpha}\Big) , \tag{2.40}$$

where $\alpha > 2$. Then

$$\mathsf{P}\Big(\gamma_2(T, d_\omega) \ge \frac{1}{K}\gamma_q(T,d)\Big) \ge \frac{3}{4} , \tag{2.41}$$

where

$$\frac{1}{q} = \frac{1}{2} - \frac{1}{\alpha} ,$$

and where K depends on α and b only.

The number $3/4$ plays no special role.

Proof of Theorem 2.3.1. Using Theorem 1.3.6, we can assume that T is finite. Consider $Z = \sup_{t \in T} Y_t$. From Theorem 2.1.1 we have

$$\mathsf{E}' Z \ge \frac{1}{L} \gamma_2(T, d_\omega)$$

and since $\mathsf{E}' Z \ge 0$, taking expectation and using (2.38) and (2.41) proves that $\mathsf{E} Z \ge \gamma_q(T,d)/K(p)$. □

Proof of Theorem 2.3.3. Replacing d by $b^{1/\alpha}d$, we can and do assume that $b = 1$. To prove (2.41) we will prove that if $U \subset \Omega$ satisfies $\mathsf{P}(U) \geq 1/4$, then

$$\mathsf{E}(\mathbf{1}_U \gamma_2(T, d_\omega)) \geq \frac{1}{L}\gamma_q(T, d) \ . \tag{2.42}$$

Since $U = \{\gamma_2(T, d_\omega) < \gamma_q(T, d)/L\}$ violates (2.42), we must have $\mathsf{P}(U) < 1/4$, and this proves (2.41).

We fix U once and for all with $\mathsf{P}(U) \geq 1/4$. Given a probability measure μ on T and $n \geq 0$ we set

$$F_n(\mu) = \mathsf{E}\Big(\mathbf{1}_U \inf \int_T \sum_{k \geq n} 2^{k/2} \Delta(A_k(t), d_\omega) \mathrm{d}\mu(t)\Big)$$

where the infimum is taken over all admissible sequences $(\mathcal{A}_n)_{n \geq 0}$ of T. Given $A \subset T$, we set

$$F_n(A) = \sup F_n(\mu) \ ,$$

where the supremum is over all probability measures μ supported by A. Using that $\int f \mathrm{d}\mu \leq \sup f$, we see that

$$F_0(T) \leq \mathsf{E}(\mathbf{1}_U \gamma_2(T, d_\omega)) \ .$$

We claim that

$$\Delta(T, d) \leq K F_0(T) \ . \tag{2.43}$$

(Here and in the rest of the proof, K denotes a number depending on α only, that need not be the same at each occurrence.) To see this, we simply note that since $\mathcal{A}_0 = \{T\}$, we have $A_0(t) = T$ for each t, so that

$$\begin{aligned} F_0(T) &\geq \mathsf{E}(\mathbf{1}_U \Delta(T, d_\omega)) \\ &\geq \epsilon \Delta(T, d) \mathsf{P}(U \cap \{\Delta(T, d_\omega) \geq \epsilon \Delta(T, d)\}) \\ &\geq \frac{1}{K} \Delta(T, d) \end{aligned} \tag{2.44}$$

using (2.40) for ϵ small enough.

Thus (2.42), and hence Theorem 2.3.3 will follow from Theorem 1.3.2 (used for $r = 4$, $\beta = 1$, $\theta(n) = 2^{n/q}/K$, $\xi = 2^{1/q}$ and $\tau = 3$) and Lemma 1.3.3 provided we prove that the functionals satisfy the growth condition of Definition 1.2.5. The purpose of taking $\tau = 3$ is that it greatly helps to check this condition, although this will became apparent only at the end of our calculations. To prove the growth condition, we consider $n \geq 0$, $m = N_{n+3}$, and points $(t_\ell)_{\ell \leq m}$ in T, with

$$\ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq 4a > 0 \ . \tag{2.45}$$

We consider sets $H_\ell \subset B(t_\ell, a)$, and we want to show that

$$F_n\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge \frac{2^{n/q}a}{K} + \min_{\ell\le m} F_{n+1}(H_\ell)\ . \tag{2.46}$$

(Thus a does not have exactly the same meaning as in (1.31).) Consider $c < \min_{\ell\le m} F_{n+1}(H_\ell)$, and consider for each ℓ a probability μ_ℓ supported by H_ℓ, and such that $F_{n+1}(\mu_\ell) > c$. Consider

$$\mu = \frac{1}{m}\sum_{\ell\le m} \mu_\ell\ . \tag{2.47}$$

This is a probability, which is supported by $\bigcup_{\ell\le m} H_\ell$. To prove (2.46), it suffices to prove that

$$F_n(\mu) \ge \frac{2^{n/q}a}{K} + c\ . \tag{2.48}$$

Using that $\inf(f(x)+g(x)) \ge \inf f(x) + \inf g(x)$, we see that

$$F_n(\mu) \ge \mathrm{I} + \mathrm{II}$$

where

$$\mathrm{I} = F_{n+1}(\mu) = \mathsf{E}\Big(\mathbf{1}_U \inf \int \sum_{k\ge n+1} 2^{k/2}\Delta(A_k(t), d_\omega)\mathrm{d}\mu(t)\Big)$$

$$\mathrm{II} = \mathsf{E}\Big(\mathbf{1}_U \inf \int 2^{n/2}\Delta(A_n(t), d_\omega)\mathrm{d}\mu(t)\Big)\ ,$$

where both infimum are over all admissible sequences $(\mathcal{A}_n)$ of T. Using (2.47), we have

$$\mathrm{I} \ge \frac{1}{m}\sum_{\ell\le m} F_{n+1}(\mu_\ell) \ge c$$

so all what remains to prove is that

$$\mathrm{II} \ge \frac{2^{n/q}a}{K}\ . \tag{2.49}$$

Given a partition $\mathcal{A}_n$, let us define the random subset D of T by

$$D = \bigcup\big\{A \in \mathcal{A}_n\ ;\ \Delta(A, d_\omega) \le 2^{-(n+3)/\alpha}a\big\}\ .$$

For $t \notin D$, we have $\Delta(A_n(t), d_\omega) \ge 2^{-(n+3)/\alpha}a$, and thus, since $1/2 - 1/\alpha = 1/q$,

$$\int 2^{n/2}\Delta(A_n(t), d_\omega)\mathrm{d}\mu(t) \ge \frac{2^{n/q}}{K} a\mu(T\backslash D)\ .$$

Now

$$\mathsf{E}(\mathbf{1}_U \inf \mu(T\backslash D)) = \mathsf{E}(\mathbf{1}_U(1 - \sup \mu(D))) \ge \frac{1}{4} - \mathsf{E}\sup \mu(D)\ , \tag{2.50}$$

where as usual the infimum and the supremum are over the choice of $(\mathcal{A}_n)$. Thus all we now have to do is to prove that

$$\mathsf{E} \sup \mu(D) \le \frac{1}{8} \,. \tag{2.51}$$

Let us define

$$\mathcal{D} = \{A \in \mathcal{A}_n \; ; \; A \subset D\} = \big\{A \in \mathcal{A}_n \; ; \; \Delta(A, d_\omega) \le 2^{-(n+3)/\alpha} a\big\} \,.$$

Thus, by the Cauchy-Schwarz inequality, we have

$$\mu(D) = \sum_{A \in \mathcal{D}} \mu(A) \le \Big(\operatorname{card} \mathcal{D} \sum_{A \in \mathcal{D}} \mu^2(A)\Big)^{1/2} \,. \tag{2.52}$$

Now, by definition of $\mathcal{D}$, for $s, t \in A \in \mathcal{D}$ we have $d_\omega(s,t) \le 2^{-(n+3)/\alpha} a$, so that

$$A^2 \subset V_\omega := \big\{(s,t) \; ; \; d_\omega(s,t) \le 2^{-(n+3)/\alpha} a\big\} \tag{2.53}$$

and since the sets A^2 are disjoint, we have $\sum \mu^2(A) = \sum \mu^{\otimes 2}(A^2) \le \mu^{\otimes 2}(V_\omega)$ and (2.52) yields, since $\operatorname{card} \mathcal{D} \le \operatorname{card} \mathcal{A}_n \le N_n$,

$$\mu(D) \le \big(N_n \mu^{\otimes 2}(V_\omega)\big)^{1/2} \,. \tag{2.54}$$

Now, if $s \in H_\ell$, $t \in H_{\ell'}$, $\ell \ne \ell'$, we have $d(s,t) \ge 2a$, so that if $H = \bigcup_{\ell \le m} H_\ell$ we have

$$H^2 \cap V_\omega \subset W_\omega \cup \bigcup_{\ell \le m} H_\ell^2 \,, \tag{2.55}$$

where

$$W_\omega = \big\{(s,t) \in T^2 \; ; \; d_\omega(s,t) \le 2^{-(n+3)/\alpha} d(s,t)\big\} \,.$$

Since $\mu(H_\ell) = 1/m$ and $\mu(H) = 1$, this yields

$$\mu^{\otimes 2}(V_\omega) \le \frac{1}{m} + \mu^{\otimes 2}(W_\omega) \,.$$

This bound is independent of $\mathcal{A}_n$, so combining with (2.54) we get

$$\sup \mu(D) \le \Big(N_n \Big(\frac{1}{m} + \mu^{\otimes 2}(W_\omega)\Big)\Big)^{1/2}$$

and, by the Cauchy-Schwarz inequality,

$$\mathsf{E} \sup \mu(D) \le \Big(N_n \Big(\frac{1}{m} + \mathsf{E}\mu^{\otimes 2}(W_\omega)\Big)\Big)^{1/2} \,. \tag{2.56}$$

Now

$$\begin{aligned}\mathsf{E}\mu^{\otimes 2}(W_\omega) &= \mathsf{E}\int \mathbf{1}_{W_\omega}(s,t)\mathrm{d}\mu(s)\mathrm{d}\mu(t)\\ &= \int \mathsf{E}(\mathbf{1}_{W_\omega}(s,t))\mathrm{d}\mu(s)\mathrm{d}\mu(t)\\ &\le \exp(-2^{n+3})\,,\end{aligned}$$

using (2.40) in the last line. Thus

$$\mathsf{E}\sup \mu(D) \le \left(\frac{2N_n}{N_{n+3}}\right)^{1/2} \le \frac{1}{8}\,. \tag{2.57}$$

□

We now turn to the case $p = 1$. We set $M_0 = 1$, $M_n = 2^{N_n}$ for $n \ge 1$. Given a metric space (T, d) we define

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

where the infimum is taken over all increasing families of partitions $(\mathcal{B}_n)$ of T with $\operatorname{card}\mathcal{B}_n \le M_n$.

Theorem 2.3.4. *Consider a finite metric space (T, d) and a random distance d_ω on T. Assume that*

$$\forall s\,,\ t\in T\,,\ \forall \epsilon > 0\,,\ P(d_\omega(s,t) < \epsilon d(s,t)) \le \exp\Bigl(-\frac{1}{\epsilon^2}\Bigr)\,.$$

Then

$$P\Bigl(\gamma_2(T,d_\omega) \ge \frac{1}{L}\gamma_\infty(T,d)\Bigr) \ge \frac{3}{4}\,.$$

Proof. The proof of Theorem 2.3.4 closely follows that of Theorem 2.3.3, so we indicate only the necessary modifications. It should be obvious that Theorem 1.3.2 holds when we replace N_n by M_n. We will use it for $\theta(n) = 2^n/L$, $r = 4$ and $\tau = 2$. We define

$$F(\mu) = \mathsf{E}\Bigl(\mathbf{1}_U \inf \int \sum_{k\ge 2^n-1} 2^{k/2}\Delta(A_k(t),d_\omega)d\mu(t)\Bigr)\,.$$

Here, and everywhere in this proof the infimum is over all admissible sequences $(\mathcal{A}_n)_{n\ge 0}$ of T. (Thus, as usual, $\operatorname{card}\mathcal{A}_n \le N_n$.) All we have to do is to prove that under the condition (2.45) (with now $m = M_{n+2}$) we have

$$\mathsf{E}\Bigl(\mathbf{1}_U \inf \int \sum_{2^n-1\le k<2^{n+1}-1} 2^{k/2}\Delta(A_k(t),d_\omega)\mathrm{d}\mu(t)\Bigr) \ge \frac{2^n}{L}a\,.$$

It suffices for this purpose to prove that for each $2^n - 1 \le k < 2^{n+1} - 1$, we have

$$\mathsf{E}\Big(\mathbf{1}_U \inf \int 2^{k/2} \Delta(A_k(t), d_\omega) \mathrm{d}\mu(t)\Big) \geq \frac{a}{L} \,,$$

This will follow from (2.51), where now

$$D = \bigcup \Big\{ A \in \mathcal{A}_k \,,\ \Delta(A, d_\omega) \leq \frac{2^{-k/2} a}{4} \Big\} \,.$$

To prove (2.51) we define

$$W_\omega = \big\{ (s,t) \in T^2 \,;\ d_\omega(s,t) \leq 2^{-k/2-2} d(s,t)) \big\} \,,$$

and copying the proof of (2.56) we obtain

$$\begin{aligned}
\mu(D) &\leq \Big(N_k \Big(\frac{1}{m} + \mathsf{E}\mu(W_\omega) \Big) \Big)^{1/2} \\
&\leq \Big(N_k \Big(\frac{1}{M_{n+2}} + \exp\big(-2^{k+4}\big) \Big) \Big)^{1/2} \,.
\end{aligned}$$

Since $k < 2^{n+1} - 1$, we have $k \leq 2^{n+2} - 3$, so $M_{n+2} \geq N_{k+3}$ and then $\mu(D) \leq (2N_k / N_{k+3})^{1/2} \leq 1/8$ as before. □

Theorem 2.3.5. *For every* 1*-stable process* $(X_t)_{t \in T}$ *we have*

$$\mathsf{P}\Big(\sup_{t \in T} (X_t - X_{t_0}) \geq \frac{1}{L} \gamma_\infty(T, d) \Big) \geq \frac{1}{L} \,.$$

To understand the formulation of this theorem, we note that we cannot use expectation to measure the size of $\sup_{t \in T} X_t$, as is shown by (2.32). Also, we observe that when T consists of 2 points t_0 and t_1, then

$$\sup_{t \in T} (X_t - X_{t_0}) = \max(X_{t_1} - X_{t_0}, 0)$$

is 0 with probability 1/2.

Lemma 2.3.6. *If* $(Y_t)_{t \in T}$ *is a Gaussian process then*

$$\mathsf{P}\Big(\sup_{t \in T} (Y_t - Y_{t_0}) \geq \frac{1}{2} \mathsf{E} \sup_{t \in T} (Y_t - Y_{t_0}) \Big) \geq \frac{1}{L} \,.$$

Proof. This is a consequence of two classical facts. Firstly, the r.v. $Z = \sup_{t \in T} (Y_t - Y_{t_0})$ satisfies $\mathsf{E}Z^2 \leq L(\mathsf{E}Z)^2$ (a weak consequence of (2.4)). Secondly a r.v. $Z \geq 0$ satisfies

$$\mathsf{P}\Big(Z \geq \frac{\mathsf{E}Z}{2} \Big) \geq \frac{1}{4} \frac{(\mathsf{E}Z)^2}{\mathsf{E}Z^2} \,, \tag{2.58}$$

the "second moment method". □

Remark 2.3.7. Since $\mathsf{E}Z^2 \le L(\mathsf{E}Z)^2$, (2.58) shows that, assuming $Y_{t_0} = 0$ for some $t_0 \in T$

$$\mathsf{P}\Big(\sup_{t \in T} Y_t \ge \frac{1}{L}\big(\mathsf{E}(\sup_{t\in T} Y_t)^2\big)^{1/2}\Big) \ge \frac{1}{L}\,. \tag{2.59}$$

Proof of Theorem 2.3.5. Combining Theorems 2.3.4 and 2.1.1, we get

$$\mathsf{P}\Big(\mathsf{E}' \sup_{t\in T}(Y_t(\omega,\omega') - Y_{t_0}(\omega,\omega')) \ge \frac{1}{L}\gamma_\infty(T,d)\Big) \ge \frac{1}{L}$$

and we apply Lemma 2.3.6 given ω. □

2.4 Further Reading: Stationarity

In contrast with the Gaussian case, the inequality of Theorem 2.3.1 cannot be reversed. This, and much more, will be apparent in Chapter 5. There is however a special case of interest where everything behaves much better. It is the case where one has a kind of "stationarity". The remarkable fact about "stationary" situations is that lower bounds such as that of Theorem 2.3.1 can often be reversed, yielding a complete understanding. To give a specific example of what "stationarity" means without going into technicalities, consider the case of a Gaussian processes, where T is a locally compact group, and where the distance induced by the process is invariant under the action of the group. This case is historically important, because it is connected with the classical topic of random Fourier series. In this case, the lower bound of Theorem 1.3.2 was discovered by X. Fernique [10]. Now that the correct approach has been found, Fernique result is however not really simpler to prove than Theorem 2.1.1.

Fernique's result was an essential ingredient of the solution of all the classical problems on random Fourier series by Marcus and Pisier [22], [23]. A somewhat simpler treatment of the work of Marcus and Pisier is given in [18], and there is no point to reproduce it here. The work of Marcus and Pisier was extended by Marcus [21] to more general situations (that involve the infinitely divisible processes that we will study in Chapter 5). Marcus fails however to obtain necessary and sufficient conditions. Obtaining these requires the ideas of "families of distances" discussed in Section 5.1, and is done in the paper [49]. This paper arguably contains results that go far beyond the classical ones (in particular in the case of random Fourier series with random coefficients that do not have second moments), and obtaining these does not require really harder work than to get the classical results. Not surprisingly however, the paper [49] is apparently still waiting for its first reader.

2.5 Order 2 Gaussian Chaos

. Consider independent standard normal sequences (g_i) , (g'_j) , i , $j \ge 1$. Given a double sequence $t = (t_{i,j})_{i,j\ge 1}$ we consider the r.v.

$$X_t = \sum_{i,j\ge 1} t_{i,j} g_i g'_j \ . \tag{2.60}$$

The series converges in ℓ^2 as soon as $\sum_{i,j\ge 1} t_{i,j}^2 < \infty$. This random variable is called a (decoupled) order 2 Gaussian chaos. There is also a theory of non-decoupled chaos, $\sum_{i,j\ge 1} t_{i,j} g_i g_j$. For the present purposes, this theory reduces to the decoupled case using well understood arguments.

Given a finite family T of double sequences $t = (t_{i,j})$, we would like to find upper and lower bounds for the quantity

$$S(T) = \mathsf{E} \sup_{t\in T} X_t \ . \tag{2.61}$$

We find it convenient to assume that the underlying probability space is a product $(\Omega \times \Omega'$, $\mathsf{P} = \mathsf{P}_0 \otimes \mathsf{P}')$, so that

$$X_t(\omega, \omega') = \sum_{i,j} t_{i,j} g_i(\omega) g'_j(\omega') \ .$$

We denote by E' integration in ω' only (i.e. conditional expectation given ω).

Conditionally on ω , X_t is a Gaussian r.v. and

$$\mathsf{E}' X_t^2 = \sum_{j\ge 1} \Big(\sum_{i\ge 1} t_{i,j} g_i(\omega) \Big)^2 \ . \tag{2.62}$$

Consider

$$\sigma_t = \sigma_t(\omega) = (\mathsf{E}' X_t^2)^{1/2} \ .$$

Then

$$\begin{aligned} \sigma_t &= \sup_\alpha \sum_{j\ge 1} \alpha_j \Big(\sum_{i\ge 1} t_{i,j} g_i(\omega) \Big) \\ &= \sup_\alpha \sum_{i\ge 1} g_i(\omega) \Big(\sum_{j\ge 1} \alpha_j t_{i,j} \Big) := \sup_\alpha g_{t,\alpha} \end{aligned} \tag{2.63}$$

where the supremum is over the sequences $\alpha = (\alpha_j)$ with $\sum_{j\ge 1} \alpha_j^2 \le 1$.

Let us set

$$\begin{aligned} \|t\| &= \sup_\alpha \Big(\sum_{i\ge 1} \Big(\sum_{j\ge 1} \alpha_j t_{i,j} \Big)^2 \Big)^{1/2} \\ &= \sup \Big\{ \sum_{i,j\ge 1} \alpha_j \beta_i t_{i,j} \ ; \ \sum_{j\ge 1} \alpha_j^2 \le 1 \, , \, \sum_{i\ge 1} \beta_i^2 \le 1 \Big\} \ . \end{aligned}$$

If we think of t as a matrix, $\|t\|$ is the operator norm of t from ℓ^2 to ℓ^2. We will also need the Hilbert-Schmidt norm of this matrix, given by

$$\|t\|_{HS} = \Big(\sum_{i,j\ge 1} t_{i,j}^2\Big)^{1/2} .$$

We note that by the Cauchy-Schwarz inequality we have $\|t\| \le \|t\|_{HS}$. We also have

$$(\mathsf{E}g_{t,\alpha}^2)^{1/2} = \Big(\sum_{i\ge 1}\Big(\sum_{j\ge 1}\alpha_j t_{i,j}\Big)^2\Big)^{1/2} \le \|t\| ,$$

and (2.4) implies that for $v > 0$,

$$\mathsf{P}(|\sigma_t - \mathsf{E}\sigma_t| \ge v) \le 2\exp\Big(-\frac{v^2}{2\|t\|^2}\Big) \tag{2.64}$$

so that

$$\mathsf{E}(\sigma_t - \mathsf{E}\sigma_t)^2 \le L\|t\|^2 .$$

Denoting by $\|\cdot\|_2$ the norm in $L^2(\Omega)$, we thus have $\|\sigma_t - \mathsf{E}\sigma_t\|_2 \le L\|t\|$, so that $|\|\sigma_t\|_2 - |\mathsf{E}\sigma_t|| \le L\|t\|$. Now

$$\|\sigma_t\|_2 = (\mathsf{E}\sigma_t^2)^{1/2} = (\mathsf{E}X_t^2)^{1/2} = \|t\|_{HS} , \tag{2.65}$$

so that (2.64) implies

$$\mathsf{P}(|\sigma_t - \|t\|_{HS}| \ge v + L\|t\|) \le 2\exp\Big(-\frac{v^2}{2\|t\|^2}\Big) . \tag{2.66}$$

Taking $v = \|t\|_{HS}/4$, and distinguishing the cases whether $L\|t\| \le \|t\|_{HS}/4$ or not, we get

$$\mathsf{P}\Big(\sigma_t \le \frac{\|t\|_{HS}}{2}\Big) \le L\exp\Big(-\frac{\|t\|_{HS}^2}{L\|t\|^2}\Big) . \tag{2.67}$$

The random distance d_ω associated to the Gaussian process X_t (at given ω) is

$$d_\omega(s,t) = \sigma_{s-t}(\omega) .$$

Considering the two distances on T defined by

$$d_1(s,t) = \|t-s\| ,\ d_2(s,t) = \|t-s\|_{HS}$$

we then have shown that

$$\mathsf{P}\Big(d_\omega(s,t) \le \frac{1}{2}d_2(s,t)\Big) \le L\exp\Big(-\frac{d_2^2(s,t)}{Ld_1^2(s,t)}\Big) . \tag{2.68}$$

Let us prove another simple fact.

Lemma 2.5.1. *For $v \geq 0$ we have*

$$\mathsf{P}(|X_t| \geq v) \leq L \exp\Big(-\frac{1}{L}\min\Big(\frac{v^2}{\|t\|_{HS}^2}\,,\,\frac{v}{\|t\|}\Big)\Big)\,. \tag{2.69}$$

Proof. Given ω, the r.v. X_t is Gaussian so that

$$\mathsf{P}'(|X_t| \geq v) \leq 2\exp\Big(-\frac{v^2}{2\sigma_t^2}\Big)\,,$$

and, given $a > 0$

$$\begin{aligned}\mathsf{P}(|X_t| \geq v) = \mathsf{E}\mathsf{P}'(|X_t| \geq v) &\leq 2\mathsf{E}\exp\Big(-\frac{v^2}{2\sigma_t^2}\Big)\\ &\leq 2\exp\Big(-\frac{v^2}{2a^2}\Big) + 2\mathsf{P}(\sigma_t \geq a)\,.\end{aligned}$$

If $a \geq L\|t\|_{HS}$, it follows from (2.66) that

$$\mathsf{P}(\sigma_t \geq a) \leq L\exp\Big(-\frac{a^2}{L\|t\|^2}\Big)\,.$$

and thus

$$\mathsf{P}(|X_t| \geq v) \leq 2\exp\Big(-\frac{v^2}{2a^2}\Big) + L\exp\Big(-\frac{a^2}{L\|t\|^2}\Big)\,. \tag{2.70}$$

To finish the proof we take $a = \max\big(L\|t\|_{HS}\,,\,\sqrt{v\|t\|}\big)$ and we observe that the last term in (2.70) is always at most $L\exp(-v/(L\|t\|))$. $\square$

As a consequence of (2.69), we have

$$\mathsf{P}(|X_s - X_t| \geq v) \leq L\exp\Big(-\frac{1}{L}\min\Big(\frac{v^2}{d_2^2(s,t)}\,,\,\frac{v}{d_1(s,t)}\Big)\Big)$$

and Theorem 1.2.7 implies the following.

Theorem 2.5.2. *For a set T of sequences $(t_{i,j})$, we have*

$$\mathsf{E}\sup_{t\in T} X_t \leq L\big(\gamma_1(T,d_1) + \gamma_2(T,d_2)\big)\,. \tag{2.71}$$

At the end of this section, we will explain why there is no hope to reverse this inequality. However, we have the following, where we recall the notation (2.61).

Theorem 2.5.3. *We have*

$$\gamma_2(T,d_2) \leq L\big(S(T) + \sqrt{S(T)\gamma_1(T,d_1)}\big)\,. \tag{2.72}$$

Combining with Theorem 2.5.2, we have the following.

Corollary 2.5.4. *If*

$$R = \frac{\gamma_1(T, d_1)}{\gamma_2(T, d_2)} ,$$

then

$$\frac{1}{L(1+R)}\gamma_2(T, d_2) \le S(T) \le L(1+R)\gamma_2(T, d_2) \tag{2.73}$$

Proof. The right-hand side is obvious from (2.71). To obtain the left-hand side, we simply write in (2.72) that, since $\sqrt{ab} \le (a+b)/2$,

$$\begin{aligned}\sqrt{S(T)\gamma_1(T, d_1)} &= \sqrt{S(T)R\gamma_2(T, d_2)} \\ &\le \frac{1}{2}\Big(\frac{1}{L}\gamma_2(T, d_2) + LS(T)R\Big)\end{aligned}$$

where L is as in (2.72), and this yields

$$\gamma_2(T, d) \le LS(T) + \frac{1}{2}\gamma_2(T, d) + LS(T)R .$$

□

Theorem 2.5.3 relies on the following abstract statement.

Theorem 2.5.5. *Consider a finite set T, provided with two distances d_1 and d_2. Consider a random distance d_ω on T, and assume that*

$$\forall s\,,\ t \in T\,,\ \mathsf{P}\Big(d_\omega(s, t) \ge \frac{1}{L_1}d_2(s, t)\Big) \ge \frac{1}{L_1} \tag{2.74}$$

$$\forall s\,,\ t \in T\,,\ \mathsf{P}\big(d_\omega(s, t) \le d_2(s, t)\big) \le L_1 \exp\Big(-\frac{d_2^2(s, t)}{d_1^2(s, t)}\Big) . \tag{2.75}$$

Consider a number M such that

$$\mathsf{P}(\gamma_2(T, d_\omega) \le M) \ge 1 - \frac{1}{2L_1} . \tag{2.76}$$

Then

$$\gamma_2(T, d_2) \le L\big(M + \sqrt{M\gamma_1(T, d_1)}\big) . \tag{2.77}$$

Proof of Theorem 2.5.3. By (2.68), the pair of distances d_1 and $d_2' = d_2/L$ satisfy (2.75). The formula (2.63) makes σ_t, and hence σ_{s-t} appear as the supremum of a Gaussian process. Applying (2.59) to this process, we see that $\mathsf{P}(\sigma_{s-t} \ge (\mathsf{E}\sigma_{s-t}^2)^{1/2}/L) \ge 1/L$. Thus, using (2.65) we see that (2.74) holds if L_1 is large enough. Since $\mathsf{E}\mathsf{E}' \sup_{t\in T} X_t = S(T)$, and since $\mathsf{E}' \sup_{t\in T} X_t \ge 0$, by Markov inequality we have

$$\mathsf{P}\Big(\mathsf{E}' \sup_{t\in T} X_t \le 2L_1 S(T)\Big) \ge 1 - \frac{1}{2L_1} .$$

It then follows from Theorem 2.1.1 that (2.76) holds for $M = LS(T)$. □

Proof of Theorem 2.5.5. We consider the subset U of Ω given by $U = \{\gamma_2(T, d_\omega) \le M\}$, so that $\mathsf{P}(U) \ge 1 - 1/(2L_1)$ by hypothesis. Let us fix once and for all an admissible sequence $(\mathcal{C}_n)_{n\ge 0}$ of partitions of T such that

$$\forall t \in T\,,\ \sum_{n\ge 0} 2^n \Delta(C_n(t), d_1) \le 2\gamma_1(T, d_1)\ .$$

We consider an integer $\tau \ge 0$, that will be chosen later. Given a probability measure μ on T, we define

$$F_n(\mu) = \mathsf{E}\Big(\mathbf{1}_U \inf \int \Big(\sum_{k\ge n} 2^{k/2}\Delta(A_k(t), d_\omega) + \sum_{\ell\ge n} 2^\ell \Delta(C_{\ell+\tau}(t, d_1))\Big)\mathrm{d}\mu(t)\Big),$$

where the infimum is over all choices of the admissible sequence $(\mathcal{A}_k)$. Given $A \subset T$, we define

$$F_n(A) = \sup\{F_n(\mu)\ ;\ \exists C \in \mathcal{C}_{n+\tau}\,,\ \mu(C \cap A) = 1\}\ .$$

Thus, since $\int f(t)\mathrm{d}\mu(t) \le \sup_{t\in T} f(t)$, we get

$$\begin{aligned} F_0(T) &\le \mathsf{E}\Big(\mathbf{1}_U \inf\Big(\sup_{t\in T}\sum_{k\ge 0} 2^{k/2}\Delta(A_k(t), d_\omega) + \sup_{t\in T}\sum_{\ell\ge 0} 2^\ell \Delta(C_{\ell+\tau}(t), d_1)\Big)\Big) \\ &\le \mathsf{E}\big(\mathbf{1}_U(\gamma_2(T, d_\omega) + 2^{-\tau+1}\gamma_1(T, d_1))\big) \qquad (2.78) \\ &\le M + 2^{-\tau+1}\gamma_1(T, d_1)\ , \end{aligned}$$

where in the second inequality we have used the fact that

$$\sup_{t\in T}\sum_{\ell\ge 0} 2^{\ell+\tau}\Delta(C_{\ell+\tau}(t), d_1) \le \sup_{t\in T}\sum_{k\ge 0} 2^k \Delta(C_k(t), d_1) \le 2\gamma_1(T, d_1)\ .$$

Consider $n \ge 0$, and set $m = N_{n+\tau+3}$. Consider points $(t_\ell)_{\ell\le m}$ of T, with $d_2(t_\ell, t_{\ell'}) \ge 4a$ when $\ell \ne \ell'$ and sets $H_\ell \subset B_2(t_\ell, a)$. We will prove that if $\tau \ge \tau_0$, where τ_0 depends only on the value of the constant L_1, we have

$$F_n\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge \frac{2^{n/2}}{L}a + \min_{\ell\le m} F_{n+1}(H_\ell)\,, \qquad (2.79)$$

but before doing this we finish the argument. Using Theorem 1.3.2 with $r = 4$, $\theta(n) = 2^{n/2}/L$, and $\tau + 3$ rather than τ, we get

$$\gamma_2(T, d_2) \le L2^{\tau/2}(F_0(T) + \Delta(T, d_2))\ . \qquad (2.80)$$

Considering s, $t \in T$ with $d_2(t, s) = \Delta(T, d_2)$, we see from (2.74) that

$$\mathsf{P}\Big(d_\omega(t, s) \ge \frac{1}{L_1}\Delta(T, d_2)\Big) \ge \frac{1}{L_1}\ .$$

Since $\gamma_2(T, d_\omega) \geq d_\omega(t, s)$, and since $1 - 1/(2L_1) + 1/L_1 > 1$, we see from (2.76) that $\Delta(T, d_2) \leq LM$. Thus (2.78) and (2.80) imply that

$$\gamma_2(T, d_2) \leq L2^{\tau/2}(M + 2^{-\tau}\gamma_1(T, d_1)) .$$

Optimization over $\tau \geq \tau_0$ then gives (2.77).

We turn to the proof of (2.79). It closely resembles the proof of (2.46). Consider $c < \inf_\ell F_{n+1}(H_\ell)$, and for $\ell \leq m$ consider a set $C_\ell \in \mathcal{C}_{n+\tau+1}$ and a probability measure μ_ℓ on $H_\ell \cap C_\ell$ such that $F_{n+1}(\mu_\ell) > c$. Since $m = N_{n+\tau+3} \geq N_{n+\tau+2} N_{n+\tau}$ we can find a subset I of $\{1, \dots, m\}$ with $\operatorname{card} I \geq N_{n+\tau+2}$ such that for all $\ell \in I$ the set C_ℓ is contained in the same element C_0 of $\mathcal{C}_{n+\tau}$. We set

$$\mu = \frac{1}{\operatorname{card} I} \sum_{\ell \in I} \mu_\ell$$

so that $\mu(\bigcup_{\ell \leq m} H_\ell \cap C_0) = 1$. Thus, all we have to do is to prove that

$$F_n(\mu) \geq \frac{2^{n/2}}{L} a + c .$$

Since for t in the support of μ we have $C_{n+\tau}(t) = C_0$, proceeding as in the proof of Theorem 2.3.3 we see that it suffices to prove that

$$2^n \Delta(C_0, d_1) + \mathsf{E}\Big(\mathbf{1}_U \inf \int 2^{n/2} \Delta(A_n(t), d_\omega) \mathrm{d}\mu(t)\Big) \geq \frac{a 2^{n/2}}{L} . \tag{2.81}$$

Consider an integer q. This integer will be determined later, and its value will depend only on the value of the constant L_1. If $\Delta(C_0, d_1) > a2^{-n/2-q}$, then (2.81) holds true, so that we can assume that $\Delta(C_0, d_1) \leq a2^{-n/2-q}$.

Given the partition $\mathcal{A}_n$, let us define the random subset D of T by

$$D = \bigcup \{A \in \mathcal{A}_n \ ; \ \Delta(A, d_\omega) \leq 2a\} .$$

For $t \notin D$, we have $\Delta(A_n(t), d_\omega) > 2a$, and thus we have

$$\int \Delta(A_n(t), d_\omega) \mathrm{d}\mu(t) \geq 2a\mu(T \backslash D) .$$

As shown in (2.50), to prove (2.81) it suffices to show that $\mathsf{E} \sup \mu(D) \leq 1/(4L_1)$. Let us define

$$\mathcal{D} = \{A \in \mathcal{A}_n \ ; \ A \subset D\} = \{A \in \mathcal{A}_n \ ; \ \Delta(A, d_\omega) \leq 2a\} .$$

Then, as in (2.54) we see that $\mu(D) \leq (N_n \mu^{\otimes 2}(V_\omega))^{1/2}$, where now $V_\omega = \{(s,t) \in C_0^2 \ ; \ d_\omega(s,t) \leq 2a\}$. We still have (2.55), where now

$$W_\omega = \big\{(s,t) \in C_0^2 \ ; \ d_\omega(s,t) \leq 2a \, ; \, d_2(s,t) \geq 2a\big\} .$$

By (2.75), for $d_2(s,t) \geq 2a$ and $d_1(s,t) \leq a2^{-n/2-q}$ we have

$$\mathsf{P}(d_\omega(s,t) \leq 2a) \leq L_1 \exp\bigl(-2^{n+2q+2}\bigr) ,$$

and rather than (2.56) we now have

$$\mathsf{E} \sup \mu(D) \leq \Bigl(N_n \Bigl(\frac{1}{N_{n+\tau+2}} + L_1 \exp\bigl(-2^{n+2q+2}\bigr)\Bigr)\Bigr)^{1/2} .$$

We see that indeed we can choose q and τ_0 depending only on L_1 such that for $\tau \geq \tau_0$ this is at most $1/(4L_1)$. □

Let us give a simple consequence of Theorem 2.5.3. We recall the covering numbers $N(T,d,\epsilon)$ of Section 1.2.

Proposition 2.5.6. *There exists a constant L such if*

$$\Delta(T,d_1) \leq \alpha , \tag{2.82}$$

then

$$\log N\bigl(T, d_2, L\sqrt{\alpha S(T)}\bigr) \leq \frac{S(T)}{\alpha} . \tag{2.83}$$

Proof. Consider a set T with $\operatorname{card} T = m$, and assume that for certain numbers ϵ and α, we have (2.82) and

$$\forall s\,,\, t \in T\,,\, s \neq t\,,\, d_2(s,t) = \|t-s\|_{HS} \geq \epsilon . \tag{2.84}$$

By Lemma 2.1.2 and Theorem 2.1.1 we have $\gamma_2(T,d_2) \geq \epsilon\sqrt{\log m}/L$. (The reader is encouraged to find a simple direct argument for this fact.) By (2.82), we have $\gamma_1(T,d_1) \leq L\alpha \log m$, as is witnessed by an admissible sequence $(\mathcal{A}_n)$ such that if $N_n \geq m$, then each set $A \in \mathcal{A}_n$ contains exactly one point. By (2.72), we have

$$\begin{aligned}\frac{\epsilon}{L}\sqrt{\log m} \leq \gamma_2(T,d_2) &\leq L\bigl(S(T) + \sqrt{S(T)\gamma_1(T,d_1)}\bigr)\\ &\leq L\bigl(S(T) + \sqrt{S(T)\alpha \log m}\bigr) .\end{aligned}$$

Thus, if $\epsilon \geq L_2\sqrt{\alpha S(T)}$, we have

$$2L\sqrt{\alpha S(T)\log m} \leq LS(T) + L\sqrt{S(T)\alpha \log m}$$

and thus $\alpha \log m \leq S(T)$.

If now T' is given satisfying (2.82), consider $T \subset T'$ that satisfies (2.84) and has a cardinality m as large as possible. Then we have shown that if $\epsilon \geq L_2\sqrt{\alpha S(T')}$ we must have $\alpha \log m \leq S(T) \leq S(T')$. Thus $\alpha \log m \leq S(T')$, and, since the cardinality of T is as large as possible, the balls centered at the points of T of radius 2ϵ cover T'. □

Remark 2.5.7. If $\epsilon = L\sqrt{\alpha S(T)}$, then $S(T)/\alpha = L^2 S^2(T)/\epsilon^2$. Thus one can reformulate Proposition 2.5.6 as follows. If T satisfies (2.82), then

$$S(T) = \mathsf{E}\sup_{t\in T} X_t \geq \frac{1}{L}\epsilon\sqrt{\log N(T, d_2, \epsilon)}$$

provided $\epsilon \geq L\sqrt{\alpha S(T)}$.

A very interesting example of a set T is as follows. Given an integer n, we take

$$T = \{t\ ;\ \|t\| \leq 1\,,\, t_{i,j} \neq 0 \Rightarrow i\,,\, j \leq n\}\ .$$

Since

$$\sum_{i,j} t_{ij} g_i g'_j \leq \Big(\sum_{i\leq n} g_i^2\Big)^{1/2}\Big(\sum_{j\leq n} g'^2_j\Big)^{1/2}\|t\|\,,$$

by the Cauchy-Schwarz inequality we have $S(T) \leq n$. On the other hand, by volume arguments, we have $\log N(T, d_1, 1/4) \geq n^2/L$, so that $\gamma_1(T, d_1) \geq n^2/L$. It is also simple to see that (see [18])

$$\log N(T, d_2, \sqrt{n}/L) \geq n^2/L\ .$$

In fact it is simple to show that $S(T)$ is about n, $\gamma_1(T, d_1)$ is about n^2 and $\gamma_2(T, d_2)$ is about $n^{3/2}$, so that in (2.73) the lower bound is tight. The upper bound however is not tight, which means that there is no hope of reversing the inequality (2.71).

For completeness let us mention the following.

Proposition 2.5.8. *For each $\epsilon > 0$, we have*

$$\epsilon(\log N(T, d_2, \epsilon))^{1/4} \leq LS(T)\ . \tag{2.85}$$

In the previous example, both sides are of order n for $\epsilon = \sqrt{n}/L$.

Research problem 2.5.9. Is it true that

$$\epsilon\sqrt{\log N(T, d_1, \epsilon)} \leq LS(T)\ ? \tag{2.86}$$

For a partial result, and a proof of Proposition 2.5.8, see [50].

It is interesting to observe that (2.85) would follow from (2.86) and (2.83). Indeed by (2.86) we would have

$$\log N(T, d_1, \alpha) \leq L\frac{S^2(T)}{\alpha^2}$$

and by (2.83) if B is a ball $B_1(t, \alpha)$ of T, we have

$$\log N\big(B, d_2, L\sqrt{\alpha S(T)}\big) \leq \frac{S(T)}{\alpha}\ .$$

Combining these, we would have

$$\log N\big(T, d_2, L\sqrt{\alpha S(T)}\big) \le L\Big(\frac{S^2(T)}{\alpha^2} + \frac{S(T)}{\alpha}\Big)$$

and taking α such that $L\sqrt{\alpha S(T)} = \epsilon$, i.e. $\alpha = \epsilon^2/LS(T)$ would prove (2.85).

To conclude this section, we describe a way to control $S(T)$ from above, which is really different from the method of Theorem 2.5.2.

Given a convex balanced subset U of ℓ^2 (that is, $\lambda U \subset U$ for $|\lambda| \le 1$, or, equivalently, $U = -U$), we write

$$g(U) = \mathsf{E} \sup_{(u_i)\in U} \sum_{i\ge 1} u_i g_i$$

$$\sigma(U) = \sup_{(u_i)\in U} \Big(\sum_{i\ge 1} u_i^2\Big)^{1/2} .$$

Given U , V convex balanced subsets of ℓ^2, we write

$$T_{U,V} = \Big\{t = (t_{i,j}) \; ; \; \forall (x_i)_{i\ge 1} \, , \, \forall (y_j)_{j\ge 1} \, ,$$
$$\sum t_{i,j} x_i y_j \le \sup_{(u_i)\in U} \sum_{i\ge 1} x_i u_i \sup_{(v_j)\in V} \sum_{j\ge 1} y_j v_j \Big\} .$$

It follows from (2.4) that, if $w > 0$,

$$\mathsf{P}\Big(\sup_{(u_i)\in U} \sum_{i\ge 1} g_i u_i \ge g(U) + w\sigma(U)\Big) \le 2\exp\big(-\frac{w^2}{2}\big)$$

so that (using that for positive numbers, when $ab > cd$ we have either $a > c$ or $b > d$)

$$\mathsf{P}\Bigg(\sup_{(u_i)\in U} \sum_{i\ge 1} g_i u_i \sup_{(v_j)\in V} \sum g'_j v_j \ge g(U)g(V) \tag{2.87}$$
$$+ \; w(\sigma(U)g(V) + \sigma(V)g(U)) + w^2\sigma(U)\sigma(V)\Bigg)$$
$$\le \; 4\exp\Big(-\frac{w^2}{2}\Big) .$$

We note that

$$\sup_{t\in T_{U,V}} X_t \le \sup_{(u_i)\in U} \sum_{i\ge 1} u_i g_i \sup_{(v_j)\in V} \sum_{j\ge 1} v_j g'_j \, ,$$

so that, if $g(U)\, ,\, g(V) \le 1\, ,\, \sigma(U)\, ,\, \sigma(V) \le 2^{-n/2}$, changing w into $2^{n/2}w$, we see from (2.87) that

$$\mathsf{P}\Big(\sup_{t\in T_{U,V}} X_t \ge (1+w)^2\Big) \le 4\exp(-2^{n-1}w^2) \tag{2.88}$$

Proposition 2.5.10. *Consider for $n \geq 0$ a family $\mathcal{C}_n$ of pairs of convex balanced subsets of ℓ^2. Assume that* $\operatorname{card} \mathcal{C}_n \leq N_n$ *and that*

$$\forall (U,V) \in \mathcal{C}_n \,,\; g(U)\,,\, g(V) \leq 1 \;;\; \sigma(U)\,,\, \sigma(V) \leq 2^{-n/2} \,.$$

Then, if

$$T = \operatorname{conv}\Big\{ \bigcup_n \bigcup_{(U,V)\in\mathcal{C}_n} T_{U,V} \Big\}$$

we have $S(T) \leq L$.

Proof. This should be obvious from (2.88) writing

$$\mathsf{P}\Big(\sup_T X_t \geq w\Big) \leq \sum_n \sum_{(U,V)\in\mathcal{C}_n} \mathsf{P}\Big(\sup_{t\in T_{U,V}} X_t \geq w\Big) \,.$$

□

Having now found two very distinct ways of controlling $S(T)$ from above, it is natural to ask whether these are essentially the only ways.

Research problem 2.5.11. Does there exist a number L with the following property: given a (finite) set T of sequences $(t_{i,j})_{i,j\geq 1}$, with $0 \in T$ and $S(T) \leq 1$, can one find a sequence $(t_n)_{n\geq 2}$ with $\|t_n\| \leq 1/\log n$, $\|t_n\|_{HS} \leq 1/\sqrt{\log n}$, and families $\mathcal{C}_n$ as in Proposition 2.5.10 such that

$$T \subset L \operatorname{conv}\Big(\bigcup\{t_n \,;\, n \geq 2\} \cup \bigcup \big\{T_{U,V} \,;\, (U,V) \in \mathcal{C}_n \,,\, n \geq 1\big\}\Big) \;?$$

The reason for the sequence (t_n) is as follows. If $0 \in T$, $\gamma_1(T,d_1) \leq 1$ and $\gamma_2(T,d_2) \leq 1$, then $T \subset L \operatorname{conv}\{t_n \;;\; n \geq 1\}$ for a sequence t_n as in Problem 2.5.11. This can be shown by a simple modification of the argument of Theorem 2.1.8.

The only (flimsy...) support for a positive answer to this problem is our failure to imagine another method to control $S(T)$ from above.

2.6 L^2 , L^1 , L^∞ Balls

In this section we consider a measured space (Ω, Σ, μ), and the classical Banach spaces $L^1 = L^1(\mu)\,,\, L^2 = L^2(\mu)\,,\, L^\infty = L^\infty(\mu)$. There will be two situations of special interest. The first one is when μ is a probability. The second one is when Ω is countable and when μ is the counting measure, $\mu(A) = \operatorname{card} A$. In this case $L^2(\mu)$ identifies with $\ell^2(\mathbb{N}^*)$. We denote by B_1 the unit ball of $L^1(\mu)$.

Lemma 2.6.1. *Consider $f \in L^2$ and $u > 0$. Then we can write $f = f_1 + f_2$ where*

$$\|f_1\|_2 \leq \|f\|_2 \;,\; \|f_1\|_\infty \leq u \;\;;\;\; \|f_2\|_2 \leq \|f\|_2 \;,\; \|f_2\|_1 \leq \frac{\|f\|_2^2}{u} \,. \tag{2.89}$$

Proof. We set $f_1 = f\mathbf{1}_{\{|f|\le u\}}$, so that the first part of (2.89) is obvious. We set $f_2 = f\mathbf{1}_{\{|f|>u\}} = f - f_1$, so that

$$u\|f_2\|_1 = \int u|f|\mathbf{1}_{\{|f|>u\}}\mathrm{d}\mu \le \int f^2\mathrm{d}\mu = \|f\|_2^2 \, . \tag{2.90}$$

□

The following is a version of Lemma 2.6.1 for classes of functions.

Theorem 2.6.2. *Consider a countable set $T \subset L^2(\mu)$, and a number $u > 0$. Assume that $S = \gamma_2(T, d_2) < \infty$. Then there is a decomposition $T \subset T_1 + T_2$ where*

$$\gamma_2(T_1, d_2) \le LS \; ; \; \gamma_1(T_1, d_\infty) \le LSu \tag{2.91}$$

$$\gamma_2(T_2, d_2) \le LS \; ; \; T_2 \subset \frac{LS}{u} B_1 \, . \tag{2.92}$$

Here d_2 is the distance induced by the norm $\|\cdot\|_2$ (etc.). Also,

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

In words, we can reconstruct T from the two sets T_1 , T_2. These two sets are not really larger than T with respect to γ_2. Moreover, for each of them we have some extra information: we control $\gamma_1(T_1, d_\infty)$, and we control the L^1 norm of the elements of T_2.

Proof. As usual, $\Delta_2(A)$ denotes the diameter of A for the distance d_2. We consider an admissible sequence of partitions $(\mathcal{A}_n)_{n\ge 0}$ such that

$$\forall t \in T \, , \, \sum_{n\ge 0} 2^{n/2}\Delta_2(A_n(t)) \le 2S \, . \tag{2.93}$$

Let us enumerate T as $(t_n)_{n\ge 0}$. By induction over n we pick points $t_A \in A$ for $A \in \mathcal{A}_n$. We make sure that for $A = A_n(t_n)$ we have $t_A = t_n$. Thus each point of T is of the type t_A for some m and $A = A_m(t)$. Let $\pi_n(t) = t_A$ where $A = A_n(t)$. For $n \ge 1$, let $f_{t,n} = \pi_n(t) - \pi_{n-1}(t)$, so that $f_{t,n}$ depends only on $A_n(t)$ and

$$\|f_{t,n}\|_2 \le \Delta_2(A_{n-1}(t)) \, . \tag{2.94}$$

Using Lemma 2.6.1 with $2^{-n/2}u\|f_{t,n}\|_2$ instead of u we can decompose $f_{t,n} = f_{t,n}^1 + f_{t,n}^2$ where

$$\|f_{t,n}^1\|_2 \le \|f_{t,n}\|_2 \, , \, \|f_{t,n}^1\|_\infty \le 2^{-n/2}u\|f_{t,n}\|_2 \tag{2.95}$$

$$\|f_{t,n}^2\|_2 \le \|f_{t,n}\|_2 \, , \, \|f_{t,n}^2\|_1 \le \frac{2^{n/2}}{u}\|f_{t,n}\|_2 \, . \tag{2.96}$$

Given $t \in T$ we set $g_{t,0}^1 = t_T$ and $g_{t,0}^2 = 0$, while if $n \ge 1$ we set

$$g_{t,n}^1 = t_T + \sum_{1 \le k \le n} f_{t,k}^1 \ , \ g_{t,n}^2 = \sum_{1 \le k \le n} f_{t,k}^2 \ .$$

We set

$$T_n^1 = \left\{ g_{t,m}^1 \ ; \ m \le n \ , \ t \in T \right\} \ ; \ T_n^2 = \left\{ g_{t,m}^2 \ ; \ m \le n \ , \ t \in T \right\}$$

$$T_1 = \bigcup_{n \ge 0} T_n^1 \ ; \ T_2 = \bigcup_{n \ge 0} T_n^2 \ .$$

We have $T \subset T_1 + T_2$. Indeed, if $t \in T$, then $t = t_A$ for some m and $A = A_m(t)$. Thus $\pi_n(t) = t$, and since $\pi_0(t) = t_T$ we have $t - t_T = \sum_{1 \le k \le m} f_{t,k}$, so that $t = g_{t,m}^1 + g_{t,m}^2$.

Since for $j = 1, 2$ the element $g_{t,n}^j$ depends only on $A_n(t)$, we have $\operatorname{card} T_n^j \le N_0 + \ldots + N_n$, so that $\operatorname{card} T_0^j = 1$ and $\operatorname{card} T_n^j \le N_{n+1}$. Consider $t^1 \in T_1$, so that $t^1 = g_{t,m}^1$ for some m and some $t \in T$. If $m \le n$ we have $t^1 = g_{t,m}^1 \in T_n^1$ so that $d(t^1, T_n^1) = 0$. If $m > n$ we have $g_{t,n}^1 \in T_n^1$, so that, using (2.94) and (2.95) we have

$$d_2(t^1, T_n^1) \le d_2(g_{t,m}^1, g_{t,n}^1) \le \sum_{k > n} \|f_{t,k}^1\|_2 \le \sum_{k > n} \Delta_2(A_{k-1}(t)) \ . \tag{2.97}$$

Hence

$$\begin{aligned} \sum_{n \ge 0} 2^{n/2} d_2(t^1, T_n^1) &\le \sum_{n \ge 0, k > n} 2^{n/2} \Delta_2(A_{k-1}(t)) \\ &\le L \sum_{k \ge 1} 2^{k/2} \Delta_2(A_{k-1}(t)) \le LS \ . \end{aligned}$$

It then follows from Theorem 1.3.5 (used for $\tau' = 1$) that $\gamma_2(T_1, d_2) \le LS$. The proof that $\gamma_2(T_2, d_2) \le LS$ is identical. The same approach works to control $\gamma_1(T_1, d_\infty)$. Indeed, we can replace (2.97) by

$$d_\infty(t^1, T_n^1) \le d_\infty(g_{t,m}^1, g_{t,n}^1) \le \sum_{k > n} \|f_{t,k}^1\|_\infty \le \sum_{k > n} 2^{-k/2} u \Delta_2(A_{k-1}(t)) \ .$$

Hence

$$\begin{aligned} \sum_{n \ge 0} 2^n d_\infty(t^1, T_n^1) &\le u \sum_{n \ge 0, k > n} 2^{n-k/2} \Delta_2(A_{k-1}(t)) \\ &\le Lu \sum_{k \ge 1} 2^{k/2} \Delta_2(A_{k-1}(t)) \le LuS \ , \end{aligned}$$

and it follows again from Theorem 1.3.5 that $\gamma_1(T_1, d_\infty) \le LS$. Finally, by (2.96) and (2.95) we have

$$\|g_{t,n}^2\|_1 \le \sum_{k \ge 1} \|f_{t,k}^2\|_1 \le \sum_{k \ge 1} \frac{2^{k/2}}{u} \Delta_2(A_{k-1}(t)) \le \frac{LS}{u} \ ,$$

so that $T_2 \subset LB_1/u$. This completes the proof. $\square$

The material of the rest of this section is a bit technical. This technical work will turn out later to be a good investment, but the first time reader should probably proceed directly to Chapter 3.

Later in the book we will need a more general theorem than Theorem 2.6.2. Some of the conditions occurring in this theorem might at this stage look completely arbitrary and their meaning will become clear only gradually. They are in particular related to the ideas of Chapter 5.

Theorem 2.6.3. *Consider a countable set T of measurable functions on Ω, a number $V \geq 2$, and assume that $0 \in T$. Consider an admissible sequence of partitions $(\mathcal{A}_n)$ of T, and for $A \in \mathcal{A}_n$ consider $j(A) \in \mathbb{Z}$ and a number $\delta(A) \in \mathbb{R}^+$, with the following properties:*

$$\forall t \in T\,,\ \lim_{n\to\infty} j(A_n(t)) = \infty \tag{2.98}$$

$$A \in \mathcal{A}_n\,,\ B \in \mathcal{A}_{n-1}\,,\ A \subset B \Rightarrow j(A) \geq j(B) \tag{2.99}$$

$$A \subset B\,,\ A \in \mathcal{A}_n\,,\ B \in \mathcal{A}_{n'}\,,\ j(A) = j(B) \Rightarrow \delta(B) \leq 2\delta(A) \tag{2.100}$$

$$\forall s\,,\ t \in A\,,\ \int (s(\omega) - t(\omega))^2 \wedge V^{-2j(A)} \mathrm{d}\mu(\omega) \leq \delta^2(A)\,, \tag{2.101}$$

where $x \wedge y = \min(x, y)$. Then we can write $T \subset T_1 + T_2 + T_3$ where

$$\gamma_2(T_1, d_2) \leq L \sup_{t\in T} \sum_{n\geq 0} 2^{n/2} \delta(A_n(t)) \tag{2.102}$$

$$\gamma_1(T_1, d_\infty) \leq L \sup_{t\in T} \sum_{n\geq 0} 2^n V^{-j(A_n(t))} \tag{2.103}$$

$$\forall t \in T_2\,,\ \forall p \geq 1\,,\ \|t\|_p^p \leq L^p \sup_{t\in T} \sum V^{2j(A_{n+1}(t)) - pj(A_n(t))} \delta^2(A_{n+1}(t))\,, \tag{2.104}$$

where the summation is over the $n \geq 0$ for which either $n = 0$ or $j(A_{n+1}(t)) > j(A_n(t))$. Moreover,

$$\forall t \in T_3\,,\ \exists s \in T\,,\ |t| \leq 5|s| \mathbf{1}_{\{2|s| \geq V^{-j(T)}\}}\,. \tag{2.105}$$

Conditions (2.98) to (2.100) are mild technical requirements. Condition (2.101) is weaker than the condition $\Delta(A, d_2) \leq \delta(A)$, in that it gives a much weaker control of the large values of $s - t$. The term T_3 of the decomposition is of secondary importance, and will be easy to control. It is required because (2.101) says little about the functions $|s|\mathbf{1}_{\{|s| \geq V^{-j(T)}\}}$. The important statements are (2.102) to (2.104). How one can use them is illustrated in the following proof.

Second proof of Theorem 2.6.2. We denote by $\Delta_2(A)$ the diameter of a set A for the L^2 norm. We consider an admissible sequence $(\mathcal{A}_n)$ of T such that

$$\forall t \in T\ ,\ \sum_{n\geq 0} 2^{n/2}\Delta_2(A_n(t)) \leq 2S\ . \tag{2.106}$$

We define $\delta(A) = \Delta_2(A)$, so that $A \subset B \Rightarrow \delta(A) \leq \delta(B)$. We take $V = 2$, and, given $u > 0$, for $A \in \mathcal{A}_n$ we define $j(A)$ as the largest integer for which

$$2^{-j(A)} \geq u2^{-n/2}\delta(A)\,, \tag{2.107}$$

so that

$$2^{-j(A)} \leq 2u2^{-n/2}\delta(A)\ , \tag{2.108}$$

and (2.98), (2.99) and (2.100) hold. As for (2.101), it is obvious by definition of $\delta(A)$. From (2.102) and (2.106) we see that $\gamma_2(T_1, d_2) \leq LS$, and from (2.103), (2.106) and (2.108) we see that $\gamma_1(T_1, d_\infty) \leq LuS$.

Using (2.104) for $p = 1$, and since $2^{2j(A_{n+1}(t))}\delta^2(A_{n+1}(t)) \leq 2^{n+1}/u^2$ by (2.107), we see from (2.108) and (2.106) that $\|t\|_1 \leq LS/u$ for $t \in T_2$.

We have $2^{-j(T)} \geq u\Delta_2(T)$ by (2.107) and since $0 \in T$, we have $\|s\|_2 \leq \Delta_2(T)$ for $s \in T$. Using (2.90) with $u = 2^{-j(T)}$ we get $\|s\mathbf{1}_{\{2|s|\geq 2^{-j(T)}\}}\|_1 \leq \|s\|_2^2/2^{-j(T)-1} \leq L\Delta_2(T)/u$. Since by (2.106), looking only at the term $n = 0$, we have $\Delta_2(T) \leq LS$, we get $\|t\|_1 \leq LS/u$ for $t \in T_3$. Setting $T_2' = T_2 + T_3$, we have shown that $\|t\|_1 \leq LS/u$ for $t \in T_2'$.

We have $T \subset T_1 + T_2'$ so that $T \subset T_1 + T_2''$ where $T_2'' = T_2' \cap (T - T_1)$, and since $\gamma_2(T, d_2) \leq LS$ and $\gamma_2(T_1, d_2) \leq LS$, we have $\gamma_2(T_2'', d_2) \leq LS$ by (2.14). $\square$

Proof of Theorem 2.6.3. This proof is rather tedious and technical, and reading it should be attempted only after motivation has been found through the subsequent applications of this principle, the first of which occurs in Chapter 4. We define

$$p(t,n) = \inf\big\{p \geq 0\ ;\ j(A_n(t)) = j(A_p(t))\big\},$$

and we observe that by (2.100) we have

$$\delta(A_{p(t,n)}(t)) \leq 2\delta(A_n(t))\ . \tag{2.109}$$

For $A \in \mathcal{A}_n$, $n \geq 1$, we choose an arbitrary point t_A in A. If $A = T$, we choose $t_A = 0$. We define

$$\pi_n(t) = t_B \quad \text{where } B = A_{p(t,n)}(t)\ .$$

It should be observed that $\pi_n(t)$ depends only on $A_n(t)$, i.e. if $s \in A_n(t)$ then $\pi_n(s) = \pi_n(t)$. We note also that $\pi_0(t) = 0$.

We define $j_n(t) = j(A_n(t))$ and

$$m(t,\omega) = \inf\big\{n \geq 0\ ;\ |\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| > V^{-j_n(t)}\big\}$$

if the set on the right is not empty and $m(t,\omega) = \infty$ otherwise. Thus

$$n < m(t,\omega) \Rightarrow |\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| \le V^{-j_n(t)} . \tag{2.110}$$

Let us note that the construction of $p(t,n)$ implies that $p(t,n+1) = p(t,n)$ as soon as $j_{n+1}(t) = j_n(t)$. Thus if $\pi_{n+1}(t) \neq \pi_n(t)$ we have $j_{n+1}(t) \ge j_n(t)+1$. Since $V \ge 2$, we deduce from (2.110) that if $n < m(t,\omega)$ then we have

$$\sum_{n \le m < m(t,\omega)} |\pi_{m+1}(t)(\omega) - \pi_m(t)(\omega)| \le 2V^{-j_n(t)} . \tag{2.111}$$

Let us define t^1 by $t^1(\omega) = \pi_{m(t,\omega)}(t)(\omega)$ if $m(t,\omega) < \infty$ and $t^1(\omega) = \lim_{n\to\infty} \pi_n(t)(\omega)$ if $m(t,\omega) = \infty$. The limit exists from (2.111) and (2.98), and since $\pi_0(t) = 0$, using (2.111) with $n = 0$ we have

$$|t^1(\omega)| \le 2V^{-j(T)} . \tag{2.112}$$

We define $T_1 = \{t^1 \,,\, t \in T\}$.

For $n \ge 0$, we define t^1_n by $t^1_n(\omega) = \pi_{n\wedge m(t,\omega)}(t)(\omega)$. It follows from the construction that $n \wedge m(t,\omega) = n \wedge m(s\,,\,\omega)$ if $A_n(s) = A_n(t)$. Thus if $U_n = \{t^1_n \ ;\ t \in T\}$, then $\operatorname{card} U_n \le N_n$. We note that $t^1(\omega) - t^1_n(\omega) = 0$ if $n \ge m(t,\omega)$, and by (2.111) that if $n < m(t,\omega)$, we have

$$|t^1(\omega) - t^1_n(\omega)| \le 2V^{-j_n(t)} .$$

Thus $\|t^1 - t^1_n\|_\infty \le 2V^{-j_n(t)}$, and hence $d_\infty(t^1, U_n) \le 2V^{-j_n(t)}$. Thus (2.103) follows from Theorem 1.3.5 with $\alpha = 1$.

We have

$$t^1_{n+1} - t^1_n = (\pi_{n+1}(t) - \pi_n(t))\mathbf{1}_{\{m(t,\cdot)>n\}}$$

and thus

$$|t^1_{n+1} - t^1_n| \le |\pi_{n+1}(t) - \pi_n(t)|\mathbf{1}_{\{|\pi_{n+1}(t)-\pi_n(t)| \le V^{-j_n(t)}\}} .$$

Now $\pi_n(t), \pi_{n+1}(t) \in A_{p(t,n)}(t)$ so that by (2.101) we have

$$\begin{aligned}\|t^1_{n+1} - t^1_n\|_2^2 &\le \int \big(\pi_n(t)(\omega) - \pi_{n+1}(t)(\omega)\big)^2 \wedge V^{-2j_n(t)} \mathrm{d}\mu(\omega)\\ &\le \delta^2(A_{p(t,n)}(t)) \le 4\delta^2(A_n(t))\end{aligned}$$

using (2.100). Thus $\|t^1_{n+1} - t^1_n\|_2 \le 2\delta(A_n(t))$ and thus

$$d_2(t^1, U_n) \le \|t^1 - t^1_n\|_2 \le \sum_{m \ge n} \|t^1_{m+1} - t^1_m\|_2 \le 2\sum_{m\ge n} \delta(A_m(t)) .$$

Since

$$\begin{aligned}\sum_{n\ge 0} 2^{n/2} \sum_{m \ge n} \delta(A_m(t)) &\le \sum_{m\ge 0} \delta(A_m(t)) \sum_{n \le m} 2^{n/2}\\ &\le L \sum_{m \ge 0} 2^{m/2} \delta(A_m(t)) ,\end{aligned}$$

we conclude by Theorem 1.3.5 again that (2.102) holds.

For $t \in T$, define $\Omega(t) = \{\omega\ ;\ |t(\omega)| \le V^{-j(T)}/2\}$ and $t^3 = (t - t^1)\mathbf{1}_{\Omega(t)^c}$. Since for $\omega \in \Omega(t)^c$ we have $|t(\omega)| \ge V^{-j(T)}/2$ and $|t^1(\omega)| \le 2V^{-j(T)}$ by (2.112), we have $|t^3| \le 5|t|\mathbf{1}_{\Omega(t)^c}$, so that $T_3 = \{t^3\ ;\ t \in T\}$ satisfies (2.105).

We set $t^2 = t - t^1 - t^3 = (t - t^1)\mathbf{1}_{\Omega(t)}$, $T_2 = \{t^2\,;\, t \in T\}$, and we turn to the proof of (2.104). We define

$$r(t,\omega) = \inf\Big\{n \ge 0\ ;\ |\pi_{n+1}(t)(\omega) - t(\omega)| \ge \frac{1}{2}V^{-j_{n+1}(t)}\Big\}$$

if the set on the right is not empty and $r(t,\omega) = \infty$ otherwise. For $0 \le n < r(t,\omega)$ and $\omega \in \Omega(t)$ we have

$$\begin{aligned}|\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| &\le |\pi_{n+1}(t)(\omega) - t(\omega)| + |\pi_n(t)(\omega) - t(\omega)| \\ &\le \frac{1}{2}(V^{-j_{n+1}(t)} + V^{-j_n(t)}) \le V^{-j_n(t)}\,,\end{aligned}$$

using that $n - 1 < r(t,\omega)$ if $n \ge 1$, and that

$$|\pi_0(t)(\omega) - t(\omega)| = |t(\omega)| \le V^{-j(T)}/2$$

if $n = 0$. Thus for $\omega \in \Omega(t)$ we have $r(t,\omega) \le m(t,\omega)$. Since $t(\omega) = \lim_{n\to\infty} \pi_n(t)(\omega) = t^1(\omega)$ when $r(t,\omega) = \infty$, we have

$$t^2 = (t - t^1)\mathbf{1}_{\Omega(t)} = \sum_{n\ge 0} t_n^2$$

where

$$t_n^2 = (t - t^1)\mathbf{1}_{\{r(t,\omega)=n\}\cap\Omega(t)}\ .$$

Now

$$\begin{aligned}|t_n^2(\omega)| &\le |t(\omega) - \pi_n(t)(\omega)|\mathbf{1}_{\{r(t,\omega)=n\}\cap\Omega(t)}(\omega) \\ &\quad + \sum_{n\le m<m(t,\omega)} |\pi_{m+1}(t)(\omega) - \pi_m(t)(\omega)| \\ &\le 3V^{-j_n(t)}\end{aligned} \tag{2.113}$$

using (2.111) and that $|t(\omega) - \pi_n(t)(\omega)| \le V^{-j_n(t)}/2$ if $r(t,\omega) \ge n$. Also, since $|\pi_{n+1}(t)(\omega) - t(\omega)| \ge V^{-j_{n+1}(t)}/2$ when $r(t,\omega) = n$, we have

$$\begin{aligned}&\mu\ (\{\omega\ ;\ r(t,\omega) = n\}) \cdot \Big(\frac{V^{-j_{n+1}(t)}}{2}\Big)^2 \\ &\le \int (\pi_{n+1}(t)(\omega) - t(\omega))^2 \wedge V^{-2j_{n+1}(t)} \mathrm{d}\mu(\omega) \\ &\le 4\delta^2(A_{n+1}(t))\,,\end{aligned}$$

using in the last inequality that $\pi_{n+1}(t)\,,\ t \in A_{p(t,n+1)}$, and (2.101), (2.109). Thus

$$\mu(\{t_n^2 \neq 0\}) \le LV^{2j_{n+1}(t)}\delta^2(A_{n+1}(t)) ,$$

so that, using (2.113)

$$\|t_n^2\|_p^p \le L^p V^{2j_{n+1}(t)}\delta^2(A_{n+1}(t))V^{-pj_n(t)} .$$

The functions $(t_n^2)_{n\ge 0}$ have disjoint supports. Moreover, $t_n^2 = 0$ unless the set $\{\omega \; ; \; r(t,\omega) = n\}$ is non-empty. When $n \ge 1$, this implies that $j_{n+1}(t) \neq j_n(t)$, for otherwise we would have $r(t,\omega) \le n-1$. Thus (2.104) is proved. □

The rest of the section is strongly influenced by ideas from Banach space theory. It is not related to any subsequent material except Section 6.3. A central result is the following, that provides a kind of classification of the elements of the unit ball B_1 of $L^1(\mu)$.

Theorem 2.6.4. *For any integer $\tau \ge 0$ there exists an admissible sequence of partitions $(\mathcal{C}_n)$ of B_1, and for each $C \in \mathcal{C}_n$ an integer $\ell(C) \in \mathbb{Z}$, such that if we set*

$$\ell(f,n) = \ell(C_n(f)) \tag{2.114}$$

we have

$$\forall f \in B_1 ,\; \mu\big(\{|f| > 2^{-\ell(f,n)}\}\big) \le 2^{n+\tau} \tag{2.115}$$

$$\forall f \in B_1 ,\; \sum_{n\ge 0} 2^{n-\ell(f,n)} \le 6\cdot 2^{-\tau} . \tag{2.116}$$

Proof. Let us define

$$u_n(f) = \inf\big\{u > 0 \; ; \; \mu(\{|f| > u\}) \le 2^n\big\}$$

so that

$$\mu(\{|f| > u_n(f)\}) \le 2^n . \tag{2.117}$$

We now claim that

$$\sum_{n\ge 1} 2^n u_n(f) \le 2 . \tag{2.118}$$

Indeed, we have

$$\begin{aligned}\frac{1}{2}\sum_{n\ge 1} 2^n u_n(f) &= \sum_{n\ge 1}(2^n - 2^{n-1})u_n(f) = \sum_{n\ge 1} 2^n(u_n(f) - u_{n+1}(f)) \\ &\le \sum_{n\ge 1}\int_{u_{n+1}(f)}^{u_n(f)} \mu(\{|f| \ge t\})\,\mathrm{d}t \le \|f\|_1 \le 1 .\end{aligned}$$

For $n \ge 1$, we define

$$\ell(f,n) = \sup\big\{\ell \; ; \; \ell \le 2n+\tau \; ; \; 2^{-\ell} \ge u_{n+\tau}(f)\big\} .$$

Since $u_n(f) \le 2^{-n}$ by Markov's inequality, we have

$$n + \tau \le \ell(f,n) \le 2n + \tau \; . \tag{2.119}$$

Since $2^{-\ell(f,n)} \le 2^{-2n-\tau} + 2u_{n+\tau}(f)$, by (2.118) we have

$$\sum_{n \ge 1} 2^{n-\ell(f,n)} \le 5 \cdot 2^{-\tau} \; . \tag{2.120}$$

We define $\mathcal{C}_0 = \{B_1\}$, and $\ell(B_1) = \tau$. Given integers ℓ_m for $1 \le m \le n$ such that $m + \tau \le \ell_m \le 2m + \tau$ we consider the set

$$\big\{ f \in B_1 \; ; \; \forall m \, , \, 1 \le m \le n \, , \, \ell(f,m) = \ell_m \big\} \; . \tag{2.121}$$

By (2.119) these sets form a partition $\mathcal{C}_n$ of B_1, and the sequence $(\mathcal{C}_n)$ increases. Moreover,

$$\operatorname{card} \mathcal{C}_n \le (n+1)! \le N_n \, ,$$

so that the sequence $(\mathcal{C}_n)$ is admissible.

If $C \in \mathcal{C}_n$ is given by (2.121), we define $\ell(C) = \ell_n$. We observe that for $f \in C$, we have $\ell(C_n(f)) = \ell(C) = \ell_n = \ell(f,n)$, so that (2.114) holds. For $f \in C$, we have

$$\mu(\{|f| > 2^{-\ell_n}\}) = \mu(\{|f| > 2^{-\ell(f,n)}\}) \le 2^{n+\tau}$$

by (2.117) and since $2^{-\ell(f,n)} \ge u_{n+\tau}(f)$. This proves (2.115). Finally (2.116) follows from (2.120). $\square$

In the rest of the section, $\Omega = \mathbb{N}^*$ and μ is the counting measure, so that $L^2 = \ell^2(\mathbb{N}^*)$, and $L^1 = \ell^1(\mathbb{N}^*)$.

Given a finite subset I of $\mathbb{N}^*$, and a number $a > 0$, we define

$$V(I,a) = \Big\{ x \in \ell^2 \; ; \; i \notin I \Rightarrow x_i = 0 \; ; \; \sum_{i \in I} x_i^2 \le a^2 \Big\} \; .$$

Thus

$$V(I,a) \subset aB_2 \, , \tag{2.122}$$

where of course

$$B_2 = \Big\{ x \; ; \; \sum_{i \ge 1} x_i^2 \le 1 \Big\} \; ; \; B_1 = \Big\{ x \; ; \; \sum_{i \ge 1} |x_i| \le 1 \Big\} \; .$$

We recall that for $T \subset \ell^2$ we write

$$g(T) = \mathsf{E} \sup_{t \in T} X_t = \mathsf{E} \sup_{t \in T} \sum_{i \ge 1} t_i g_i$$

and we note that

$$g(V(I,a)) \le a\sqrt{\operatorname{card} I} \tag{2.123}$$

since

$$\sum_{i\ge 1} x_i g_i \le a\Big(\sum_{i\in I} g_i^2\Big)^{1/2}$$

for $x \in V(I,a)$.

One can interpret Theorem 2.1.8 and Proposition 2.1.7 as stating that, from the point of view of Gaussian processes, the important sets are the convex hulls of a sequence going to 0 with the proper rate of convergence. This idea is behind the following definition.

Definition 2.6.5. *We say that a subset T of ℓ^2 is an explicit unconditional* GB-*set (*EIGB-*set) if there is a family $\mathcal{F}$ of pairs (I,a), where I is a finite subset of $\mathbb{N}^*$ and $a > 0$, such that $a\sqrt{\operatorname{card} I} \le 1$ for each $(I,a) \in \mathcal{F}$ and that*

$$T \subset T' = \overline{\operatorname{conv}} \bigcup_{\mathcal{F}} V(I,a) \tag{2.124}$$

$$\forall v > 0\,,\ \operatorname{card}\{(I,a) \in \mathcal{F}\ ;\ a \ge v\} \le \exp\Big(\frac{1}{v^2}\Big) - 1\,. \tag{2.125}$$

In (2.124) $\overline{\operatorname{conv}}$ means the closed convex hull. We observe that T' is unconditional in the sense that if $x = (x_i)_{i\ge 1} \in T'$, then $(|x_i|)_{i\ge 1} \in T'$, and vice versa. The name "explicit" in Definition 2.6.5 refers to the fact that we have an explicit description of T' through the family $\mathcal{F}$. This definition is inspired by the result of Theorem 2.1.8, taking into account that we want here unconditionality. (A GB-set U is simply a set such that $g(U) < \infty$, a name that we will not use elsewhere.)

Proposition 2.6.6. *If T is a* EIGB *set then $T \subset B_1$.*

Proof. By the Cauchy-Schwarz inequality we have $V(I,a) \subset B_1$ if $a\sqrt{\operatorname{card} I} \le 1$. □

Proposition 2.6.7. *There exist a number L such that $g(T) \le L$ for each* EIGB *set T.*

The proof relies on a general principle that we spell out now.

Proposition 2.6.8. *Consider subsets T_n of ℓ^2, and assume that $T_n \subset a_n B_2$. Then*

$$g\Big(\bigcup_{n\ge 1} T_n\Big) \le \sup_n \Big(g(T_n) + L a_n \sqrt{\log(n+1)}\Big) + L \sup_n a_n\,. \tag{2.126}$$

This is a generalization of Proposition 2.1.7 which we recover when the sets T_n consist of a single point.

Proof. We can assume each set T_n finite. By Lemma 2.1.3 (the concentration inequality for the supremum of a Gaussian process) we have

$$\mathsf{P}\Big(\sup_{T_n} X_t \geq g(T_n) + u\Big) \leq 2\exp\Big(-\frac{u^2}{2a_n^2}\Big)$$

so that

$$\mathsf{P}\Big(\sup_{T_n} X_t \geq g(T_n) + 2a_n\sqrt{\log(n+1)} + ua_n\Big) \leq 2\exp\Big(-\frac{u^2}{2} - 2\log(n+1)\Big)$$

and thus

$$\mathsf{P}\bigg(\sup_{\bigcup T_n} X_t \geq \sup_n \Big(g(T_n) + 2a_n\sqrt{\log(n+1)}\Big) + u\sup_n a_n\bigg) \leq L\exp\Big(-\frac{u^2}{2}\Big);$$

from which it follows that

$$g\Big(\bigcup_n T_n\Big) \leq \sup_n \Big(g(T_n) + 2a_n\sqrt{\log(n+1)}\Big) + L\sup_n a_n\ .$$

□

Proof of Proposition 2.6.7. We order the pairs (I,a) of $\mathcal{F}$ as a sequence (I_n,a_n) such that the sequence (a_n) is non-increasing. Then (2.125) used for $v = a_n$ implies $n+1 \leq \exp(1/a_n^2)$, i.e. $a_n \leq 1/\sqrt{\log(n+1)}$. Thus (2.126) implies $g\big(\bigcup_n V(I_n,a_n)\big) \leq L$. □

Theorem 2.6.9. *Consider a subset T of ℓ^2. Assume that for a certain number S we have $\gamma_2(T,d_2) \leq S$ and $T \subset SB_1$. Then there exists a* EIGB *set T_1 with $T \subset LST_1$.*

This is a kind of converse of Propositions 2.6.6 and 2.6.7.

Proof. By homogeneity we can assume that $S = 1$. We consider an admissible sequence $(\mathcal{B}_n)$ with

$$\sup_{t\in T}\sum_{n\geq 0} 2^{n/2}\Delta_2(B_n(t)) \leq 2\,, \tag{2.127}$$

and the admissible sequence $(\mathcal{C}_n)$ provided by Theorem 2.6.4 when $\tau = 0$ (we recall that Δ_2 means that the diameter is for the distance in ℓ^2). We consider the increasing sequence of partitions $(\mathcal{A}_n)_{n\geq 0}$ where $\mathcal{A}_n$ is generated by $\mathcal{B}_n$ and $\mathcal{C}_n$, so $\operatorname{card}\mathcal{A}_n \leq N_{n+1}$. The numbers $\ell(t,n)$ of (2.114) depend only on $A_n(t)$.

For every $A \in \mathcal{A}_n$, we pick an arbitrary element $x(A)$ of A, and we set

$$J(A) = \big\{i \in \mathbb{N}^*\ ;\ |x_i(A)| > 2^{-\ell(x(A),n)}\big\}\,,$$

so that $\operatorname{card} J(A) \leq 2^n$ by (2.115). For $n \geq 1$ and $A \in \mathcal{A}_n$, consider the unique element $B \in \mathcal{A}_{n-1}$ such that $A \subset B$, and set

$$I(A) = J(A)\backslash J(B)$$

so that $\operatorname{card} I(A) \le 2^n$ and

$$i \in I(A) \Rightarrow |x_i(B)| \le 2^{-\ell(x(B),n-1)} . \tag{2.128}$$

For $t \in T$, we write $\pi_n(t) = x(A_n(t))$, and, if $n \ge 1$, we write $I_n(t) = I(A_n(t))$. We observe that since $x(A_{n-1}(t)) \in A_{n-1}(t)$ we have $\ell(x(A_{n-1}), n-1) = \ell(t, n-1)$. Thus it follows from (2.128) used for $B = A_{n-1}(t)$ that

$$\|\pi_{n-1}(t)\mathbf{1}_{I_n(t)}\|_\infty \le 2^{-\ell(t,n-1)} , \tag{2.129}$$

and thus, since $\operatorname{card} I_n(t) \le 2^n$

$$\|\pi_{n-1}(t)\mathbf{1}_{I_n(t)}\|_2 \le 2^{n/2-\ell(t,n-1)} .$$

Since $\|t - \pi_{n-1}(t)\|_2 \le \Delta_2(A_{n-1}(t))$ we thus have

$$\|t\mathbf{1}_{I_n(t)}\|_2 \le c(t,n) := \Delta_2(A_{n-1}(t)) + 2^{n/2-\ell(t,n-1)} , \tag{2.130}$$

and hence

$$t\mathbf{1}_{I_n(t)} \in 2^{n/2}c(t,n)V(I_n(t), 2^{-n/2}) .$$

For $n = 0$ we set $I_0(t) = J(T)$, and the previous relation still holds for $c(t,0) = \Delta_2(T) + 1$. We claim now that

$$t = \sum_{n\ge 0} t\mathbf{1}_{I_n(t)} . \tag{2.131}$$

Since the sets $(I_n(t))_{n\ge 0}$ are disjoint, it suffices to show that

$$|t_i| > 0 \Rightarrow i \in \bigcup_{n\ge 0} I_n(t) = \bigcup_{n\ge 0} J(A_n(t)) . \tag{2.132}$$

To prove this, consider i with $|t_i| > 0$ and n large enough that $\Delta_2(A_n(t)) < |t_i|/2$. Then for all $x \in A_n(t)$ we have $|x_i - t_i| \le |t_i|/2$ and hence $|x_i| > |t_i|/2$. Since $\ell(x,n) \ge n-3$ by (2.116), if n is large enough, for all $x \in A_n(t)$ we have $2^{-\ell(x,n)} < |x_i|$. This proves (2.132) and hence (2.131).

Thus we have written $t = \sum_{n\ge 0} t_n$ where

$$t_n = t\mathbf{1}_{I_n(t)} \in b(t,n)V(I_n(t), 2^{-n/2}) , \tag{2.133}$$

for $b(t,n) = 2^{n/2}c(t,n)$. By (2.127) and (2.116) we have $\sum_{n\ge 0} b(t,n) \le L$, so (2.133) implies that

$$t \in L\,\overline{\operatorname{conv}} \bigcup_{\mathcal{F}} V(I,a) .$$

where $\mathcal{F}$ consists of the pairs $(I(A), 2^{-n/2-1})$ for $A \in \mathcal{A}_n$ and $n \ge 0$. Since $\operatorname{card} I(A) \le 2^n$, we have

$$(I,a) \in \mathcal{F} \Rightarrow a\sqrt{\operatorname{card} I} \le 1 \,.$$

Consider $u > 0$, and the largest integer m with $2^{-m/2-1} \ge u$. If $(I,a) \in \mathcal{F}$ and $a \ge u$, then $I = I(A)$ for some $A \in \mathcal{A}_n$, where $n \le m$. Thus there are at most

$$N_0 + \cdots + N_{m+1} \le N_{m+2} \le \exp\Big(\frac{1}{u^2}\Big) - 1$$

choices for A, so that

$$\operatorname{card}\big\{(I,a) \in \mathcal{F}\,;\ a \ge u\big\} \le \exp\Big(\frac{1}{u^2}\Big) - 1\,,$$

Thus $\overline{\operatorname{conv}} \bigcup V(I,a)$ is an EIGB set. □

Theorem 2.6.9 has the following striking consequence.

Theorem 2.6.10. *Consider a Banach space W with an unconditional basis $(e_i)_{i\ge1}$. Assume that $\mathsf{E}\|\sum_{i\ge1} g_i e_i\| = S < \infty$. Then we can find a family $\mathcal{F}$ of pairs (I,a), where I is a finite subset of $\mathbb{N}^*$ and a is a number, with the following properties*

$$\forall x \in W\,,\ x = \sum_{i\ge1} x_i e_i\,,\ \|x\| \le LS \sup_{(I,a)\in\mathcal{F}} a\Big(\sum_{i\in I} x_i^2\Big)^{1/2} \tag{2.134}$$

$$\forall (I,a) \in \mathcal{F}\,,\ a\sqrt{\operatorname{card} I} \le 1 \tag{2.135}$$

$$\forall u > 0\,,\ \operatorname{card}\big\{(I,a) \in \mathcal{F}\,;\ a \ge u\big\} \le \exp\Big(\frac{1}{u^2}\Big) - 1\,. \tag{2.136}$$

Thus, we have $\|x\| \le LS\mathcal{N}(x)$, where $\mathcal{N}(x) = \sup_{(I,a)\in\mathcal{F}} a(\sum_{i\in I} x_i^2)^{1/2}$, and $\mathsf{E}\mathcal{N}(\sum_{i\ge1} g_i e_i) \le L$ (as should be apparent after the following proof, using Proposition 2.6.7). The norms of the type $\mathcal{N}$ are essentially the only unconditional norms for which $\mathsf{E}\mathcal{N}(\sum_{i\ge1} g_i e_i) < \infty$.

Proof. We have

$$\mathsf{E}\Big\|\sum_{i\ge1} g_i e_i\Big\| = g(T)$$

where

$$T = \big\{(x^*(e_i))_{i\ge1}\,,\ x^* \in W^*\,,\ \|x^*\| \le 1\big\}\,.$$

By unconditionality, we have

$$\sup_{\|x^*\|\le1} \sum_{i\ge1} x^*(e_i) g_i = \sup_{\|x^*\|\le1} \sum_{i\ge1} |x^*(e_i)||g_i|$$

so that

$$\mathsf{E} \sup_{\|x^*\|\le1} \sum_{i\ge1} |x^*(e_i)||g_i| = S\,,$$

and, by Jensen's inequality, we get $T \subset LSB_1$. By Theorem 2.1.1 we have $\gamma_2(T,d_2) \le LS$. By Theorem 2.6.9 there exists a family $\mathcal{F}$ that satisfies (2.135), (2.136) and $T \subset LST_1$, where

$$T_1 = \overline{\mathrm{conv}} \bigcup_{\mathcal{F}} V(I,a) \; .$$

Thus, by duality, if $x = \sum_{i\ge 1} x_i e_i \in W$, for any n we have

$$\Big\|\sum_{i\le n} x_i e_i\Big\| \le LS \sup_{t\in T_1} \sum_{i\le n} t_i x_i \le LS \sup_{\mathcal{F}} a\Big(\sum_{i\in I} x_i^2\Big)^{1/2}$$

and this proves (2.134) since $\|x\| = \sup_n \|\sum_{i\le n} x_i e_i\|$. $\square$

For $I \subset \mathbb{N}^*$ and $a > 0$, we write

$$W(I,a) = \big\{x = (x_i)_{i\ge 1} \; ; \; i \notin I \Rightarrow x_i = 0 \, ; \, i \in I \Rightarrow |x_i| \le a\big\} \; .$$

Definition 2.6.11. *We say that a subset T of ℓ^2 is a strong unconditional* GB *set (*SIGB *set) if there exists a family $\mathcal{F}$ of pairs (I,a), where I is a finite subset of* $\mathbb{N}^*$, $a > 0$, $a\,\mathrm{card}\, I \le 1$ *for every* $(I,a) \in \mathcal{F}$ *and*

$$T \subset \overline{\mathrm{conv}} \bigcup_{\mathcal{F}} W(I,a) \tag{2.137}$$

$$\forall\, u > 0\, , \; \mathrm{card}\{(I,a) \in \mathcal{F} \; ; \; a \ge u\} \le \exp\Big(\frac{1}{u}\Big) - 1 \; . \tag{2.138}$$

We have

$$x \in W(I,a) \Rightarrow \sum x_i^2 \le a^2 \mathrm{card}\, I$$

and thus $W(I\,,a) \subset V(I, a\sqrt{\mathrm{card}\, I})$. Moreover, if $a\mathrm{card}\, I \le 1$, we have $a\sqrt{\mathrm{card}\, I} \le \sqrt{a}$. This shows that a SIGB set is an EIGB set.

Theorem 2.6.12. *Consider $T \subset SB_1$, and assume that $\gamma_1(T,d_\infty) \le S$. Then we can find a* SIGB *set T_1 with $T \subset LST_1$.*

Proof. By homogeneity we can assume that $S = 1$. We proceed as in the proof of Theorem 2.6.9, but we can now assume

$$\forall t \in T\, , \sum_{n\ge 0} 2^n \Delta_\infty(A_n(t)) \le 2 \; .$$

Using (2.129) rather than (2.130) we get

$$\|t\mathbf{1}_{I_n(t)}\|_\infty \le c(t,n) := \Delta_\infty(A_n(t)) + 2^{-\ell(t,n)}$$

so that

$$t\mathbf{1}_{I_n(t)} \in 2^n c(t,n) W(I_n(t), 2^{-n})$$

and the proof is finished exactly as before. $\square$

As a consequence, under the conditions of Theorem 2.6.12, we have $\gamma_2(T, d_2) \leq LS$. It is of course of interest to give a more direct (and more general) proof of this fact.

Theorem 2.6.13. *Consider a measure space (Ω, μ) and $T \subset SB_1$, such that $\gamma_1(T, d_\infty) < \infty$. Then*

$$\gamma_2(T, d_2) \leq L\sqrt{S\gamma_1(T, d_\infty)}\,. \tag{2.139}$$

Proof. If we replace μ by $a\mu$, we replace d_1 by ad_1, d_2 by $\sqrt{a}d_2$ and we do not change d_∞. By an appropriate choice of a, we see that it is enough to prove (2.139) when $S = \gamma_1(T, d_\infty)$, and by homogeneity we can assume $S = \gamma_1(T, d_\infty) = 2$. With the notations of the proof of Theorem 2.6.4, let us write

$$a_n(f) = \left\| f \mathbf{1}_{\{|f| \leq u_n(f)\}} \right\|_2 \,.$$

Thus

$$\begin{aligned} a_n(f) &\leq \sum_{r \geq n} \left\| f \mathbf{1}_{\{u_{r+1}(f) \leq |f| \leq u_r(f)\}} \right\|_2 \\ &\leq \sum_{r \geq n} u_r(f) 2^{(r+1)/2} \end{aligned}$$

using (2.117). Thus

$$\sum_{n \geq 0} 2^{n/2} a_n(f) \leq \sqrt{2} \sum_{r \geq n \geq 1} u_r(f) 2^{r/2+n/2} \leq L$$

by first summing over n and using (2.118). Following the proof of Theorem 2.6.4 we then construct an admissible sequence $(\mathcal{C}_n)$ of partitions of B_1 and for $C \in \mathcal{C}_n$ a number $\ell(C)$ such that

$$f \in C \Rightarrow a_n(f) \leq 2^{-\ell(C)}$$

$$\forall f \in B_1\,, \sum_{n \geq 0} 2^{n/2 - \ell(f,n)} \leq L \tag{2.140}$$

where $\ell(f, n) = \ell(C_n(f))$ depends only on $C_n(f)$.

Proceeding as in the proof of Theorem 2.6.9 we find an increasing sequence of $(\mathcal{A}_n)$ of partitions of T such that $\operatorname{card} \mathcal{A}_n \leq N_{n+1}$,

$$\sup_{t \in T} \sum_{n \geq 0} 2^n \Delta_\infty(A_n(t)) \leq 2\gamma_1(T, d_\infty)\,, \tag{2.141}$$

and, moreover, the numbers $\ell(f, n)$ depend only on $A_n(f)$.

Consider f, $g \in A_n(t)$, and set $\Delta = \Delta_\infty(A_n(t))$, so that $\|f - g\|_\infty \leq \Delta$. Thus

$$\|f-g\|_2 \le \|\min(|f-g|,\Delta)\|_2 \tag{2.142}$$
$$\le \|\min(|f|,\Delta)\|_2 + \|\min(|g|,\Delta)\|_2 \,.$$

Now

$$\min(|f|,\Delta) \le \Delta \mathbf{1}_{\{|f|\ge u_n(f)\}} + |f|\mathbf{1}_{\{|f|\le u_n(f)\}}$$

and, using (2.117),

$$\|\min(|f|,\Delta)\|_2 \le 2^{n/2}\Delta + a_n(f)$$

and thus, combining with (2.142),

$$\Delta_2(A_n(t)) \le 2\big(2^{n/2}\Delta_\infty(A_n(t)) + 2^{-\ell(t,n)}\big) \,.$$

Combining with (2.140) and (2.141) yields

$$\sum_{n\ge 0} 2^{n/2}\Delta_2(A_n(t)) \le L \,.$$

Since $\Delta_2(T) \le L$, appealing to Lemma 1.3.3 finishes the proof. □

2.7 Donsker Classes

Throughout this section we consider a probability space (Ω,μ), and a bounded subset $\mathcal{F}$ of $\mathcal{L}^2(\mu)$, which, following the standard notation, we will denote by $\mathcal{F}$ rather than T. To avoid (well understood) measurability problems, we assume that $\mathcal{F}$ is countable. (Thus, there is no need to really distinguish between $\mathcal{L}^2(\mu)$ and $L^2(\mu)$.)

Consider i.i.d r.v. $(X_i)_{i\ge 1}$ valued in Ω, of law μ. We set $\mu(f) = \int f \mathrm{d}\mu$ and

$$S_N(\mathcal{F}) = E\sup_{f\in\mathcal{F}}\Big|\frac{1}{\sqrt{N}}\sum_{i\le N}(f(X_i)-\mu(f))\Big| \tag{2.143}$$

$$S(\mathcal{F}) = \sup_N S_N(\mathcal{F}) \,. \tag{2.144}$$

The question of understanding when $S(\mathcal{F}) < \infty$ is central to the study of Donsker classes, which are classes of functions on which the central limit theorem holds uniformly. The precise definition of Donsker classes includes a number of technicalities that are not related to the topic of this book, and since there exists an abundant (and most excellent) literature on this topic, (see e.g. [8]) we will not deal with them, but will concentrate on the study of upper bounds for $S(\mathcal{F})$ and, more generally, for $S_N(\mathcal{F})$ at a given value of N. To avoid trivial complications we will often assume $\mu(f) = \int f\mathrm{d}\mu = 0$ for each f in $\mathcal{F}$.

The following classical result will play a fundamental role.

Lemma 2.7.1. *(Bernstein's inequality). Consider independent r.v. $(Y_i)_{i\geq 1}$ with $\mathsf{E}Y_i = 0$ and a number U with $|Y_i| \leq U$ for each i. Then, for $v > 0$,*

$$\mathsf{P}\Big(|\sum_{i\geq 1} Y_i| \geq v\Big) \leq 2\exp\Big(-\min\Big(\frac{v^2}{4\sum_{i\geq 1} EY_i^2}, \frac{v}{2U}\Big)\Big)\,.$$

Proof. For $|x| \leq 1$, we have

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

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

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

so that $\mathsf{E}\exp\lambda Y_i \leq 1 + \lambda^2\mathsf{E}Y_i^2 \leq \exp\lambda^2\mathsf{E}Y_i^2$, and thus

$$\mathsf{E}\exp\lambda\sum_{i\geq 1} Y_i = \prod_{i\geq 1}\mathsf{E}\exp\lambda Y_i \leq \exp\lambda^2\sum_{i\geq 1}\mathsf{E}Y_i^2\,.$$

Now

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

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

$$\frac{1}{U^2}\sum_{i\geq 1}\mathsf{E}Y_i^2 - \frac{v}{U} \leq \frac{Uv}{2U^2} - \frac{v}{U} \leq -\frac{v}{2U}\,.$$

□

Proposition 2.7.2. *If $0 \in \mathcal{F}$ we have*

$$S_N(\mathcal{F}) \leq L\Big(\gamma_2(\mathcal{F}, d_2) + \frac{1}{\sqrt{N}}\gamma_1(\mathcal{F}, d_\infty)\Big)\,, \tag{2.145}$$

where d_2 and d_∞ are the distances induced by the norms of L^2 and L^∞ respectively.

Proof. We combine Bernstein's inequality with Theorem 1.2.7 to get, since $0 \in \mathcal{F}$, that

$$\begin{aligned}\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\leq N} f(X_i)\Big| &\leq \mathsf{E}\sup_{f,f'\in\mathcal{F}}\Big|\sum_{i\leq N} f(X_i) - f'(X_i)\Big| \\ &\leq L\big(\gamma_2(\mathcal{F}, 2\sqrt{N}d_2) + \gamma_1(\mathcal{F}, 2d_\infty)\big)\,.\end{aligned}$$

To conclude, we use that $\gamma_2(\mathcal{F}, 2\sqrt{N}d_2) = 2\sqrt{N}\gamma_2(\mathcal{F}, d_2)$ and $\gamma_1(\mathcal{F}, 2d_\infty) = 2\gamma_1(\mathcal{F}, d_\infty)$. □

Proposition 2.7.2 provides us with a method to bound $S_N(\mathcal{F})$ from above. There is however a completely different method, namely the inequality

$$S_N(\mathcal{F}) \le \mathsf{E} \sup_{f\in\mathcal{F}} \frac{1}{\sqrt{N}} \sum_{i\le N} |f(X_i)| \,. \tag{2.146}$$

One should point out that the very idea of the central limit theorem is that there is cancelation between terms of opposite signs, while (2.146), where there is no such cancelation, is of a different nature.

One can of course combine (2.145) and (2.146) to control $S_N(\mathcal{F})$.

Proposition 2.7.3. *Consider classes $\mathcal{F}, \mathcal{F}_1$ and $\mathcal{F}_2$ of functions in $L^2(\mu)$ with $\mu(f) = 0$ for $f \in \mathcal{F}_1$, $f \in \mathcal{F}_2$, and assume that $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$. Assume that $0 \in \mathcal{F}_1$. Then*

$$\begin{aligned} S_N(\mathcal{F}) = \mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\le N} f(X_i) \Big| &\le L\Big(\gamma_2(\mathcal{F}_1, d_2) + \frac{1}{\sqrt{N}} \gamma_1(\mathcal{F}_1, d_\infty)\Big) \\ &\quad + \mathsf{E} \sup_{f\in\mathcal{F}_2} \frac{1}{\sqrt{N}} \sum_{i\le N} |f(X_i)| \,. \end{aligned}$$

The following question seems related to the Bernoulli problem of Chapter 4.

Research problem 2.7.4. Consider a class $\mathcal{F}$ of functions in $L^2(\mu)$ with $\mu(f) = 0$ for $f \in \mathcal{F}$. *Given an integer N*, can we find a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ with $0 \in \mathcal{F}_1$ such that the following properties hold:

$$\gamma_2(\mathcal{F}_1, d_2) \le L S_N(\mathcal{F})$$

$$\gamma_1(\mathcal{F}_1, d_\infty) \le L\sqrt{N} S_N(\mathcal{F})$$

$$\mathsf{E} \sup_{f\in\mathcal{F}_2} \frac{1}{\sqrt{N}} \sum_{i\le N} |f(X_i)| \le L S_N(\mathcal{F})\,?$$

A positive answer to this problem would mean that there is essentially no other method to control $S_N(\mathcal{F})$ from above than the method of Proposition 2.7.3.

The main result of this section is a kind of partial answer to Research Problem 2.7.4.

Theorem 2.7.5. *Consider a class $\mathcal{F}$ of functions in $\mathcal{L}^2(\mu)$, with $\mu(f) = 0$ for $f \in \mathcal{F}$. Then we can find a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ where $0 \in \mathcal{F}_1$,*

$$\gamma_2(\mathcal{F}_1, d_2) \le L\gamma_2(\mathcal{F}) \tag{2.147}$$

$$\gamma_1(\mathcal{F}_1, d_\infty) \le L\sqrt{N}\gamma_2(\mathcal{F}) \tag{2.148}$$

$$\mathsf{E} \sup_{f\in\mathcal{F}_2} \frac{1}{\sqrt{N}} \sum_{i\le N} |f(X_i)| \le L(S_N(\mathcal{F}) + \gamma_2(\mathcal{F}))\,. \tag{2.149}$$

To understand the link between this theorem and Research Problem 2.7.4 we prove the following easy fact.

Lemma 2.7.6. *If $\mu(f)=0$ for each f in $\mathcal{F}$, we have $\gamma_2(\mathcal{F}) \le LS(\mathcal{F})$.*

Of course here $\mathcal{F}$ is viewed as a subset of $L^2(\mu)$, with the corresponding distance.

Proof. Consider a finite subset T of $\mathcal{F}$. By the ordinary central limit theorem, the joint law of $(N^{-1/2}\sum_{i\le N} f(X_i))_{f\in T}$ converges to the law of a Gaussian process $(g_f)_{f\in T}$ and thus

$$\mathsf{E}\sup_{f\in T} g_f \le S(\mathcal{F}) . \tag{2.150}$$

The construction of the process $(g_f)_{f\in T}$ shows that for $f_1, f_2 \in T$ we have $\mathsf{E} g_{f_1} g_{f_2} = \int f_1 f_2 \mathrm{d}\mu$. If we identify $L^2(\mu)$ with $\ell^2(\mathbb{N}^*)$, and since the law of a Gaussian process is determined by its covariance, the left-hand side of (2.150) is exactly $g(T)$. This shows that $g(T) \le S(\mathcal{F})$, and the result follows by Theorem 2.1.1 and (1.51). $\square$

As a consequence, we have the following characterization of classes for which $S(\mathcal{F}) < \infty$.

Theorem 2.7.7. *Consider a class of functions $\mathcal{F}$ of $L^2(\mu)$ and assume that $\mu(f)=0$ for each $f\in\mathcal{F}$. Then we have $S(\mathcal{F})<\infty$ if and only if there exists a number A and for each N there exists a decomposition $\mathcal{F}\subset\mathcal{F}_1+\mathcal{F}_2$ (depending on N) where $0\in\mathcal{F}_1$ such that*

$$\gamma_2(\mathcal{F}_1, d_2) \le A$$

$$\gamma_1(\mathcal{F}_1, d_\infty) \le \sqrt{N} A$$

$$\mathsf{E}\sup_{f\in\mathcal{F}_2} \frac{1}{\sqrt{N}} \sum_{i\le N} |f(X_i)| \le A.$$

Proof of Theorem 2.7.5. We use the decomposition of Theorem 2.6.2 with $u=\sqrt{N}$. This produces a decomposition $\mathcal{F}\subset\mathcal{F}_1+\mathcal{F}_2$, where $\mathcal{F}_1$ satisfies (2.147) and (2.148), while $\mathcal{F}_2 \subset L\gamma_2(\mathcal{F})B_1/\sqrt{N}$. Moreover the construction is such that $\mathcal{F}_2\subset\mathcal{F}-\mathcal{F}_1$, and so that by (2.145) we have

$$\mathsf{E}\sup_{f\in\mathcal{F}_2}\Big|\sum_{i\le N} f(X_i)\Big| \le (S+L\gamma_2(\mathcal{F}))\sqrt{N} .$$

Then (2.149) follows from the next result. $\square$

Theorem 2.7.8. *(The Giné-Zinn Theorem) For a class $\mathcal{F}$ of functions with $\mu(f)=0$ for f in $\mathcal{F}$ we have*

$$\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\le N}|f(X_i)| \le N\sup_{f\in\mathcal{F}}\int|f|\mathrm{d}\mu + 4\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\le N} f(X_i)\Big| . \tag{2.151}$$

While simple, this is very useful. In order to avoid repetition, we will prove some more general facts. We consider pairs (R_i, X_i) of r.v., with $X_i \in \Omega$, $R_i \geq 0$, and we assume that these pairs are independent. We consider a Bernoulli sequence $(\epsilon_i)_{i\geq 1}$, that is an i.i.d. sequence with $\mathsf{P}(\epsilon_i = \pm 1) = 1/2$. We assume that these sequences are independent of the r.v. (R_i, X_i). We assume that for each ω, only finitely many of the r.v. $R_i(\omega)$ are not zero.

Lemma 2.7.9. *For a countable class of functions $\mathcal{F}$ we have*

$$\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} \big(R_i f(X_i) - \mathsf{E}(R_i f(X_i))\big)\Big| \leq 2\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} \epsilon_i R_i f(X_i)\Big| \tag{2.152}$$

$$\begin{aligned}\mathsf{E} \sup_{f\in\mathcal{F}} \sum_{i\geq 1} R_i |f(X_i)| \leq \sup_{f\in\mathcal{F}} \sum_{i\geq 1} \mathsf{E}(R_i |f(X_i)|) \\ + 2\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} \epsilon_i R_i f(X_i)\Big| \,.\end{aligned} \tag{2.153}$$

If $\mathsf{E}(R_i f(X_i)) = 0$ for each $i \geq 1$, then

$$\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} \epsilon_i R_i f(X_i)\Big| \leq 2\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} R_i f(X_i)\Big| \,. \tag{2.154}$$

Proof of Theorem 2.7.8. We take $R_i = 1$ if $i \leq N$ and $R_i = 0$ if $i \geq N$, and we combine (2.153) and (2.154). □

Proof of Lemma 2.7.9. Consider an independent copy $(S_i, Y_i)_{i\geq 1}$ of the sequence $(R_i, X_i)_{i\geq 1}$, that is independent of the sequence $(\epsilon_i)_{i\geq 1}$. Then, by Jensen's inequality,

$$\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} \big(R_i f(X_i) - \mathsf{E}(R_i f(X_i))\big)\Big| \leq \mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} (R_i f(X_i) - S_i f(Y_i))\Big| \,.$$

Since the sequences $(R_i f(X_i) - S_i f(Y_i))$ and $(\epsilon_i (R_i f(X_i) - S_i f(Y_i)))$ of r.v. have the same law, we have

$$\begin{aligned}\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} (R_i f(X_i) - S_i f(Y_i))\Big| &= \mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} \epsilon_i (R_i f(X_i) - S_i f(Y_i))\Big| \\ &\leq 2\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\geq 1} \epsilon_i R_i f(X_i)\Big|\end{aligned}$$

and we have proved (2.152). To prove (2.153), we write

$$\begin{aligned}\sum_{i\geq 1} R_i |f(X_i)| \leq \sum_{i\geq 1} \mathsf{E}(R_i |f(X_i)|) \\ + \sum_{i\geq 1} \big(R_i |f(X_i)| - \mathsf{E}(R_i |f(X_i)|)\big) \,,\end{aligned}$$

we take the supremum over f and expectation, and we use (2.152) to get

$$\begin{aligned}\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\geq 1}R_i|f(X_i)| \leq \sup_{f\in\mathcal{F}}\sum_{i\geq 1}\mathsf{E}(R_i|f(X_i)|) \\ + 2\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\epsilon_i R_i|f(X_i)|\Big| .\end{aligned}$$

We then conclude with the comparison theorem for Bernoulli processes ([53], Theorem 2.1) that implies that

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\epsilon_i R_i|f(X_i)|\Big| \leq \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\epsilon_i R_i f(X_i)\Big| .$$

To prove (2.154), we work conditionally on the sequence $(\epsilon_i)_{i\geq 1}$. Setting $I = \{i \geq 1\ ;\ \epsilon_i = 1\}$ and $J = \{i \geq 1\ ;\ \epsilon_i = -1\}$, we have

$$\begin{aligned}\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\leq N}\epsilon_i R_i f(X_i)\Big| \leq \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in I}R_i f(X_i)\Big| \\ + \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in J}R_i f(X_i)\Big| .\end{aligned}$$

Now, by Jensen's inequality, we have

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in I}R_i f(X_i)\Big| \leq \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}R_i f(X_i)\Big|$$

since $\mathsf{E}R_i f(X_i) = 0$. □

The following is a very powerful practical method to control $S(\mathcal{F})$.

Theorem 2.7.10. *(Ossiander's bracketing theorem) Consider a countable class $\mathcal{F}$ of functions in $\mathcal{L}^2(\mu)$. Consider an admissible sequence $\mathcal{A}_n$ of partitions of $\mathcal{F}$. For $A \in \mathcal{A}_n$, define the function h_A by*

$$h_A(\omega) = \sup_{f,f'\in A}|f(\omega) - f'(\omega)| . \tag{2.155}$$

Assume that for a number S we have

$$\sup_{t\in\mathcal{F}}\sum_{n\geq 0}2^{n/2}\|h_{A_n(t)}\|_2 \leq S . \tag{2.156}$$

Then

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\frac{1}{\sqrt{N}}\sum_{i\leq N}(f(X_i) - \mu(f))\Big| \leq LS . \tag{2.157}$$

Strictly speaking Ossiander [31] proved this result only under entropy conditions, but once one understands the principles of the generic chaining it is immediate to adapt her proof to the present setting. The reader will observe that $\Delta(A) \le \|h_A\|_2$ for all A, so that (2.156) implies that $\gamma_2(\mathcal{F}, d_2) \le 2$. This alone is however not sufficient to prove (2.157).

Theorem 2.7.11. *Consider a countable set $T \subset \mathcal{L}^0(\Omega, \mu)$, where μ is a positive measure (that need not be a probability). Consider a number $u > 0$. Consider an admissible sequence $(\mathcal{A}_n)$ of partitions of T and for $A \in \mathcal{A}_n$ define h_A by (2.155). Assume that*

$$\forall t \in T\ ,\ \sum_{n \ge 0} 2^{n/2} \|h_{A_n(t)}\|_2 \le S\ . \tag{2.158}$$

Then we can find two sets $T_1\ ,\ T_2 \subset \mathcal{L}^0(\Omega, \mu)$ with the following properties:

$$\gamma_2(T_1, d_2) \le LS\ ,\ \gamma_1(T_1, d_\infty) \le LuS \tag{2.159}$$

$$\gamma_2(T_2, d_2) \le LS\ ,\ \gamma_1(T_2, d_\infty) \le LuS \tag{2.160}$$

$$s \in T_2 \Rightarrow s \ge 0\ ,\ \|s\|_1 \le LS/u\,, \tag{2.161}$$

in such a way that

$$T \subset T_1 + T_2',\ \textit{where}\ T_2' = \{t\ ;\ \exists s \in T_2\ ,\ |t| \le s\}\ . \tag{2.162}$$

This theorem is related to Theorem 2.6.2, and the reader should be familiar with the proof of that previous result before reading the following argument.

Proof. For $n \ge 0$ and $t \in T$ we define

$$\Omega'(t,n) = \left\{ h_{A_n(t)} \le 2^{-n/2} u \|h_{A_n(t)}\|_2 \right\}\ ,$$

so that by Markov's inequality we have

$$\mathsf{P}(\Omega'(t,n)^c) \le \frac{2^n}{u^2}\ .$$

We define

$$\Omega(t,n) = \bigcap_{0 \le \ell \le n} \Omega'(t,\ell)$$

so that

$$\mathsf{P}(\Omega(t,n)^c) \le \frac{2^{n+1}}{u^2}\ . \tag{2.163}$$

We construct the points t_A and $\pi_n(t)$ as in the proof of Theorem 2.6.2, and for $n \ge 1$ we define

$$f^1_{t,n} = (\pi_n(t) - \pi_{n-1}(t)) \mathbf{1}_{\Omega(t,n-1)}\ ,$$

so that, since $|\pi_n(t) - \pi_{n-1}(t)| \leq h_{A_{n-1}(t)}$, and since $|h_{A_{n-1}}(t)(\omega)| \leq 2^{-(n-1)/2}u\|h_{A_{n-1}(t)}\|_2$ for $\omega \in \Omega(t, n-1) \subset \Omega'(t, n-1)$, we have

$$\|f^1_{t,n}\|_2 \leq \|h_{A_{n-1}(t)}\|_2 \ , \ \|f^1_{t,n}\|_\infty \leq 2^{-(n-1)/2}u\|h_{A_{n-1}(t)}\|_2 \ .$$

We set $g^1_{t,0} = t_T$ and for $n \geq 1$ we set $g^1_{t,n} = t_T + \sum_{1\leq k\leq n} f^1_{t,k}$. Moreover we define

$$T^1_n = \{g^1_{t,m} \ ; \ m \leq n \ , \ t \in T\} \ ; \ T_1 = \bigcup_{n\geq 0} T^1_n \ ,$$

and we prove (2.159) as in the proof of Theorem 2.6.2. Let us define $f^2_{t,0} = 0$ and for $n \geq 1$

$$f^2_{t,n} = 2^{-(n-1)/2}u\|h_{A_{n-1}(t)}\|_2 \mathbf{1}_{\Omega(t,n)^c} \ ,$$

so that using (2.163) we see that

$$\begin{aligned}
\|f^2_{t,n}\|_2 &\leq \|h_{A_{n-1}(t)}\|_2 \\
\|f^2_{t,n}\|_\infty &\leq L2^{-n/2}u\|h_{A_{n-1}(t)}\|_2 \\
\|f^2_{t,n}\|_1 &\leq \frac{L2^{n/2}}{u}\|h_{A_{n-1}(t)}\|_2 \ .
\end{aligned}$$

Let us define

$$w = h_T \mathbf{1}_{\{h_T \geq u\|h_T\|_2\}} \ ,$$

so that $\|w\|_1 \leq \|h_T\|_2/u$. Let us further define $g^2_{t,0} = w$ and for $n \geq 1$

$$g^2_{t,n} = w + \sum_{1\leq k\leq n} f^2_{t,k} \ .$$

Define finally

$$T^2_n = \{g^2_{t,m} \ ; \ m \leq n \ , \ t \in T\} \ ; \ T_2 = \bigcup_{n\geq 0} T^2_n \ .$$

It should be obvious from (2.158) that (2.161) holds, and that (2.160) can be proved as (2.159). We turn to the proof of (2.162). Any t in T is of the type $t = t_A$ for some m and $A = A_m(t)$, so that $t = t_T + \sum_{1\leq k\leq m}(\pi_k(t) - \pi_{k-1}(t))$. Let $t^2 = t - g^1_{t,m}$, so $t = g^1_{t,m} + t^2$. Since $g^1_{t,m} \in T_1$ is suffices to show that $t^2 \in T'_2$, and to show this, we show that $|t^2| = |t - g^1_{t,m}| \leq g^2_{t,m} \in T_2$. Since $t = t_T + \sum_{1\leq k\leq m}(\pi_k(t) - \pi_{k-1}(t))$, and since $f^1_{t,k} = \pi_k(t)(\omega) - \pi_{k-1}(t)(\omega)$ for $\omega \in \Omega(t, k-1)$, the definition of $g^1_{t,m}$ shows that if $t^2(\omega) \neq 0$, then for some $1 \leq k \leq m$ we have $\omega \notin \Omega(t, k-1)$. Consider the smallest such number k. If $k = 1$, we have $\omega \notin \Omega(t, p)$ for $p \geq 0$, and thus $g^1_{t,m}(\omega) = t_T(\omega)$ so that $t^2(\omega) = t(\omega) - t_T(\omega)$. Since $\omega \in \Omega(t, 0)^c = \Omega'(t, 0)^c = \{h_T > u\|h_T\|_2\}$, we have $|t^2(\omega)| = |t(\omega) - t_T(\omega)| \leq h_T(\omega) = w(\omega) \leq g^2_{t,m}(\omega)$. If $k > 1$, since k is the smallest possible, and since the sequence of sets $\Omega(t, n)$ decreases as n

increases, we have $\omega \in \Omega(t,\ell-1)$ for $\ell < k$ and $\omega \notin \Omega(t,\ell-1)$ for $\ell \geq k$ so that

$$g^1_{t,m}(\omega) = t_T(\omega) + \sum_{1\leq \ell\leq k-1} (\pi_\ell(t)(\omega) - \pi_{\ell-1}(t)(\omega)) = \pi_{k-1}(t)(\omega)$$

and

$$\begin{aligned}|t^2(\omega)| = |t(\omega) - g^1_{t,m}(\omega)| &= |t(\omega) - \pi_{k-1}(t)(\omega)| \\ &\leq h_{A_{k-2}(t)}(\omega)\mathbf{1}_{\Omega(t,k-1)^c}(\omega) \\ &\leq f^2_{t,k-1}(\omega) \leq g^2_{t,m}(\omega)\ ,\end{aligned}$$

where the first inequality uses that $|t - \pi_{k-1}(t)| \leq h_{A_{k-2}}(t)$ and $\omega \in \Omega(t,k-1)^c$, and the second that, since $\omega \in \Omega(t,k-2) \subset \Omega'(t,k-2)$, we have $h_{A_{k-2}(t)}(\omega) \leq 2^{-(k-2)/2}u\|h_{A_{k-2}(t)}\|_2$. □

Proof of Theorem 2.7.10. We apply Theorem 2.7.12 with $u = \sqrt{N}$, $T = \mathcal{F}$. The decomposition $\mathcal{F} \subset T_1 + T_2'$ shows that it suffices to prove (2.157) when $\mathcal{F}$ is one of the classes T_1 or T_2'. When $\mathcal{F} = T_1$, this follows from Bernstein's inequality and Theorem 1.2.7, since by (2.159) we have

$$\gamma_2(T_1,d_2) \leq LS\,;\ \gamma_1(T_1,d_\infty) \leq LS\sqrt{N}.$$

When $\mathcal{F} = T_2'$, we write

$$\begin{aligned}\sup_{f\in T_2'}\Big|\frac{1}{\sqrt{N}}\sum_{i\leq N}(f(X_i)-\mu(f))\Big| &\leq \sup_{f\in T_2'}\Big(\sqrt{N}|\mu(f)| + \frac{1}{\sqrt{N}}\sum_{i\leq N}|f(X_i)|\Big) \\ &\leq \sup_{f\in T_2}\Big(\sqrt{N}\mu(f) + \frac{1}{\sqrt{N}}\sum_{i\leq N}f(X_i)\Big) \\ &\leq \sup_{f\in T_2}\Big(2\sqrt{N}\mu(f) + \frac{1}{\sqrt{N}}\sum_{i\leq N}(f(X_i)-\mu(f))\Big)\,.\end{aligned}$$

By (2.161) we have $\mu(f) = \|f\|_1 \leq LS/\sqrt{N}$ for $f \in T_2$. Thus it suffices to show that

$$\mathsf{E}\sup_{f\in T_2}\Big|\frac{1}{\sqrt{N}}\sum_{i\leq N}(f(X_i)-\mu(f))\Big| \leq LS\ .$$

This follows as in the case of T_1 from the fact that $\gamma_2(T_2,d_2) \leq LS$ and $\gamma_1(T_2,d_\infty) \leq LS\sqrt{N}$ by (2.160). □

To cover further needs, we will also prove a general principle that is to Theorem 2.7.11 what Theorem 2.6.3 is to Theorem 2.6.2. This result will be used only in chapter 5, and its proof is better omitted at first reading.

Theorem 2.7.12. *Consider a countable set $T \subset \mathcal{L}^0(\Omega,\mu)$, where μ is a positive measure (that need not be a probability). Assume that $0 \in T$, and* *consider a number $V \geq 2$. Consider an admissible sequence $(\mathcal{A}_n)$ of partitions of T and for $A \in \mathcal{A}_n$ define h_A by (2.155). For $A \in \mathcal{A}_n$ consider $j(A) \in \mathbb{Z} \cup \{\infty\}$ and $\delta(A) \in \mathbb{R}^+$. Assume the following properties*

$$\forall t \in T\,,\ \lim_{n\to\infty} j(A_n(t)) = \infty \tag{2.164}$$

$$A \in \mathcal{A}_n\,,\ B \in \mathcal{A}_{n-1}\,,\ A \subset B \Rightarrow j(A) \geq j(B) \tag{2.165}$$

$$A \in \mathcal{A}_n\,,\ B \in \mathcal{A}_{n'}\,,\ A \subset B\,,\ j(A) = j(B) \Rightarrow \delta(B) \leq 2\delta(A) \tag{2.166}$$

$$\forall A \in \mathcal{A}_n\,,\ \int h_A^2 \wedge V^{-2j(A)} \mathrm{d}\mu \leq \delta^2(A)\,. \tag{2.167}$$

Then we can find two sets $T_1\,,\ T_2 \subset \mathcal{L}^0(\Omega, \mu)$ with the following properties, where $j_n(t) = j(A_n(t))$

$$\gamma_2(T_1, d_2) \leq L \sup_{t\in T} \sum_{n\geq 0} 2^{n/2}\delta(A_n(t)) \tag{2.168}$$

$$\gamma_1(T_1, d_\infty) \leq L \sup_{t\in T} \sum_{n\geq 0} 2^n V^{-j_n(t)} \tag{2.169}$$

$$\gamma_2(T_2, d_2) \leq L \sup_{t\in T} \sum_{n\geq 1} 2^{n/2} V^{j_n(t)-j_{n-1}(t)}\delta(A_n(t)) \tag{2.170}$$

$$\gamma_1(T_2, d_\infty) \leq L \sup_{t\in T} \sum_{n\geq 0} 2^n V^{-j_n(t)} \tag{2.171}$$

$$s \in T_2 \Rightarrow s \geq 0\,,\ \|s\|_1 \leq L \sup_{t\in T} \sum_{n\geq 1} V^{2j_n(t)-j_{n-1}(t)}\delta^2(A_n(t))\,, \tag{2.172}$$

in such a way that $T \subset T_1 + T_2' + T_3$, where

$$\begin{aligned} T_2' &= \{t\,;\ \exists s \in T_2\,,\ |t| \leq s\} \\ T_3 &= \left\{t\,;\ |t| \leq h_T \mathbf{1}_{\{h_T \geq V^{-j(T)}\}}\right\}. \end{aligned}$$

Second proof of Theorem 2.7.10. We set $\delta(A) = \|h_A\|_2$, and we define $j(A)$ as the largest element of $\mathbb{Z} \cup \{\infty\}$ for which

$$2^{-j(A)} \geq \sqrt{N} 2^{-n/2}\delta(A)\,,$$

so that (2.164) to (2.167) obviously hold true. We note that by definition of $j(A)$ we have

$$2^{-j(A)} \leq \sqrt{N} 2^{-n/2+1}\delta(A)\,,$$

and hence we have

$$2^{j_n(t)}\delta(A_n(t))\sqrt{N} \leq 2^{n/2}\ ;\ 2^{-j_{n-1}(t)} \leq \sqrt{N} 2^{-n/2+3/2}\delta(A_{n-1}(t))\,. \tag{2.173}$$

We apply Theorem 2.7.12 with $V = 2$, $T = \mathcal{F}$. The decomposition $\mathcal{F} \subset T_1 + T_2' + T_3$ shows that it suffices to prove (2.157) when $\mathcal{F}$ is one of the classes T_1, T_2', T_3. When $\mathcal{F} = T_1$, this follows from Bernstein's inequality and Theorem 1.2.7, since by (2.168) and (2.169) we have

$$\gamma_2(T_1, d_2) \le LS\,;\ \gamma_1(T_1, d_\infty) \le LS\sqrt{N}.$$

When $\mathcal{F} = T_2'$, we proceed as in the first proof of Theorem 2.7.10, since by (2.172) and (2.173) we have $\mu(f) = \|f\|_1 \le LS/\sqrt{N}$ for $f \in T_2$ and since $\gamma_2(T_2, d_2) \le LS$ and $\gamma_1(T_2, d_\infty) \le LS\sqrt{N}$ by (2.170) and (2.171), using (2.173) again. The case $\mathcal{F} = T_3$ is very simple after one notices that $\|h_T \mathbf{1}_{\{h_T \ge 2^{-j(T)}\}}\|_1 \le LS/\sqrt{N}$, and is left to the reader. $\square$

Proof of Theorem 2.7.12. This result will be used only in chapter 5, and this proof is better omitted at first reading. We define

$$\begin{aligned} p(t,n) &= \inf\{p \ge 0\,;\ j_p(t) = j_n(t)\} \\ A(t,n) &= A_{p(t,n)}(t) \end{aligned}$$

and we observe that by (2.166) we have

$$\delta(A(t,n)) \le 2\delta(A_n(t))\ .$$

For $n \ge 1$ and $A \in \mathcal{A}_n$ we choose $t_A \in A$ arbitrary. We choose $t_T = 0$. We define $\pi_n(t) = t_{A(t,n)}$. We note that

$$\pi_{n+1}(t) \ne \pi_n(t) \Rightarrow A(t, n+1) \ne A(t,n) \Rightarrow j_{n+1}(t) > j_n(t)\ . \tag{2.174}$$

We define

$$\begin{aligned} \Omega_{t,n} &= \{h_{A(t,n)} > V^{-j_n(t)}\} \\ \Omega &= \{h_T > V^{-j(T)}\} \\ m(t,\omega) &= \inf\{n \ge 0\,;\ \omega \in \Omega_{t,n}\} \end{aligned}$$

if the set on the right is not empty, and $m(t,\omega) = \infty$ otherwise. We note that

$$n < m(t,\omega) \Rightarrow |\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| \le h_{A(t,n)}(\omega) \le V^{-j_n(t)}\ . \tag{2.175}$$

When $m(t,\omega) < \infty$ we define

$$t^1(\omega) = \pi_{m(t,\omega)}(t)(\omega)\mathbf{1}_{\Omega^c}(\omega)\,,$$

and if $m(t,\omega) = \infty$ we define

$$t^1(\omega) = t(\omega)\mathbf{1}_{\Omega^c}(\omega) = \lim_{n\to\infty} \pi_n(t)(\omega)\mathbf{1}_{\Omega^c}(\omega)\ .$$

The second equality follows from (2.164) and the fact that

$$n < m(t,\omega) \Rightarrow |t(\omega) - \pi_n(t)(\omega)| \le h_{A(t,n)}(\omega) \le V^{-j_n(t)} .$$

We set $T_1 = \{t^1 \ ; \ t \in T\}$. The proof of (2.168) and (2.169) is then the same as in the case of Theorem 2.6.3. We set

$$t'^2 = (t - t^1)\mathbf{1}_{\Omega^c} = \sum_{m\ge 0}(t - \pi_m(t))\mathbf{1}_{\{m(t,\cdot)=m\}\setminus\Omega} . \tag{2.176}$$

We have $\{m(t,\cdot) = m\} \subset \Omega_{t,m}$, and, by (2.167),

$$\mu(\Omega_{t,m}) \le V^{2j_m(t)}\delta^2(A_m(t)) . \tag{2.177}$$

For $m \ge 1$, we observe that the set $\{m(t,\cdot) = m\}$ is empty when $j_m(t) = j_{m-1}(t)$, because then $\Omega_{t,m} = \Omega_{t,m-1}$.

Also, if $m(t,\omega) = m$ then $\omega \notin \Omega_{t,m-1}$, and thus

$$|t(\omega) - \pi_m(t)(\omega)| \le h_{A(t,m-1)}(\omega) \le V^{-j_{m-1}(t)} .$$

Thus, for $m \ge 1$

$$(t - \pi_m(t))\mathbf{1}_{\{m(t,\cdot)=m\}\setminus\Omega} \le u(t,m) ,$$

where

$$u(t,m) = V^{-j_{m-1}(t)}\mathbf{1}_{\Omega_{t,m}\setminus\Omega} \tag{2.178}$$

if $j_m(t) \ne j_{m-1}(t)$ and $u(t,m) = 0$ otherwise. We observe that for every $t \in T$ we have $\Omega = \{m(t,\cdot) = 0\}$, so that in (2.176) the term of the summation corresponding to $m = 0$ is 0, and thus

$$|t'^2| \le \sum_{m\ge 1} u(t,m) . \tag{2.179}$$

We set $T_2 = \{t^2 := \sum_{m\ge 1} u(t,m) \ ; \ t \in T\}$.

We observe that $u(t,m)$ depends only on $A_m(t)$. Thus if

$$U_n = \Big\{ \sum_{1\le m\le n} u(t,m) \ ; \ t \in T \Big\} ,$$

we have $\operatorname{card} U_n \le N_n$. By (2.178) we have

$$d_\infty(t^2, U_n) \le \sum_{m>n} \|u(t,m)\|_\infty \le 2V^{-j_n(t)}$$

so that (2.171) follows from Theorem 1.3.5.

By (2.177) and (2.178) we have

$$d_2(t^2, U_n) \le \sum_{m>n} \|u(t,m)\|_2 \le \sum_{m>n} V^{j_m(t)-j_{m-1}(t)}\delta(A_m(t))$$

and thus

$$\sum_{n \geq 0} 2^{n/2} d_2(t^2, U_n) \leq L \sum_{m \geq 1} 2^{m/2} V^{j_m(t) - j_{m-1}(t)} \delta(A_m(t)) \,,$$

and (2.170) follows from Theorem 1.3.5. Moreover,

$$\|t^2\|_1 \leq \sum_{m \geq 1} \|u(t,m)\|_1 \leq \sum_{m \geq 1} V^{2j_m(t) - j_{m-1}(t)} \delta^2(A_m(t)) \,,$$

using again (2.177) and (2.178). Finally, $t^3 := t\mathbf{1}_\Omega = t - t^1 - t'^2 \in T_3$ for $t \in T$, so that $t = t^1 + t'^2 + t^3 \in T_1 + T'_2 + T_3$. $\square$

3 Matching Theorems

3.1 The Ellipsoid Theorem

As pointed out after Proposition 2.2.2, 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".

An ellipsoid (2.19) is the unit ball of the norm

$$\|x\|_{\mathcal{E}} = \Big(\sum_{i\geq 1} \frac{x_i^2}{a_i^2}\Big)^{1/2} . \tag{3.1}$$

Lemma 3.1.1. *We have*

$$\|x\|_{\mathcal{E}}\,,\ \|y\|_{\mathcal{E}} \leq 1 \Rightarrow \Big\|\frac{x+y}{2}\Big\|_{\mathcal{E}} \leq 1 - \frac{\|x-y\|_{\mathcal{E}}^2}{8} . \tag{3.2}$$

Proof. By the parallelogram identity we have

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

so that

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

and

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

□

Since (3.2) is the only property of ellipsoids we will use, it clarifies matters to state the following definition.

Definition 3.1.2. *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*

$$\|x\|\,,\ \|y\| \leq 1 \Rightarrow \Big\|\frac{x+y}{2}\Big\| \leq 1 - \eta\|x-y\|^p . \tag{3.3}$$

For example, for $q < \infty$ the classical Banach space $L^q(\mu)$ is p-convex where $p = \min(2, q)$. The reader is referred to [20] for this result and any other classical facts about Banach spaces.

In this section we will 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 $\|\cdot\|_\sim$.

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

$$\gamma_{\alpha,\beta}(T,d) = \Bigl(\inf \sup_t \sum_{n\ge 0} \bigl(2^{n/\alpha} \Delta(A_n(t))\bigr)^\beta\Bigr)^{1/\beta} , \tag{3.4}$$

where α and β are positive numbers, and where the infimum is over all admissible sequences. Thus, with the notation of Definition 1.2.5, we have $\gamma_{\alpha,1}(T,d) = \gamma_\alpha(T)$. The importance of these functionals is that in certain conditions they will nicely relate to $\gamma_2(T,d)$ through Hölder's inequality.

Theorem 3.1.3. *If T is the unit ball of a p-convex Banach space, if η is as in (3.3) and if the distance d on T is induced by another norm $\|\cdot\|_\sim$, then*

$$\gamma_{\alpha,p}(T,d) \le K(\alpha,p,\eta) \sup_{n\ge 0} 2^{n/\alpha} e_n(T,d) . \tag{3.5}$$

The point of this theorem is that, for a general metric space (T, d), it is true that

$$\gamma_{\alpha,p}(T,d) \le K(\alpha)\Bigl(\sum_{n\ge 0} \bigl(2^{n/\alpha} e_n(T,d)\bigr)^p\Bigr)^{1/p} , \tag{3.6}$$

which, if one does not mind the worst constant, is weaker in an essential way than (3.5), but that in general it is essentially impossible to improve on (3.6). The proofs of (3.6) and it optimality are left as an easy but instructive exercise. Another easy observation is that

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

Corollary 3.1.4. (The Ellipsoid Theorem.) *Consider the ellipsoid (2.19), where the sequence a_i is not necessarily non-increasing. Consider $\alpha \ge 1$. Then we have*

$$\gamma_{\alpha,2}(\mathcal{E}) \le K(\alpha) \sup_{\varepsilon>0} \varepsilon(\mathrm{card}\{i\,;\, a_i \ge \varepsilon\})^{1/\alpha}. \tag{3.7}$$

Proof. Without loss of generality we can assume that the sequence (a_i) is non-decreasing. We apply Theorem 3.1.3 to the case $\|\cdot\| = \|\cdot\|_{\mathcal{E}}$, and where $\|\cdot\|_\sim$ is the norm of ℓ^2, and we get

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

using (2.25) in the last inequality. Now, taking $\varepsilon = a_{2^n}$ we see that we have

$$2^{n/\alpha} a_{2^n} \le \sup_{\varepsilon > 0} \varepsilon (\operatorname{card}\{i\,;\, a_i \ge \varepsilon\})^{1/\alpha} .$$

□

The restriction $\alpha \ge 1$ is inessential and can be removed by a suitable modification of (2.25).

The Ellipsoid Theorem will be our main tool to construct matchings. The more general Theorem 3.1.3 will have equally far-reaching consequences in Section 6.3.

We will deduce Theorem 3.1.3 from the following general result.

Theorem 3.1.5. *Under the hypothesis of Theorem 3.1.3, consider a sequence* $(\theta(n))_{n\ge 0}$*, such that*

$$\forall n \ge 0\,,\ \theta(n) \le \eta \Big(\frac{1}{4 e_n(T,d)}\Big)^p \tag{3.8}$$

and that, for certain numbers $1 < \xi \le 2$, $r \ge 4$ *we have*

$$\forall n \ge 0\,,\ \xi\theta(n) \le \theta(n+1) \le \frac{r^p}{2}\theta(n)\ . \tag{3.9}$$

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

$$\sup_{t\in T} \sum_{n\ge 0} \theta(n) \Delta(A_n(t), d)^p \le L \frac{(2r)^p}{\xi - 1}\ . \tag{3.10}$$

The abstraction here might make it hard for the reader to realize at once that this is a very powerful and precise statement. Not only it implies Corollary 3.1.4, but also (2.22), as is shown after the statement of Theorem 3.1.6.

Proof. We will use Theorem 1.3.2 for $\tau = 1$, $\beta = p$ and the functionals $F_n = F$ given by

$$F(A) = 1 - \inf\{\|v\|\,;\ v \in \operatorname{conv} A\}\ .$$

To prove that these functionals satisfy the growth condition of Definition 1.2.5 we consider $n \ge 0$, $m = N_{n+1}$, and points $(t_\ell)_{\ell \le m}$ in T, such that $d(t_\ell, t_{\ell'}) \ge a$ whenever $\ell \ne \ell'$. Consider also sets $H_\ell \subset T \cap B_d(t_\ell, a/r)$, where the index d emphasizes that the ball is for the distance d rather than for the norm $\|\cdot\|$. Set

$$u = \inf\Big\{\|v\|\,;\ v \in \operatorname{conv} \bigcup_{\ell \le m} H_\ell\Big\} = 1 - F\Big(\bigcup_{\ell\le m} H_\ell\Big)\,, \tag{3.11}$$

and consider

$$u' > \max_{\ell \le m} \inf\{\|v\|\,;\ v \in \operatorname{conv} H_\ell\} = 1 - \min_{\ell\le m} F(H_\ell)\ . \tag{3.12}$$

For $\ell \le m$ consider $v_\ell \in \mathrm{conv} H_\ell$ with $\|v_\ell\| \le u'' = \min(u', 1)$. It follows from (3.3) that for $\ell\,,\,\ell' \le m$,

$$\Big\|\frac{v_\ell + v_{\ell'}}{2u''}\Big\| \le 1 - \eta\Big\|\frac{v_\ell - v_{\ell'}}{u''}\Big\|^p\,. \tag{3.13}$$

Moreover, since $(v_\ell + v_{\ell'})/2 \in \mathrm{conv}\bigcup_{\ell\le m} H_\ell$, we have $u \le \|v_\ell + v'_\ell\|/2$, and (3.13) implies

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

so that

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

and hence the points $w_\ell = R^{-1}(v_\ell - v_1)$ belong to T. Now, since $r \ge 4$, we have $d(v_\ell, v_{\ell'}) \ge a/2$ for $\ell \ne \ell'$, and, since the distance d arises from a norm, we have $d(w_\ell, w_{\ell'}) \ge R^{-1}a/2$ for $\ell \ne \ell'$, and thus $e_{n+1}(T,d) \ge R^{-1}a/4$.

Since $u' - u \ge u'' - u \ge \eta R^p$ it follows that

$$u' \ge u + \eta\Big(\frac{a}{4e_{n+1}(T,d)}\Big)^p\,.$$

Since u' is arbitrary in (3.12) we get, using (3.11), that

$$F\Big(\bigcup_{\ell\le n} H_\ell\Big) \ge \min_{\ell\le n} F(H_\ell) + \eta\Big(\frac{a}{4e_{n+1}(T,d)}\Big)^p\,.$$

This completes the proof of (1.31). To finish the proof one uses (1.33) and one observes that $F_0(T) = F(T) = 1$ and that $\theta(0)\Delta^p(T) \le \theta(0)2^p e_0^p(T) \le \eta 2^{-p} \le 1$, using (3.8) for $n = 0$ in the last inequality. $\square$

Proof of Theorem 3.1.3. Let $S = \sup_{n\ge1} 2^{n/\alpha} e_n(T,d)$. Then the sequence

$$\theta(n) = \eta\frac{2^{np/\alpha}}{(4S)^p}$$

satisfies (3.8), and also (3.9) for $\xi = \min(2, 2^{p/\alpha})$ and whenever $r \ge 2^{1/p+1/\alpha}$. We then construct the desired admissible sequence by setting $\mathcal{B}_0 = \{T\}$ and $\mathcal{B}_n = \mathcal{A}_{n-1}$ for $n \ge 1$. $\square$

Another consequence of Theorem 3.1.5 is the following generalization of Theorem 3.1.3. When applied to ellipsoids, it yields very precise results, and in particular (2.22). It will not be used in the sequel, and could be omitted at first reading.

Theorem 3.1.6. *Consider* $\beta\,,\,\beta'\,,\,p > 0$ *with*

$$\frac{1}{\beta} = \frac{1}{\beta'} + \frac{1}{p}\,. \tag{3.14}$$

Then, under the conditions of Theorem 3.1.3 we have

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

The case of Theorem 3.1.5 is the case where $\beta' = \infty$. To recover (2.22) we simply use the case where $\alpha = 2, \beta = 1, \beta' = p = 2$ and (2.25).

Proof. For $n \ge 0$, set

$$d(n) = \eta\Big(\frac{1}{4e_n(T,d)}\Big)^p .$$

Consider $a = p/(2\alpha)$, $b = 2p/\alpha$, and set

$$\theta(n) = \min\Big(\inf_{k\ge n} d(k)2^{a(n-k)} , \inf_{k\le n} d(k)2^{b(n-k)}\Big) .$$

Then we have

$$2^a\theta(n) \le \theta(n+1) \le 2^b\theta(n) . \tag{3.15}$$

For example, to prove the left-hand side, we note that

$$\begin{aligned} 2^a \inf_{k\ge n} d(k)2^{a(n-k)} &\le \inf_{k\ge n+1} d(k)2^{a(n+1-k)} \\ 2^b \inf_{k\le n} d(k)2^{b(n-k)} &\le \inf_{k\le n} d(k)2^{b(n+1-k)} \end{aligned}$$

and we observe that $\theta(n+1)$ is the minimum of the right-hand sides of the two previous inequalities. Thus (3.9) holds for $\xi = \min(2, 2^a)$ and $r = \max(4, 2^{(b+1)/p})$ and by Theorem 3.1.5 we can find an increasing sequence $(\mathcal{A}_n)$ of partitions of T with $\operatorname{card}\mathcal{A}_n \le N_{n+1}$ and

$$\sup_{t\in T}\sum_{n\ge 0}\theta(n)\Delta(A_n(t))^p \le K(\alpha,p) . \tag{3.16}$$

Now we use (3.14) and Hölder's inequality to see that

$$\Big(\sum_{n\ge 0}\big(\Delta(A_n(t))2^{n/\alpha}\big)^\beta\Big)^{1/\beta} \le \Big(\sum_{n\ge 0}\theta(n)\Delta(A_n(t))^p\Big)^{1/p}\Big(\sum_{n\ge 0}\frac{2^{n\beta'/\alpha}}{\theta(n)^{\beta'/p}}\Big)^{1/\beta'} . \tag{3.17}$$

If we set $c = \beta'/\alpha$, we have $a\beta'/p = c/2$ and $b\beta'/p = 2c$, so that

$$\theta(n)^{-\beta'/p} \le \sum_{k\ge n} d(k)^{-\beta'/p}2^{c(k-n)/2} + \sum_{k\le n} d(k)^{-\beta'/p}2^{2c(k-n)}$$

and

$$\begin{aligned} \sum_{n\ge 0}\frac{2^{n\beta'/\alpha}}{\theta(n)^{\beta'/p}} &\le \sum_{n,k;k\ge n} d(k)^{-\beta'/p}2^{c(k+n)/2} + \sum_{n,k;k\le n} d(k)^{-\beta'/p}2^{c(2k-n)} \\ &\le K(c)\sum_{k\ge 0} d(k)^{-\beta'/p}2^{ck} \end{aligned}$$

by performing the summation in n first. Thus, recalling the value of $d(k)$,

$$\sum_{n\geq 0}\frac{2^{n\beta'/\alpha}}{\theta(n)^{\beta'/p}}\leq K(p,\beta,\eta)\sum_{k\geq 0}\big(2^{k/\alpha}e_k(T,d)\big)^{\beta'} .$$

Combining with (3.16) and (3.17) concludes the proof. □

3.2 Matchings

The rest of this chapter is devoted to the following problem. Consider N r.v. $X_1,\dots,X_N$ 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 $X_1,\dots,X_N$? 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 π of $\{1,\dots,N\}$ such that the points X_i and $Y_{\pi(i)}$ are "close". There are of course 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 3.3), that the maximum distance $d(X_i,Y_{\pi(i)})$ be as small as possible, (Section 3.4), or one can use more complicated measures of "closeness" (Section 3.5). The case where $d=2$ is very special, and will be the only one we study. The reader having never thought of the matter might think that the points $X_1,\dots,X_N$ 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. Of course, such a statement is by no mean obvious at this stage.

Matching problems in dimension $d\geq 2$ are very interesting, but do not share this feature, and this makes them somewhat easier. Essentially the final solution to these problems is given in [57]. One of the very different features between the case $d=2$ and the case $d>2$ is that for $d=2$ we are not concerned with what happens at a scale less that $((\log N)/N)^{1/d}$, while in dimension $d>2$ we are concerned with what happens at a scale as small as $N^{-1/d}$, and the very local irregularities play the essential role there. In a sense the heart of [57] is the study of bounds of processes with Poisson tails. The main result of this paper lies as deep as anything presented in this book. (The methods are somewhat similar to those of Section 3.5, but simpler). Unfortunately we did not find the energy to rewrite this work, which consequently is likely to keep awaiting its first reader.

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 $(Y_i)_{i\le N}$ 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 when N is not a square, one can simply cut $[0,1]^2$ into horizontal strips of width k/N, where k is about $\sqrt{N}$ (and depends on the strip), use vertical cuts to cut such a strip into k rectangles of area $1/N$, and put a point Y_i in each rectangle. There is an elegant approach that dispenses of 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 $[0,1]^2$.) Since this approach does not help with the proofs, we will not use it.

The basic tool to construct matchings is the following classical fact.

Proposition 3.2.1. *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 π of $\{1,\ldots,N\}$. Then*

$$M(C) = \sup \sum_{i\le N} (w_i + w'_i) \;, \tag{3.18}$$

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

$$\forall i\,,\, j \le N\,,\; w_i + w'_j \le c_{ij} \;. \tag{3.19}$$

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" (3.18).

Proof. Let us denote by a the right-hand side of (3.18). If the families $(w_i)_{i\le N}$, $(w'_i)_{i\le N}$ satisfy (3.19), then for any permutation π of $\{1,\ldots,N\}$, we have

$$\sum_{i\le N} c_{i\pi(i)} \ge \sum_{i\le N} (w_i + w'_i)$$

and thus

$$\sum_{i\le N} c_{i\pi(i)} \ge a \,,$$

so that $M(C) \ge a$.

The converse relies on the Hahn-Banach Theorem. Consider the subset $\mathcal{C}$ of $\mathbb{R}^{N\times N}$ that consists of the vectors $(x_{ij})_{i,j\le N}$ for which there exists numbers $(w_i)_{i\le N}$, and $(w'_i)_{i\le N}$ such that

$$\sum_{i\le N} (w_i + w'_i) > a \tag{3.20}$$

$$\forall i\,,\, j \le N\,,\; x_{ij} \ge w_i + w'_j \;. \tag{3.21}$$

Then, by definition of a, we have $(c_{ij})_{i,j\le N} \notin \mathcal{C}$. Since $\mathcal{C}$ is an open convex subset of $\mathbb{R}^{N\times N}$, we can separate the point $(c_{ij})_{i,j\le N}$ from $\mathcal{C}$ by a linear functional, i.e. we can find numbers $(p_{ij})_{i,j\le N}$ such that

$$\forall (x_{ij}) \in \mathcal{C}\,,\ \sum_{i,j\le N} p_{ij}c_{ij} < \sum_{i,j\le N} p_{ij}x_{ij}\ . \tag{3.22}$$

Since by definition of $\mathcal{C}$, and in particular (3.21), this remains true when one increases x_{ij}, we see that $p_{ij}\ge 0$, and because of the strict inequality in (3.22) we see that not all the numbers p_{ij} are 0. Thus there is no loss of generality to assume that $\sum_{i,j\le N} p_{ij} = N$. Consider families $(w_i)_{i\le N}$, $(w'_i)_{i\le N}$ that satisfy (3.20). Then if $x_{ij} = w_i + w'_j$, the point $(x_{ij})_{i,j\le N}$ belongs to $\mathcal{C}$ and using (3.22) for this point we see that

$$\sum_{i,j\le N} p_{ij}c_{ij} \le \sum_{i,j\le N} p_{ij}(w_i + w'_j)\ . \tag{3.23}$$

If $(y_i)_{i\le N}$ are numbers with $\sum_{i\le N} y_i = 0$, we have

$$\begin{aligned}\sum_{i,j\le N} p_{ij}c_{ij} &\le \sum_{i,j\le N} p_{ij}(w_i + y_i + w'_j) \\ &\le \sum_{i,j\le N} p_{ij}(w_i + w'_j) + \sum_{i\le N} y_i\Big(\sum_{j\le N} p_{ij}\Big)\end{aligned} \tag{3.24}$$

as we see from (3.23), replacing w_i by $w_i + y_i$. But (3.24) forces all the sums $\sum_{j\le N} p_{ij}$ to the equal, and since $\sum_{i,j\le N} p_{ij} = N$, we have $\sum_{j\le N} p_{ij} = 1$, for all i. Similarly, we have $\sum_{i\le N} p_{ij} = 1$ for all j, i.e. the matrix $(p_{ij})_{i,j\le N}$ is bistochastic. Thus (3.23) becomes

$$\sum_{i,j\le N} p_{ij}c_{ij} \le \sum_{i\le N} (w_i + w'_i)$$

so that $\sum_{i,j\le N} p_{ij}c_{ij} \le a$. The set of bistochastic matrices is a convex set, so the infimum of $\sum_{i,j\le N} p_{ij}c_{ij}$ over this convex set is obtained at an extreme point. The extreme points are of the type $p_{ij} = \mathbf{1}_{\{\pi(i)=j\}}$ for a permutation π of $\{1,\dots,N\}$, so that we can find such a permutation with $\sum_{i\le N} c_{i\pi(i)} \le a$. □

The following is a well-known, and rather useful, result of combinatorics.

Corollary 3.2.2. *(Hall's marriage Lemma). Assume that to each $i\le N$ we associate a subset $A(i)$ of $\{1,\dots,N\}$ and that, for each subset I of $\{1,\dots,N\}$ we have*

$$\operatorname{card}\Big(\bigcup_{i\in I} A(i)\Big) \ge \operatorname{card} I\ . \tag{3.25}$$

Then we can find a permutation π of $\{1,\dots,N\}$ such that

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

Proof. We set $c_{ij} = 0$ if $j \in A(i)$ and $c_{ij} = 1$ otherwise. We want to prove, with the notations of Proposition 3.2.1, that $M(C) = 0$. Using (3.18), it suffices to prove that given numbers $u_i(= -w_i)$, $v_i(= w'_i)$ we have

$$\forall i, \forall j \in A(i)\,,\; v_j \le u_i \Rightarrow \sum_{i \le N} v_i \le \sum_{i \le N} u_i\,. \tag{3.26}$$

By adding a suitable constant, we can assume v_i and $u_i \ge 0$ for all i, and thus

$$\sum_{i \le N} u_i = \int_0^\infty \operatorname{card}\{i \le N\,;\; u_i \ge t\} \mathrm{d}t \tag{3.27}$$

$$\sum_{i \le N} v_i = \int_0^\infty \operatorname{card}\{i \le N\,;\; v_i \ge t\} \mathrm{d}t\,. \tag{3.28}$$

Given t, using (3.25) for $I = \{i \le N\,;\; u_i < t\}$ and that $v_j \le u_i$ if $j \in A(i)$, we see that

$$\operatorname{card}\{j \le N\,;\; v_j < t\} \ge \operatorname{card}\{i \le N\,;\; u_i < t\}$$

and thus

$$\operatorname{card}\{i \le N\,;\; u_i \ge t\} \le \operatorname{card}\{i \le N\,;\; v_i \ge t\}\,.$$

Combining with (3.27) and (3.28) this proves (3.26). □

There are other proofs of Hall's lemma, based on different ideas, see [3], § 2.

3.3 The Ajtai, Komlòs, Tusnàdy Matching Theorem

Theorem 3.3.1. *[1]. 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 $N \ge 2$)*

$$\mathsf{E} \inf_\pi \sum_{i \le N} d(X_i, Y_{\pi(i)}) \le L\sqrt{N \log N} \tag{3.29}$$

where the infimum is over all permutations of $\{1, \dots, N\}$ 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 $\sqrt{\log N}$.

It is shown in [1] that (3.29) can be reversed, i.e.

$$\mathsf{E} \inf_\pi \sum_{i \le N} d(X_i, Y_{\pi(i)}) \ge \frac{1}{L}\sqrt{N \log N}\,. \tag{3.30}$$

As will be apparent later, the left-hand side of (3.30) is essentially the expected value of the supremum of a stochastic process. Finding lower bounds for such quantities is one of our main endeavors. It would be of interest to prove (3.30) using our methods, but this has yet to be done.

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. Theorem 3.3.1 is a consequence of the following fact, interesting in its own right. We denote by λ the uniform measure on $[0,1]^2$.

Theorem 3.3.2. *We have*

$$\mathsf{E} \sup_{f \in \mathcal{C}} \Big| \sum_{i \le N} \Big(f(X_i) - \int f \mathrm{d}\lambda\Big) \Big| \le L\sqrt{N \log N} \,. \tag{3.31}$$

This theorem is a prime example of a natural situation where Dudley's entropy bound cannot yield the correct result. We will let the reader convince herself that Dudley's entropy bound cannot yield better than a bound $L\sqrt{N}\log N$. This is closely related to the fact that, as explained in Section 2.2, Dudley's entropy bound is not sharp on all ellipsoids.

Research problem 3.3.3. 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.

Proof of Theorem 3.3.1. We use Proposition 3.2.1 with $c_{ij} = d(X_i, Y_j)$, so that

$$\inf_\pi \sum_{i \le N} d(X_i, Y_{\pi(i)}) = \sup \sum_{i \le N} (w_i + w'_i) \,, \tag{3.32}$$

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

$$\forall i\,,\, j \le N \,,\ w_i + w'_j \le d(X_i, Y_j) \,. \tag{3.33}$$

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

$$f(x) = \min_{j \le N} (-w'_j + d(x, Y_j)) \,. \tag{3.34}$$

It is 1-Lipschitz, since it is the minimum of functions that are themselves 1-Lipschitz. By definition we have $f(Y_j) \le -w'_j$ and by (3.33) for $i \le N$ we have $w_i \le f(X_i)$, so that

$$\begin{aligned}\sum_{i\le N}(w_i+w_i') &\le \sum_{i\le N}(f(X_i)-f(Y_i))\\ &\le \Big|\sum_{i\le N}(f(X_i)-\int f d\lambda)\Big| + \Big|\sum_{i\le N}(f(Y_i)-\int f\mathrm{d}\lambda)\Big| \,.\end{aligned}$$

It should be obvious that the last term is $\le L\sqrt{N}$, so that, using (3.32) 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 (3.31). □

Consider the class $\mathcal{C}_0$ consisting of functions $f : [0,1]^2 \to \mathbb{R}$ that are differentiable and satisfy

$$\sup\Big|\frac{\partial f}{\partial x}\Big| \le 1 \; ; \; \sup\Big|\frac{\partial f}{\partial y}\Big| \le 1$$

$$\int f\mathrm{d}\lambda = 0 \; ; \; \forall u\,,\, 0\le u\le 1\,,\, f(u,0)=f(u,1)\,,\, f(0,u)=f(1,u)\,. \tag{3.35}$$

The main ingredient in our proof of (3.31) is the following, where we use the functional $\gamma_{2,2}$ of (3.4), and where the underlying distance is the distance induced by $L^2([0,1]^2)$.

Proposition 3.3.4. *We have* $\gamma_{2,2}(\mathcal{C}_0) < \infty$.

Proof. The very beautiful idea (due to Coffman and Shor [6]) is to represent $\mathcal{C}_0$ as a subset of an ellipsoid using Fourier transforms. Fourier transforms associates to each function f on $L_2([0,1]^2)$ the complex numbers $c_{p,q}(f)$ given by

$$c_{p,q}(f) = \iint_{[0,1]^2} f(x_1,x_2)\exp 2i\pi(px_1+qx_2)\mathrm{d}x_1\mathrm{d}x_2\,. \tag{3.36}$$

For our purpose the fundamental fact is that

$$\|f\|_2 = \Big(\sum_{p,q\in\mathbb{Z}} |c_{p,q}(f)|^2\Big)^{1/2}, \tag{3.37}$$

so that if

$$\mathcal{D} = ((c_{p,q}(f))_{p,q\in\mathbb{Z}} \; ; \; f\in\mathcal{C}_0)\,,$$

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 (3.36), integration by parts and (3.35), we get

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

Using (3.37) 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

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

We view each complex number $c_{p,q}$ as a pair $(x_{p,q}, y_{p,q})$ of real numbers, and $|c_{p,q}|^2 = x^2_{p,q} + y^2_{p,q}$. For $u \ge 1$, we have

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

We then deduce from Corollary 3.1.4 that $\gamma_{2,2}(\mathcal{E}, d) < \infty$ □

We set

$$\mathcal{C}_1 = \{f \in \mathcal{C}\ ;\ \|f\|_\infty \le 2\}\ ,$$

so that $\mathcal{C}_0 \subset \mathcal{C}_1$. The following is classical.

Lemma 3.3.5. *For each $\epsilon > 0$ we have*

$$N(\mathcal{C}_1, d_\infty, \epsilon) \le \exp\Big(\frac{L}{\epsilon^2}\Big)\ , \tag{3.38}$$

where $N(\mathcal{C}_1, d_\infty, \epsilon)$ is the smallest number of balls of radius ϵ needed to cover $\mathcal{C}_1$.

Proof. We first show that given $h \in \mathcal{C}$, an integer k and letting

$$A = \big\{f \in \mathcal{C}\ ;\ \|f - h\|_\infty \le 2^{1-k}\big\}\,, \tag{3.39}$$

we have

$$N(A, d_\infty, 2^{-k}) \le \exp(L_0 2^{2k})\ . \tag{3.40}$$

Consider a subset C of A that is maximal with respect to the property that if $f_1\,,\ f_2 \in C$ and $f_1 \ne f_2$ then $d_\infty(f_1, f_2) > 2^{-k}$. Then each point of A is within distance $\le 2^{-k}$ of C, so $N(A, d_\infty, 2^{-k}) \le \operatorname{card} C$.

Consider the subset U of $[0,1]^2$ that consists of the points of the type $y = (\ell_1 2^{-k-3}, \ell_2 2^{-k-3})$ for $\ell_1\,,\ \ell_2 \in \mathbb{N}\,,\ 1 \le \ell_1\,,\ \ell_2 \le 2^{k+3}$, so that $\operatorname{card} U \le 2^{2k+6}$. If $f_1\,,\ f_2 \in C$ and $f_1 \ne f_2$, there is $x \in [0,1]^2$ with $|f_1(x) - f_2(x)| > 2^{-k}$. Consider $y \in U$ with $d(x,y) \le 2^{-k-2}$. Then for $j = 1\,,2$ we have $|f_j(x) - f_j(y)| \le 2^{-k-2}$ so that $|f_1(y) - f_2(y)| > 2^{-k-1}$. Thus, if we see C as a subset of $\mathbb{R}^U$, we have shown that any two distinct points of C are at distance at least 2^{-k-1} of each other for the supremum norm. The balls for this norm of radius 2^{-k-2} centered at the points of C have disjoint interiors, and (from (3.39)) are entirely contained in a certain ball of radius $9\cdot 2^{-k-2}$, so that, by volume considerations, we have $\operatorname{card} C \le 9^{\operatorname{card} U}$, and this prove (3.40).

We now prove by induction over $k \ge 0$ that

$$N(\mathcal{C}_1, d_\infty, 2^{1-k}) \le \exp(L_0 2^{2k}) . \tag{3.41}$$

This certainly holds true for $k = 0$. For the induction step, we use the induction hypothesis to cover $\mathcal{C}_1$ by $\exp(L_0 2^{2k})$ sets A of the type (3.39) and we use (3.40) for each of these sets. This completes the induction. Finally, (3.38) follows from (3.41). $\square$

It is useful to reformulate (3.38) as follows. For $n \ge 0$, we have

$$e_n(\mathcal{C}_1, d_\infty) \le L 2^{-n/2} . \tag{3.42}$$

Proposition 3.3.6. *We have*

$$\mathsf{E} \sup_{f \in \mathcal{C}_0} \Big| \sum_{i \le N} f(X_i) \Big| \le L \sqrt{N \log N} . \tag{3.43}$$

Proof. Consider the largest integer m with $2^{-m} \ge 1/N$. By (3.42), and since $\mathcal{C}_0 \subset \mathcal{C}_1$, we can find a subset T of $\mathcal{C}_0$ with $\operatorname{card} T \le N_m$ and

$$\forall f \in \mathcal{C}_0 \, , \; d_\infty(f, T) \le L 2^{-m/2} \le L/\sqrt{N} .$$

Thus

$$\mathsf{E} \sup_{f \in \mathcal{C}_0} \Big| \sum_{i \le N} f(X_i) \Big| \le \mathsf{E} \sup_{f \in T} \Big| \sum_{i \le N} f(X_i) \Big| + L\sqrt{N} . \tag{3.44}$$

By Proposition 3.3.4 we have $\gamma_{2,2}(T) \le L$, so that there is an admissible sequence $(\mathcal{A}_n)$ of T for which

$$\forall t \in T \, , \; \sum_{n \ge 0} 2^n \Delta(A_n(t), d_2)^2 \le L . \tag{3.45}$$

Since $\operatorname{card} T \le N_m$, we can assume that $A_m(t) = \{t\}$ for each t, so that in (3.45) the sum is really over $n \le m-1$. Since $\sum_{0 \le n \le m} a_n \le \sqrt{m} (\sum_{0 \le n \le m} a_n^2)^{1/2}$ by the Cauchy- Schwarz inequality, we have shown that

$$\forall t \in T \, , \; \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t), d_2) \le L\sqrt{m} \le L\sqrt{\log N} \, , \tag{3.46}$$

and thus $\gamma_2(T, d_2) \le L\sqrt{\log N}$.

Using (1.49) with (3.42) and $T_n = T$ for $n \ge m$, we see that $\gamma_1(T, d_\infty) \le L 2^{m/2} \le L\sqrt{N}$. Thus (3.43) follows from Proposition 2.7.2. $\square$

We consider the class $\mathcal{C}_2$ of functions of the type

$$f(x_1, x_2) = x_1 g(x_2)$$

where $g : [0,1] \to \mathbb{R}$ is 1-Lipschitz, $g(0) = g(1)$ and $|g| \le 1$.

Proposition 3.3.7. *We have*

$$\mathsf{E}\sup_{f\in\mathcal{C}_2}\Big|\sum_{i\le N}(f(X_i)-\int f\mathrm{d}\lambda)\Big|\le L\sqrt{N}\,.$$

Proof. We observe that for 2 functions g and g^* we have $|x_1g(x_2)-x_1g^*(x_2)|\le d_\infty(g,g^*)$. Mimicking the proof of the entropy estimate (3.38), we see that for $\epsilon>0$ we have $N(\mathcal{C}_2,d_\infty,\epsilon)\le\exp(L/\epsilon)$ and hence $e_n(\mathcal{C}_2,d_\infty)\le L2^{-n}$. Thus, by Theorem 1.3.5 we have $\gamma_2(\mathcal{C}_2,d_2)\le\gamma_2(\mathcal{C}_2,d_\infty)\le L$. We consider the largest integer m such that $2^{-m}\ge 1/N$. We choose $T\subset\mathcal{C}_2$ with $\operatorname{card}T\le N_m$ and

$$\forall f\in T_2\,,\ d_\infty(f,T)\le L2^{-m}\,.$$

As in the proof of Proposition 3.3.6, we see that we have $\gamma_1(T,d_\infty)\le Lm$ and we conclude by Bernstein's inequality and Theorem 1.2.7, using an inequality similar to (3.44), with huge room to spare. □

Proof of Theorem 3.3.2. We first observe that in (3.31) the supremum is the same if we replace the class $\mathcal{C}$ of 1-Lipschitz functions by the class of differentiable 1-Lipschitz functions. For a function f on $[0,1]^2$, we set $\Delta=f(1,1)-f(1,0)-f(0,1)+f(0,0)$ and we decompose

$$f=f_1+f_2+f_3+f_4\,,\tag{3.47}$$

where

$$\begin{aligned}f_4(x_1,x_2)&=x_1x_2\Delta\\ f_3(x_1,x_2)&=x_2(f(x_1,1)-f(x_1,0)-\Delta x_1)\\ f_2(x_1,x_2)&=x_1(f(1,x_2)-f(0,x_2)-\Delta x_2)\\ f_1&=f-f_2-f_3-f_4\,.\end{aligned}$$

It is straightforward to see that $f_1(x_1,0)=f_1(x_1,1)$ and $f_1(0,x_2)=f_1(0,x_2)$, so that if f is 2-Lipschitz and differentiable, f_1 is L-Lipschitz, differentiable, and $f_1-\int f_1\mathrm{d}\lambda$ satisfies (3.35). We then write

$$\Big|\sum_{i\le N}(f(X_i)-\int f\mathrm{d}\lambda)\Big|\le\sum_{j\le 4}\mathcal{D}_j$$

where $\mathcal{D}_j=|\sum_{i\le N}(f_j(X_i)-\int f_j\mathrm{d}\lambda)|$ and we conclude by Propositions 3.3.6 and 3.3.7. □

3.4 The Leighton-Shor Grid Matching Theorem.

Theorem 3.4.1. *[19]. If the points $(Y_i)_{i\le N}$ are evenly spread and if $(X_i)_{i\le N}$ are i.i.d uniform over $[0,1]^2$, then (for $N\ge 2$), with probability at least $1-L\exp(-(\log N)^{3/2}/L)$ we have*

$$\inf_{\pi} \sup_{i \le N} d(X_i, Y_{\pi(i)}) \le L \frac{(\log N)^{3/4}}{\sqrt{N}} , \tag{3.48}$$

and thus

$$\mathsf{E} \inf_{\pi} \sup_{i \le N} d(X_i, Y_{\pi(i)}) \le L \frac{(\log N)^{3/4}}{\sqrt{N}} . \tag{3.49}$$

To deduce (3.49) from (3.48) one simply uses any matching in the (rare) event that (3.48) fails.

It is proved in [19] that the inequality (3.49) can be reversed. It would be interesting to prove this by our methods, but this has yet to be done.

We consider the largest integer ℓ_1 with $2^{-\ell_1} \ge (\log N)^{3/4}/\sqrt{N}$, and the grid G of $[0,1]^2$ of mesh width $2^{-\ell_1}$ defined by

$$G = \left\{(x_1, x_2) \in [0,1]^2 \ ; \ 2^{\ell_1} x_1 \in \mathbb{N} \ \text{ or } \ 2^{\ell_1} x_2 \in \mathbb{N}\right\} .$$

A *vertex* of the grid is a point (x_1, x_2) with $2^{\ell_1} x_1 \in \mathbb{N}$, $2^{\ell_1} x_2 \in \mathbb{N}$. 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.

A *simple curve* is the image of a continuous map $\varphi : [0,1] \to \mathbb{R}^2$ that 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 $\varphi(0) = \varphi(1)$. A closed simple curve separates $\mathbb{R}^2$ in two regions. One of these is bounded. It is called the interior of C and is denoted by $\overset{o}{C}$.

The key ingredient to Theorem 3.4.1 is as follows.

Theorem 3.4.2. *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*

$$\Big| \sum_{i \le N} \big(\mathbf{1}_{\overset{o}{C}}(X_i) - \lambda(\overset{o}{C})\big) \Big| \le L \ell(C) \sqrt{N} (\log N)^{3/4} , \tag{3.50}$$

where $\lambda(\overset{o}{C})$ *is the area of* $\overset{o}{C}$ *and* $\ell(C)$ *is the length of* C.

In particular, this result allows the control for each number A of the supremum of the left-hand side of (3.50) over all the simple curves C traced on G with $\ell(C) \le A$.

This will be deduced from the following.

Proposition 3.4.3. *Consider a vertex* w *of* G *and* $k \in \mathbb{Z}$. *Define* $\mathcal{C}(w,k)$ *as the set of closed simple curves traced on* G *that contain* w *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* $C \in \mathcal{C}(w,k)$ *we have*

$$\Big| \sum_{i \le N} \big(\mathbf{1}_{\overset{o}{C}}(X_i) - \lambda(\overset{o}{C})\big) \Big| \le L 2^k \sqrt{N} (\log N)^{3/4} . \tag{3.51}$$

Proof of Theorem 3.4.2. Since there are at most $(2^{\ell_1}+1)^2$ choices for w, we can assume with probability at least

$$1 - L(2^{\ell_1}+1)^2(2\ell_1+4)\exp(-(\log N)^{3/2}/L) \geq 1 - L'\exp\bigl(-(\log N)^{3/2}/L'\bigr)$$

that (3.51) occurs for all choices of $C \in \mathcal{C}(w,k)$, for any w and any k with $-\ell_1 \leq k \leq \ell_1 + 2$.

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

$$2^{-\ell_1} \leq \ell(C) \leq 2(2^{\ell_1}+1) \leq 2^{\ell_1+2}\,,$$

so if k is the smallest integer for which $\ell(C) \leq 2^k$, then $-\ell_1 \leq k \leq \ell_1 + 2$, so that we can use (3.51) and since $2^k \leq 2\ell(C)$ the proof is finished. □

Lemma 3.4.4. *We have* $\operatorname{card}\mathcal{C}(w,k) \leq 2^{2^{k+\ell_1+1}} = N_{k+\ell_1+1}$.

Proof. A curve $C \in \mathcal{C}(w,k)$ 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 $\operatorname{card}\mathcal{C}(w,k) \leq 4^{2^{k+\ell_1}} = N_{k+\ell_1+1}$. □

On the set of closed simple curves traced on G, we define the distance d_1 by $d_1(C,C') = \lambda(\overset{o}{C}\,\Delta\,\overset{o}{C}{}')$.

Proposition 3.4.5. *We have*

$$\gamma_{1,2}(\mathcal{C}(w,k),\, d_1) \leq L2^{2k}\,. \tag{3.52}$$

This is the main ingredient of Proposition 3.4.3; we will prove it later.

Lemma 3.4.6. *Consider a metric space* (T,d) *with* $\operatorname{card} T \leq N_m$. *Then*

$$\gamma_2(T,\sqrt{d}) \leq m^{3/4}\gamma_{1,2}(T,d)^{1/2}\,. \tag{3.53}$$

Proof. Consider an admissible sequence $(\mathcal{A}_n)$ of T such that

$$\forall t \in T\,,\ \sum_{n\geq 0}(2^n\Delta(A_n(t),d))^2 \leq \gamma_{1,2}^2(T,d)\,. \tag{3.54}$$

Without loss of generality we can assume that $A_m(t) = \{t\}$ for each t, so that in (3.54) the sum is over $n \leq m-1$. Now

$$\Delta(A,\sqrt{d}) \leq \Delta(A,d)^{1/2}$$

so that, using Hölder's inequality,

$$\begin{aligned}\sum_{0\le n\le m-1} 2^{n/2}\Delta(A_n(t),\sqrt{d}) &\le \sum_{0\le n\le m-1} (2^n\Delta(A_n(t),d))^{1/2}\\ &\le m^{3/4}\Big(\sum_{n\ge 0}\big(2^n\Delta(A_n(t),d)\big)^2\Big)^{1/4}\\ &\le m^{3/4}\gamma_{1,2}(T,d)^{1/2}\;.\end{aligned}$$

□

On the set of simple curves traced on G we consider the distance

$$d_2(C_1,C_2)=\sqrt{N}\|\mathbf{1}_{\overset{\circ}{C}_1}-\mathbf{1}_{\overset{\circ}{C}_2}\|_2=\big(Nd_1(C_1,C_2)\big)^{1/2}\;, \tag{3.55}$$

so that

$$\gamma_2(\mathcal{C}(w,k),d_2)\le\sqrt{N}\gamma_2(\mathcal{C}(w,k),\sqrt{d_1}).$$

When $k\le\ell_1+2$ we have $m:=k+\ell_1+1\le L\log N$, so that combining Proposition 3.4.5 with Lemmas 3.4.4 and 3.4.6 we see that

$$\gamma_2(\mathcal{C}(w,k),d_2)\le L2^k\sqrt{N}(\log N)^{3/4}\;. \tag{3.56}$$

Proof of Proposition 3.4.3. We use Theorem 1.2.9 with $T=\mathcal{C}(w,k)$. It follows from Bernstein's inequality that the process $X_C=L^{-1}\sum_{i\le N}(\mathbf{1}_{\overset{\circ}{C}}(X_i)-\lambda(C))$ satisfies (1.21) where d_2 is given by (3.55) and d_1 is the distance δ given by $\delta(C,C')=1$ if $C\ne C'$ and $\delta(C,C')=0$ if $C=C'$. By Lemma 3.4.4 we have $\gamma_1(T,\delta)\le L2^{k+\ell_1}\le L2^k\sqrt{N}$ and $\sum_n e_n(T,\delta)\le k+\ell_1+1$, since $e_n(T,\delta)\le 1$ and $e_n(T,\delta)=0$ for $n\ge k+\ell_1+1$. Also, from (3.52) we have (see (3.54)) $e_n(T,d_1)\le L2^{2k}2^{-n}$, so that $e_n(T,d_2)\le L\sqrt{N}\,2^k2^{-n/2}$ and $\sum_{n\ge 0}e_n(T,d_2)\le L\sqrt{N}\,2^k$. We simply use (1.27) with $u_1=(\log N)^{3/2}$, $u_2=(\log N)^{3/4}$ to obtain the desired bound. □

Lemma 3.4.7. *Consider the set $\mathcal{L}$ of functions $f:[0,1]\to\mathbb{R}$ such that $f(0)=f(1/2)=f(1)=0$, 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. We use again Fourier transforms, and that

$$\|f\|_2=\Big(\sum_{p\in\mathbb{Z}}|c_p(f)|^2\Big)^2 \tag{3.57}$$

where

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

By integration by parts, $-2\pi ipc_p(f)=c_p(f')$, so that, using (3.57) for f', we have $\sum_{p\in\mathbb{Z}}p^2|c_p(f)|^2\le\sum_{p\in\mathbb{Z}}|c_p(f')|^2\le 1$ and since $|c_0(f)|\le\|f\|_2\le 1$, the result follows by Corollary 3.1.4 as in the case of Proposition 3.3.4. □

Lemma 3.4.8. *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. We proceed as in Theorem 1.3.6, b). It is straight forward to extend the second proof of Theorem 1.3.5 to the case of $\gamma_{\alpha,\beta}$. □

Proof of Proposition 3.4.5. To $f \in \mathcal{L}$ we associate the curve $W(f)$ traced out by the map

$$u \mapsto \Big(w_1 + 2^{k+1} f(\frac{u}{2}), w_2 + 2^{k+1} f(\frac{u+1}{2})\Big)\,,$$

where $(w_1, w_2) = w$, so that $\mathcal{C}(w, k) \subset W(\mathcal{L})$. We set $T = W^{-1}(\mathcal{C}(w, k))$. Consider f_0 and f_1 in T and the map $h : [0, 1]^2 \to [0, 1]^2$ given by

$$\begin{aligned} h(u, v) = \Big(\ & w_1 + 2^{k+1}\big(v f_0(\frac{u}{2}) + (1 - v) f_1(\frac{u}{2})\big)\,, \\ & w_2 + 2^{k+1}\big(v f_0(\frac{1+u}{2}) + (1 - v) f_1(\frac{1+u}{2})\big)\Big)\,. \end{aligned}$$

The area of $h([0, 1]^2)$ is at most $\iint_{[0,1]^2} |Jh(u, v)| \mathrm{d}u\mathrm{d}v$, where Jh is the Jacobian of h, and a straightforward computation gives

$$\begin{aligned} Jh(u, v) = 2^{2k+1}\Big(&\big(v f_0'(\frac{u}{2}) + (1 - v) f_1'(\frac{u}{2})\big)\big(f_0(\frac{1+u}{2}) - f_1(\frac{1+u}{2})\big) \\ &- \big(v f_0'(\frac{1+u}{2}) + (1 - v) f_1'(\frac{1+u}{2})\big)\big(f_0(\frac{u}{2}) - f_1(\frac{u}{2})\big)\Big)\,, \end{aligned}$$

so that, since $|f_0'| \le 1\,,\ |f_1'| \le 1$,

$$|Jh(u, v)| \le 2^{2k+1}\Big(\big|f_0(\frac{u}{2}) - f_1(\frac{u}{2})\big| + \big|f_0(\frac{1+u}{2}) - f_1(\frac{1+u}{2})\big|\Big)$$

and, by the Cauchy-Schwarz inequality,

$$\iint |Jh(u, v)| \mathrm{d}u\mathrm{d}v \le L 2^{2k} \|f_0 - f_1\|_2\,. \tag{3.58}$$

If x does not belong to the range of h, both curves $W(f_0)$ and $W(f_1)$ "turn the same number of times around x", so that either $x \in \overset{o}{W}(f_0) \cap \overset{o}{W}(f_1)$ or $x \notin \overset{o}{W}(f_0) \cup \overset{o}{W}(f_1)$. Thus the range of h contains $\overset{o}{W}(f_0) \varDelta \overset{o}{W}(f_1)$, and by (3.58) we have

$$d_1(W(f_0), W(f_1)) \le 2^{2k} \|f_0 - f_1\|_2\,.$$

Lemmas 3.4.7 and 3.4.8 finish the proof. □

We say that a simple curve C traced on G is a *chord* if it is the range of $[0,1]$ by a continuous map φ where $\varphi(0)$ and $\varphi(1)$ belong to the boundary of $[0,1]^2$. A chord divides $[0,1]^2$ into two regions R_1 and R_2, and

$$\sum_{i\le N}\big(\mathbf{1}_{R_1}(X_i)-\lambda(R_1)\big)=-\sum_{i\le N}\big(\mathbf{1}_{R_2}(X_i)-\lambda(R_2)\big)\,.$$

We define

$$\mathcal{D}(C)=\Big|\sum_{i\le N}(\mathbf{1}_{R_1}(X_i)-\lambda(R_1))\Big|=\Big|\sum_{i\le N}(\mathbf{1}_{R_2}(X_i)-\lambda(R_2))\Big|\,.$$

If C is a chord, there exists a closed simple curve C' on G such that $R_1=\overset{o}{C'}$ or $R_2=\overset{o}{C'}$ and $\ell(C')\le 4\ell(C)$. Thus, the following is a consequence of Theorem 3.4.2.

Theorem 3.4.9. *With probability at least* $1-L\exp(-(\log N)^{3/4}/L)$, *for each chord C we have*

$$\mathcal{D}(C)\le L\ell(C)\sqrt{N}(\log N)^{3/4}\,. \tag{3.59}$$

Proof of Theorem 3.4.1. Consider a number $\ell_2<\ell_1$, to be determined later, and the grid $G'\subset G$ of mesh width $2^{-\ell_2}$.

Given a union R of squares of G', we denote by R' the union of the squares of G' such one of the 4 edges that form their boundary is entirely contained in R (recall that squares include their boundaries). The main argument is to establish that if (3.50) and (3.59) hold, and provided ℓ_2 has been chosen appropriately, then for any choice of R we have

$$N\lambda(R')\ge \operatorname{card}\{i\le N\,;\ X_i\in R\}\,. \tag{3.60}$$

Let us say that a domain R is *decomposable* if $R=R_1\cup R_2$ where R_1 and R_2 are non-empty unions of squares of G', and when every square of G' included in R_1 has at most one vertex belonging to R_2. (Equivalently, $R_1\cap R_2$ is finite.) We can write $R=R_1\cup\ldots\cup R_k$ where each R_j is undecomposable (i.e. not decomposable) and where any two of these sets have a finite intersection.

We claim that

$$\frac{1}{4}\sum_{\ell\le k}\lambda(R'_\ell\backslash R_\ell)\le\lambda(R'\backslash R)\,. \tag{3.61}$$

To see this, let us set $\mathcal{S}_\ell=R'_\ell\backslash R_\ell$, so that $\mathcal{S}_\ell$ is the union of the squares $\mathcal{D}$ of G' that have one of the edges that form their boundary contained in R_ℓ but are not themselves contained in R_ℓ. Obviously we have $\mathcal{S}_\ell\subset R'$. When $\ell\ne\ell'$, the sets R_ℓ and $R_{\ell'}$ have a finite intersection, so that a square $\mathcal{D}$ contained in $\mathcal{S}_\ell$ cannot be contained in $R_{\ell'}$, since it has an entire edge contained in R_ℓ. Since $\mathcal{D}$ is not contained in R_ℓ either it is not contained in R. Thus the interior of $\mathcal{S}_\ell$ is contained in $R'\backslash R$. Moreover, a given square $\mathcal{D}$ of G' can be contained in a set $\mathcal{S}_\ell$ for at most 4 values of ℓ (one for each of the edges of $\mathcal{D}$). This proves (3.61).

To prove that (3.60) holds for any domain R, it suffices to prove that when R is an undecomposable domain we have (pessimistically)

$$\frac{N}{4}\lambda(R'\backslash R) \geq \operatorname{card}\{i \leq N\ ;\ X_i \in R\} - N\lambda(R)\ . \tag{3.62}$$

Indeed, writing (3.62) for $R = R_\ell$, summing over $\ell \leq k$ and using (3.61) implies (3.60).

We turn to the proof of (3.62). The boundary S of R is a subset of G'. By inspection of the cases, one sees that

$$\begin{aligned}&\text{If a vertice } w \text{ of } G' \text{ belongs to } S, \text{ either 2 or 4 of}\\ &\text{the edges of } G' \text{ incident to } w \text{ are contained in } S.\end{aligned} \tag{3.63}$$

Any subset S of G' that satisfies (3.63) is a union of closed simple curves, any two of them intersecting only at vertices of G'. To see this, it suffices to construct a closed simple curve C contained in S, to remove C from S and to iterate, since $S\backslash C$ still satisfies (3.63). The construction goes as follows. Starting with an edge w_1w_2 in S, we find successively edges $w_2w_3\,,\ w_3w_4\,,\ \ldots$ with $w_k \neq w_{k-2}$, and we continue the construction until the first time $w_k = w_\ell$ for some $\ell \leq k-2$ (in fact $\ell \leq k-3$). Then the edges $w_\ell w_{\ell+1}\,,\ w_{\ell+1}w_{\ell+2}\,,\ \ldots\,,\ w_{k-1}w_k$ define a closed simple curve contained in S.

Thus the boundary of an undecomposable set R is a union of closed simple curves $C_1, \ldots, C_k$, any two of them having at most a finite intersection.

We next show that for each ℓ, R is either contained in $\overset{o}{C}_\ell$ (so that C_ℓ is then the "outer boundary" of R) or else $\overset{o}{C}_\ell \cap R = \emptyset$ (in which case $\overset{o}{C}_\ell$ is "a hole" in R). Indeed otherwise R would be the union of the 2 non-empty sets $R\backslash\overset{o}{C}_\ell$ and $R \cap \overset{o}{C}_\ell$, and these two sets cannot have an edge of the grid G' in common, because this edge would have to be contained in C_ℓ, but could not be on the boundary of R.

Without loss of generality we assume that C_1 is the outer boundary of R, so that

$$R = \overset{o}{C}_1 \backslash \bigcup_{2 \leq \ell \leq k} \overset{o}{C}_\ell\ . \tag{3.64}$$

Let $R_\ell^{\sim}$ be the union of the squares of G' that have at least one edge contained in C_ℓ. Thus, as in (3.61), we have

$$\sum_{\ell \leq k} \lambda(R_\ell^{\sim}\backslash R_\ell) \leq 4\lambda(R'\backslash R)$$

and to prove (3.62) it suffices to show that for each $\ell \leq k$ we have

$$\left|\mathrm{card}\{i \le N \; ; \; X_i \in \overset{o}{C}_\ell\} - \lambda(\overset{o}{C}_\ell)\right| \le N 2^{-4} \lambda(R_\ell^\sim \backslash R) \, . \tag{3.65}$$

For $\ell \ge 2$, C_ℓ does not intersect the boundary of $[0,1]^2$. Each edge contained in C_ℓ is in the boundary of R. One of the 2 squares of G' that contain this edge is included in $R'_\ell \backslash R$, and the other in R. Since a given square contained in $R_\ell^\sim \backslash R$ must arise in this manner from one of its 4 edges, we have

$$\lambda(R_\ell^\sim \backslash R) \ge \frac{1}{4} 2^{-\ell_2} \ell(C_\ell) \tag{3.66}$$

so that (3.65) follows from (3.50) provided

$$2^{-\ell_2} \ge \frac{2^6 L}{\sqrt{N}} (\log N)^{3/4} \, , \tag{3.67}$$

where L is the constant of (3.50).

When $\ell = 1$, (3.66) need not be true because parts of C_1 might be traced on the boundary of $[0,1]^2$. In that case we simply decompose C_1 in a union of chords and of parts of the boundary of $[0,1]^2$ to deduce (3.65) from (3.59).

Thus we have proved that (3.50) and (3.59) imply (3.60). Now, since the sequence $(Y_i)_{i \le N}$ is evenly spread, it should be obvious that, provided

$$2^{-\ell_2} \ge \frac{10}{\sqrt{N}} \tag{3.68}$$

we have

$$\mathrm{card}\{i \le N \; ; \; Y_i \in (R')'\} \ge N \lambda(R')$$

and by (3.60) we have

$$\mathrm{card}\{i \le N \; ; \; Y_i \in (R')'\} \ge \mathrm{card}\{i \le N \; ; \; X_i \in R\} \, . \tag{3.69}$$

Consequently, if

$$A(i) = \left\{ j \le N \; ; \; d(X_i, Y_j) \le 2\sqrt{2} \cdot 2^{-\ell_2} \right\} \, ,$$

then for each subset I of $1, \ldots, N$ we have

$$\mathrm{card} \bigcup_{i \in I} A(i) \ge \mathrm{card}\, I \, .$$

This is seen by using (3.69) for the domain R that is the union of the squares of G' that contain at least a point X_i , $i \in I$.

The Marriage Lemma (Corollary 3.2.2) then shows that we can find a matching π for which $Y_{\pi(i)} \in A_i$, so that

$$\sup_{i \le N} d(X_i, Y_{\pi(i)}) \le 2\sqrt{2} \cdot 2^{-\ell_2} \le \frac{L}{\sqrt{N}} (\log N)^{3/4} \, ,$$

by taking for ℓ_2 the largest integer that satisfies (3.67) and (3.68). Since this is true whenever (3.50) and (3.59) occur, the proof of (3.48) is complete. □

3.5 Shor's Matching Theorem

Theorem 3.5.1. *Consider a number* $0 < \alpha < 1/2$*, an integer* $N \geq 2$*, and evenly spread points* $(Y_i)_{i \leq N}$ *of* $[0,1]^2$*. Set* $Y_i = (Y_i^1, Y_i^2)$*. Consider i.i.d points* $(X_i)_{i \leq N}$ *uniform over* $[0,1]^2$ *and set* $X_i = (X_i^1, X_i^2)$*. Then with probability* $\geq 1 - N^{-10}$ *there exists a matching* π *such that*

$$\sum_{i \leq N} \exp\left(\sqrt{\frac{N}{\log N}} \frac{|X_i^1 - Y_{\pi(i)}^1|}{K(\alpha)}\right)^{\alpha} \leq 2N \tag{3.70}$$

$$\sup_{i \leq N} |X_i^2 - Y_{\pi(i)}^2| \leq K(\alpha) \sqrt{\frac{\log N}{N}} \,. \tag{3.71}$$

Of course the power N^{10} plays no special role. Since $\exp |x|^\alpha \geq |x|/K(\alpha)$, it follows from (3.70) that

$$\sum_{i \leq N} |X_i^1 - Y_{\pi(i)}^1| \leq L\sqrt{N \log N} \,. \tag{3.72}$$

The existence of a matching satisfying (3.71) and (3.72) is due to P. Shor [40]. Of course (3.71) and (3.72) show that Theorem 3.5.1 improves upon Theorem 3.3.1.

When α increases, the conclusion of Theorem 3.5.1 becomes stronger. This is a simple consequence of the fact that if $\alpha \leq \alpha'$, there exists a number $K = K(\alpha, \alpha')$ such that

$$x \geq 0 \Rightarrow \exp \frac{x^\alpha}{K} \leq \frac{3}{2} + \frac{1}{4} \exp x^{\alpha'} \,.$$

We conjecture that Theorem 3.5.1 remains true for $\alpha = 2$. (So that the present version of Theorem 3.5.1 is quite far from being optimal.) This is a special case of the following conjecture.

Research problem 3.5.2. (The ultimate matching conjecture). Prove or disprove the following. Consider α_1, $\alpha_2 > 0$ with $1/\alpha_1 + 1/\alpha_2 = 1/2$. Then with high probability we can find a matching π such that, for $j = 1, 2$, we have

$$\sum_{i \leq N} \exp\left(\sqrt{\frac{N}{\log N}} \frac{|X_i^j - Y_{\pi(i)}^j|}{L}\right)^{\alpha_j} \leq 2N \,.$$

Noting that

$$\sum_{i \leq N} \exp a_i^4 \leq 2N \Rightarrow \max_{i \leq N} |a_i| \leq L(\log N)^{1/4} \,,$$

we see that the case $\alpha_1 = \alpha_2 = 4$ would provide a very neat generalization of Theorems 3.3.1 and 3.4.1.

The proof of Theorem 3.5.1 relies on a result of the same nature as (3.31), but for a more complicated class of functions. The basic new idea idea is to decompose each function in this class as a sum of many functions in simpler classes. This is done in Proposition 3.5.4 below. These simpler classes are then studied separately. Unfortunately this process apparently produces an irretrievable loss of information, and for this reason it seems probable that a proof of the ultimate matching conjecture cannot come from a refinement of the present approach, but rather needs a drastic new idea. Despite its shortcomings, the proof of Theorem 3.5.1 is significantly more involved than the other proofs of this chapter, and could be omitted at first reading. In fact, the author should probably have attempted to find a simple proof of a weaker result (e.g. not trying to control exponential moments in (3.70)), and should have refrained from presenting the version that gives the best know result in this book. On the other hand, it is nice to show once in a while how hard one has tried, and it is not certain that a proof of a weaker result would be essentially simpler.

We consider $\alpha < 1/2$ fixed once and for all and the function

$$\xi(x) = |x|(\log(e + |x|))^{1/\alpha} .$$

We consider an integer p and the set $G = \{1, \ldots, 2^p\}^2$. We consider the class $\mathcal{H}$ of functions $h : G \to \mathbb{R}$ such that

$$\sum |h(k, \ell + 1) - h(k, \ell)| + \sum \xi(h(k + 1, \ell) - h(k, \ell)) \le 2^{2p} . \tag{3.73}$$

The first summation is over $1 \le k \le 2^p$, $1 \le \ell \le 2^p - 1$, and the second summation is over $1 \le k \le 2^p - 1$, $1 \le \ell \le 2^p$. To lighten notation we will not mention any more that it is always understood that when a quantity such as $h(k, \ell + 1) - h(k, \ell)$ occurs in a summation, we consider only the values of ℓ with $\ell + 1 \le 2^p$.

For $(k, \ell) \in G$, we consider a integer $n(k, \ell)$, with

$$\sum_{(k,\ell)\in G} n(k, \ell) = N \tag{3.74}$$

and we assume that for a certain integer $m_0 \ge p$ we have

$$m_0 \le n(k, \ell) \le 2m_0 . \tag{3.75}$$

Thus

$$N 2^{-2p-1} \le m_0 \le N 2^{-2p} . \tag{3.76}$$

For a function h on G, we write

$$\mathsf{E} h = \frac{1}{N} \sum_{(k,\ell)\in G} n(k, \ell) h(k, \ell) . \tag{3.77}$$

Of course, the introduction of all these objects might look mysterious until we derive Theorem 3.5.1 from Theorem 3.5.3, at the end of this section.

The central ingredient to our approach is the following.

Theorem 3.5.3. *Consider independent r.v.* U_i *valued in* G*, with* $\mathsf{P}(U_i = (k,\ell)) = n(k,\ell)/N$*. Then, with probability* $\geq 1 - \exp(-96p)$*, we have*

$$\forall h \in \mathcal{H}\,, \Big|\sum_{i \leq N} (h(U_i) - \mathsf{E}h)\Big| \leq K(\alpha)\sqrt{pm_0}\, 2^{2p}\ . \tag{3.78}$$

Of course the number 96 plays no special role. We consider the class $\mathcal{H}_1$ consisting of the functions $h : G \to \mathbb{R}$ such that

$$\forall k\,,\ \ell\,,\ |h(k+1,\ell) - h(k,\ell)| \leq 1\ ;\ |h(k,\ell+1) - h(k,\ell)| \leq 1\ . \tag{3.79}$$

Given an integer $j \geq 2$, we set $a(j) = j^{-1/(2\alpha)} 2^j$, $b(j) = j^{1/(2\alpha)} 2^j$, and, for a number $V > 0$ we consider the class $\mathcal{H}_j(V)$ of functions $h : G \to \mathbb{R}$ such that

$$\forall k\,,\ \ell\,,\ |h(k+1,\ell) - h(k,\ell)| \leq a(j)\,,\ |h(k,\ell+1) - h(k,\ell)| \leq b(j) \tag{3.80}$$

$$\mathrm{card}\{(k,\ell) \in G\ ;\ h(k,\ell) \neq 0\} \leq V\ . \tag{3.81}$$

Proposition 3.5.4. *If* $h \in \mathcal{H}$ *we can find a sequence* $(V(j))_{j\geq 2}$ *with the following properties*

$$\sum_{j \geq 2} 2^j V(j) \leq K(\alpha) 2^{2p}\,,\ V(j) \leq 2^{2p-1} \tag{3.82}$$

$$h = \sum_{j \geq 1} h_j\,,\ h_1 \in L\mathcal{H}_1\,,\ h_j \in L\mathcal{H}_j(V(j)) \quad \text{for} \quad j \geq 2\ . \tag{3.83}$$

Thus, we can decompose h as a sum of terms that satisfy simple conditions, and that will be studied separately.

We will denote by I an interval of $\{1, \ldots, 2^p\}$, that is a set of the type

$$I = \{k\ ;\ k_1 \leq k \leq k_2\}\ .$$

Lemma 3.5.5. *Consider a map* $w : \{1, \ldots, 2^p\} \to \mathbb{R}^+$*, a number* $a > 0$ *and*

$$A = \Big\{k\ ;\ \exists I\,,\ k \in I\,,\ \sum_{k' \in I} w(k') \geq a\,\mathrm{card}\, I\Big\}\ .$$

Then

$$\mathrm{card}\, A \leq \frac{L}{a} \sum_{k \in A} w(k)\ .$$

Proof. This uses a discrete version of the classical Vitali covering theorem (with the same proof). Namely, a family $\mathcal{I}$ of intervals contains a disjoint family $\mathcal{I}'$ such that

$$\operatorname*{card}_{I\in\mathcal{I}}\bigcup I \le L \operatorname*{card}_{I\in\mathcal{I}'}\bigcup I = L\sum_{I\in\mathcal{I}'}\operatorname{card} I\ .$$

We use this for $\mathcal{I} = \{I\,;\,\sum_{k'\in I} w(k') \ge a\operatorname{card} I\}$, so that $A = \bigcup_{I\in\mathcal{I}} I$ and $\operatorname{card} A \le L\sum_{I\in\mathcal{I}'}\operatorname{card} I$. Since $\sum_{k'\in I} w(k') \ge a\operatorname{card} I$ for $I \in \mathcal{I}'$, and since the intervals of $\mathcal{I}'$ are disjoint and contained in A, we have $a\sum_{I\in\mathcal{I}'}\operatorname{card} I \le \sum_{k'\in A} w(k')$. □

Proof of Proposition 3.5.4. We consider $h \in \mathcal{H}$, and for $j \ge 2$ we define

$$A(j) = \Big\{(k,\ell)\in G\,;\,\exists I\,,\,k\in I\,,\,\sum_{k'\in I}|h(k'+1,\ell)-h(k',\ell)| \ge a(j)\operatorname{card} I\Big\}\ .$$

Consider $r \le 2^p$. We claim that

$$(k,\ell)\notin A(j) \Rightarrow |h(k,\ell)-h(r,\ell)| \le a(j)|r-k|\ . \tag{3.84}$$

To see this, assuming for specificity that $r > k$, we note that

$$|h(k,\ell)-h(r,\ell)| \le \sum_{k'\in I}|h(k'+1,\ell)-h(k',\ell)| < a(j)\operatorname{card} I$$

where $I = \{k, k+1, \ldots, r-1\}$, and where the last inequality follows from the fact that $k \in I$ and $(k,\ell) \notin A(j)$.

It follows from Lemma 3.5.5 (applied for each ℓ) that

$$\operatorname{card} A(j) \le \frac{L}{a(j)}\sum_{(k,\ell)\in A(j)} w(k,\ell) \tag{3.85}$$

where $w(k,\ell) = |h(k+1,\ell)-h(k,\ell)|$. We observe that

$$\sum_{(k,\ell)\in A(j)} w(k,\ell) \le \frac{a(j)}{2L}\operatorname{card} A(j) + \sum_{k,\ell} w(k,\ell)\mathbf{1}_{\{w(k,\ell)\ge a(j)/2L\}} \tag{3.86}$$

and substituting this in the right-hand side of (3.85) we get

$$a(j)\operatorname{card} A(j) \le 2L\sum_{k,\ell} w(k,\ell)\mathbf{1}_{\{w(k,\ell)\ge a(j)/2L\}}\ . \tag{3.87}$$

Since the statement of Theorem 3.5.1 becomes stronger when α increases, we can assume without loss of generality that $\alpha \ge 1/4$, so that $-1/2\alpha \ge -2$ and hence $a(j) = j^{-1/2\alpha}2^j \ge j^{-2}2^j$. Thus $x \ge a(j)/2L \Rightarrow j \le L\log(e+x)$, and since $\sum_{j\le j_0} j^{-1+1/\alpha} \le L j_0^{1/\alpha}$, for $x \ge 0$ we have

$$\sum_j j^{-1+1/\alpha}\mathbf{1}_{\{x\geq a(j)/2L\}} \leq L(\log(e+x))^{1/\alpha}\ . \tag{3.88}$$

Since $\alpha < 1/2$, we have $j^{-1+1/\alpha}a(j) \geq 2^j$, so that multiplying (3.87) by $j^{-1+1/\alpha}$ and summing over $j \geq 2$ we get that, using (3.73) and (3.88)

$$\sum_{j\geq 2} 2^j \operatorname{card} A(j) \leq L2^{2p}\ . \tag{3.89}$$

We define

$$B(j) = \Big\{(k,\ell)\ ;\ \exists I\ ,\ \ell\in I\ ,\ \sum_{\ell'\in I}|h(k,\ell'+1)-h(k,\ell')| \geq b(j)\operatorname{card} I\Big\}\ .$$

As in (3.84) we see that if $r, s, \ell \leq 2^p$, we have

$$(r,s)\notin B(j) \Rightarrow |h(r,\ell)-h(r,s)| \leq b(j)|\ell - s|\ . \tag{3.90}$$

Using Lemma 3.5.5 and (3.73), we see that $b(j)\operatorname{card} B(j) \leq L2^{2p}$, so that

$$2^j \operatorname{card} B(j) \leq L2^{2p} j^{-1/(2\alpha)}$$

and, since $1/(2\alpha) > 1$, this implies that

$$\sum_{j\geq 2} 2^j \operatorname{card} B(j) \leq K(\alpha)2^{2p}\ . \tag{3.91}$$

Thus, if $C(j) = A(j)\cup B(j)$, we have from (3.89) and (3.91) that

$$\sum_{j\geq 2} 2^j \operatorname{card} C(j) \leq L2^{2p}\ . \tag{3.92}$$

We define

$$g_j(k,\ell) = \min\big\{h(r,s) + a(j)|k-r| + b(j)|\ell-s|\ ;\ (r,s)\notin C(j)\big\}\ .$$

It should be obvious that

$$|g_j(k+1,\ell) - g_j(k,\ell)| \leq a(j) \tag{3.93}$$

$$|g_j(k,\ell+1) - g_j(k,\ell)| \leq b(j)\ . \tag{3.94}$$

Consider $(k,\ell)\notin C(j)$ and $(r,s)\notin C(j)$. Combining (3.84) and (3.90) we see that

$$|h(k,\ell)-h(r,s)| \leq a(j)|k-r| + b(j)|\ell-s|$$

and this shows that

$$(k,\ell)\notin C(j) \Rightarrow g_j(k,\ell) = h(k,\ell)\ . \tag{3.95}$$

We define $g'_j = g_{j+1} - g_j$ so that, by (3.95), and since $C(j+1) \subset C(j)$,

$$g'_j(k,\ell) \neq 0 \Rightarrow (k,\ell) \in C(j) \,.$$

Thus, if $V'(j) = \operatorname{card} C(j)$, we have $g'_j \in L\mathcal{H}_j(V'(j))$. By (3.92) there exists $j_0 = j_0(\alpha)$ (independent of h) such that $\operatorname{card} V'(j) \le 2^{2p-1}$ for $j \ge j_0$. We define $h_j = g'_j$ and $V(j) = V'(j)$ for $j \ge j_0$, $h_j = 0$, $V(j) = 0$ for $2 \le j \le j_0$, and $h_1 = g_{j_0}$. Thus $h_1 \in K(\alpha)\mathcal{H}_1$ and $h_j \in L\mathcal{H}_j(V(j))$ for $j \ge 2$.

It should be obvious that for j large enough (e.g. $j \ge 2^{2p}$) we have $C(j) = \emptyset$, so that $g_j = h$ and this shows that $h = \sum_{j\ge 1} h_j$. The proof is complete.□

The following simple result will take care of certain lower-order effects.

Proposition 3.5.6. *Consider an integer q, a number $S > 0$ and the class $\mathcal{G}(q,S)$ of functions $h : \{1,\ldots,q\} \to \mathbb{R}$ that satisfy*

$$\begin{aligned} \forall k \le q-1 \,, \quad & |h(k+1) - h(k)| \le S \\ \forall k \le q \,, \quad & |h(k)| \le 2Sq \,. \end{aligned}$$

Then

$$N(\mathcal{G}(q,S), d_\infty, \epsilon) \le \exp\Big(\frac{LSq}{\epsilon}\Big) \tag{3.96}$$

$$\gamma_2(\mathcal{G}(q,S), d) \le LSq^{3/2} \,, \tag{3.97}$$

where d denotes the Euclidean distance in $\mathbb{R}^q$ and d_∞ the supremum distance.

Proof. The proof of (3.96) is as in (3.38). Since $d \le \sqrt{q} d_\infty$, (3.97) follows from (3.96) and Theorem 1.3.5. □

The following is closely related to Proposition 3.3.4 and its relevance to Theorem 3.5.3 should be obvious.

Proposition 3.5.7. *Consider integers q_1, $q_2 \le 2^p$, a number $S > 0$ and the class $\mathcal{G}(q_1,q_2,S)$ of functions $h : G' = \{0,\ldots,q_1\} \times \{0,\ldots,q_2\} \to \mathbb{R}$ that satisfy*

$$\forall k\,,\, \ell \quad |h(k+1,\ell) - h(k,\ell)| \le \frac{S}{q_1} \;;\; |h(k,\ell+1) - h(k,\ell)| \le \frac{S}{q_2} \tag{3.98}$$

$$\forall k\,,\, \ell\,,\, |h(k,\ell)| \le 2S \,. \tag{3.99}$$

Then

$$\gamma_2(\mathcal{G}(q_1,q_2,S), d) \le L\sqrt{p} S \sqrt{q_1 q_2} \tag{3.100}$$

$$e_n(\mathcal{G}(q_1,q_2,S), d) \le LS\sqrt{q_1 q_2}\, 2^{-n/2} \,, \tag{3.101}$$

where d is the Euclidean distance on $\mathbb{R}^{G'}$.

Proof. A key idea is that there is a "main contribution" corresponding to the class $\mathcal{G}_1$ of functions that satisfy (3.98) and

$$\sum_{k,\ell} h(k,\ell) = 0 \tag{3.102}$$

$$\forall \ell \le q_2\,,\ h(0,\ell) = h(q_1,\ell) \tag{3.103}$$

$$\forall k \le q_1\,,\ h(k,0) = h(k,q_2)\,. \tag{3.104}$$

We take care of this main contribution first. Consider the groups $H_1 = \mathbb{Z}/q_1\mathbb{Z}\,,\, H_2 = \mathbb{Z}/q_2\mathbb{Z}$ and the class $\mathcal{G}$ of functions h from the product $H_1 \otimes H_2$ of H_1 and H_2 to $\mathbb{R}$, of average 0, and that satisfy (3.98), where now $k \in H_1\,,\, \ell \in H_2$ and 1 denotes the image of $1 \in \mathbb{Z}$ in either H_1 or H_2. We use Fourier transforms in the group $H_1 \otimes H_2$. For integers $r_1\,,\, r_2$, we define

$$c_{r_1 r_2}(h) = \sum_{(k,l)\in H_1\otimes H_2} \exp\Big(2i\pi\big(\frac{r_1}{q_1}k + \frac{r_2}{q_2}\ell\big)\Big) h(k,\ell)\,, \tag{3.105}$$

and we have the Plancherel formula

$$\|h\|_2^2 = \sum_{0\le r_1<q_1, 0\le r_2<q_2} |c_{r_1 r_2}(h)|^2\,. \tag{3.106}$$

Changing k into $k+1$ in (3.105) we get

$$c_{r_1 r_2}(h) = \exp\Big(2i\pi\frac{r_1}{q_1}\Big) \sum_{H_1\otimes H_2} \exp\Big(2i\pi\big(\frac{r_1}{q_1}k + \frac{r_2}{q_2}\ell\big)\Big) h(k+1,\ell)$$

and thus

$$\begin{aligned}
&\Big(\exp\big(-2i\pi\frac{r_1}{q_1}\big) - 1\Big) c_{r_1 r_2}(h) \\
&= \sum_{(k,\ell)\in H_1\otimes H_2} \exp\Big(2i\pi\big(\frac{r_1}{q_1}k + \frac{r_2}{q_2}\ell\big)\Big)\big(h(k+1,\ell) - h(k,\ell)\big)\,.
\end{aligned}$$

Using (3.106) for the function $h'(k,\ell) = h(k+1,\ell) - h(k,\ell)$ and the first part of (3.98) we get

$$\sum_{0\le r_1<q_1, 0\le r_2<q_2} \Big|1 - \exp\Big(-2i\pi\frac{r_1}{q_1}\Big)\Big|^2 |c_{r_1 r_2}(h)|^2 = \|h'\|_2^2 \le \frac{S^2}{q_1^2} q_1 q_2 = S^2 \frac{q_2}{q_1}\,.$$

We now use that for $0 \le r_1 < q_1$ we have

$$\Big|1 - \exp\Big(-2i\pi\frac{r_1}{q_1}\Big)\Big| \ge \frac{1}{Lq_1}\min(r_1, q_1 - r_1)$$

to get

$$\sum_{0\le r_1<q_1,0\le r_2<q_2} \min{}^2(r_1,q_1-r_1)\big|c_{r_1r_2}(h)\big|^2 \le LS^2q_1q_2 \; .$$

Proceeding in the same manner with the second variable, we get

$$\sum_{0\le r_1<q_1,0\le r_2<q_2} \big(\min{}^2(r_1,q_1-r_1)+\min{}^2(r_2,q_2-r_2)\big)|c_{r_1r_2}(h)|^2 \le LS^2q_1q_2 \; .$$

This relation and the Plancherel formula (3.106) describe $\mathcal{G}$ as isometric to a subset of an ellipsoid $\mathcal{E}$. For each integer n the number of pairs (r_1,r_2) for which the above coefficients of $|c_{r_1r_2}(h)|^2$ are $\le 2^n$ is at most $L2^n$, and moreover $r_1,r_2\le 2^p$ so that using (2.20) we get that $\gamma_2(\mathcal{G},d)\le LS\sqrt{pq_1q_2}$, and using (2.26) we get $e_n(\mathcal{G},d)\le LS2^{-n/2}\sqrt{q_1q_2}$. The same bounds hold true for $\mathcal{G}_1$ since $\mathcal{G}_1$ is isometric to a subset of $\mathcal{G}$.

To finish the proof, we use a decomposition similar to (3.47) to write $\mathcal{G}\subset L\mathcal{G}_1+\mathcal{G}_2+\mathcal{G}_3+\mathcal{G}_4$ where $\mathcal{G}_j$, for $2\le j\le 4$ is a genuinely smaller class that $\mathcal{G}_1$, that can be handled through Proposition 3.5.6. Finally, we appeal to (2.14) and to the easy relation $e_{n+1}(T_1+T_2,d)\le e_n(T_1,d)+e_n(T_2,d)$. □

Proposition 3.5.8. *Consider* $1\le k_1\le k_2\le 2^p$, $1\le \ell_1\le \ell_2\le 2^p$ *and* $R=\{k_1,\dots,k_2\}\times\{\ell_1,\dots,\ell_2\}$. *Consider independent r.v.* U_i *valued in* G, *with* $\mathsf{P}(U_i=(k,\ell))=n(k,\ell)/N$. *Then, with probability at least* $1-L\exp(-100p)$, *the following occurs. Consider any function* $h:G\to\mathbb{R}$, *and assume that*

$$h(k,\ell)=0 \;\text{ unless }\; (k,\ell)\in R \; . \tag{3.107}$$

$$(k,\ell),(k,\ell+1)\in R \Rightarrow |h(k,\ell+1)-h(k,\ell)|\le \frac{S}{\ell_2-\ell_1+1} \tag{3.108}$$

$$(k,\ell)\,,\,(k+1,\ell)\in R \Rightarrow |h(k+1,\ell)-h(k,\ell)|\le \frac{S}{k_2-k_1+1} \tag{3.109}$$

$$\forall (k,\ell)\in G\,,\;|h(k,\ell)|\le 2S \; . \tag{3.110}$$

Then we have

$$\Big|\sum_{i\le N}(h(U_i)-\mathsf{E}h)\Big| \le LS\sqrt{pm_0(k_2-k_1+1)(\ell_2-\ell_1+1)} \; . \tag{3.111}$$

Proof. By homogeneity we may and do assume $S=1$, and we denote by $\mathcal{G}$ the class of functions on G that satisfy conditions (3.107) to (3.110). As in (3.42) we see that $e_n(\mathcal{G},d_\infty)\le L2^{-n/2}$. Consider then the largest integer m such that $2^m\le pm_0(k_2-k_1+1)(\ell_2-\ell_1+1)$ and a subset T of $\mathcal{G}$ with card $T\le N_m$ and

$$\forall t\in\mathcal{G}\,,\; d_\infty(t,T)\le L2^{-m/2} \; . \tag{3.112}$$

Then by Theorem 1.3.5 we have $\gamma_1(T,d_\infty)\le L2^{m/2}$. By Proposition 3.5.7, we have

$$\gamma_2(T,d) \le L\sqrt{p(k_2-k_1+1)(\ell_2-\ell_1+1)}$$
$$\forall n \ge 0\,,\ e_n(T,d) \le L2^{-n/2}\sqrt{(k_2-k_1+1)(\ell_2-\ell_1+1)}\,.$$

We observe that by (3.75) we have

$$\mathsf{E}h^2 = \sum_{k,\ell} \frac{n(k,\ell)}{N} h^2(k,\ell) \le \frac{2m_0}{N} \sum_{k,\ell} h^2(k,\ell)\,.$$

Thus, by Bernstein's inequality, the conditions of Theorem 1.2.7 are satisfied with $d_1 = Ld_\infty$ and $d_2 = L\sqrt{m_0}d$. With this choice of distance d_1 and d_2, the quantities D_1 and D_2 of Theorem 1.2.9 satisfy $D_1 \le L$ and $D_2 \le L\sqrt{m_0(k_2-k_1+1)(\ell_2-\ell_1+1)}$. Using this theorem for $u_1 = Lp$ and $u_2 = L\sqrt{p}$, and since $p \le m_0$, we see that with probability $\ge 1-\exp(-100p)$ we have

$$\sup_{h\in T}\Big|\sum_{i\le N}(h(U_i)-\mathsf{E}h)\Big| \le L\sqrt{pm_0(k_2-k_1+1)(\ell_2-\ell_1+1)}\,. \tag{3.113}$$

Consider another function $h^* \in \mathcal{G}$. Since h and h^* are 0 outside R, we have

$$\begin{aligned}\Big|\sum_{i\le N}(h(U_i)-\mathsf{E}h) - \sum_{i\le N}(h^*(U_i)-\mathsf{E}h^*)\Big| \qquad\qquad (3.114)\\ \le \sum_{i\le N}|h(U_i)-h^*(U_i)| + N\mathsf{E}|h-h^*|\\ \le \|h-h^*\|_\infty(\mathrm{card}\{i\le N\,;\,U_i\in R\}+NA)\,,\end{aligned}$$

where $A = \mathsf{P}(U_i \in R)$. Since $\mathsf{P}(U_i = (k,\ell)) = n(k,\ell)/N \le 2m_0/N$, we note that

$$A \le \frac{2m_0}{N}\mathrm{card}\,R \le \frac{2m_0}{N}(k_2-k_1+1)(\ell_2-\ell_1+1)\,. \tag{3.115}$$

By (3.114) and (3.112) we have

$$\begin{aligned}\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i)-\mathsf{E}h)\Big| \le \sup_{h\in T}\Big|\sum_{i\le N}(h(U_i)-\mathsf{E}h)\Big| \qquad\qquad (3.116)\\ + L2^{-m/2}(NA+\mathrm{card}\{i\le N\,;\,U_i\in R\})\,.\end{aligned}$$

By Bernstein's inequality, we have

$$\mathsf{P}\Big(\sum_{i\le N}(\mathbf{1}_R(U_i)-A)\ge u\Big) \le \exp\Big(-\frac{1}{L}\min\Big(\frac{u^2}{NA},u\Big)\Big)$$

so that, taking $u = LNA$, and since $NA \ge m_0 \ge p$, we see that with probability at least $1-\exp(-100p)$ we have

$$\mathrm{card}\{i\le N\,;\,U_i\in R\} = \sum_{i\le N}\mathbf{1}_R(U_i) \le LNA \le Lm_0(k_2-k_1+1)(\ell_2-\ell_1+1)\,,$$

using (3.115) in the last inequality. Combining with (3.113), (3.116) and recalling the choice of m finishes the proof. (The reader observes that there is some room in the choice of m.) □

Proof of Theorem 3.5.3. Since in Proposition 3.5.8 there are (crudely) at most 2^{4p} choices for the quadruplet $(k_1, k_2, \ell_1, \ell_2)$, with probability at least $1 - L\exp(-96p)$, Condition (3.111) holds for all values of k_1, k_2, ℓ_1, ℓ_2. We assume that this is the case in the rest of the proof.

Consider $h \in \mathcal{H}$, and the decomposition of h provided by Proposition 3.5.4. Using (3.111) for $k_1 = \ell_1 = 1$, $k_2 = \ell_2 = 2^p$ and $S = L2^p$ we get

$$\Big|\sum_{i \le N}(h_1(U_i) - \mathsf{E}h_1)\Big| \le K(\alpha)\sqrt{pm_0}\,2^{2p} \,.$$

By (3.82) all we have to show is that if $h \in \mathcal{H}_j(V)$, $V \le 2^{2p-1}$ and $V \le K(\alpha)2^{-j}2^{2p}$ then

$$\Big|\sum_{i \le N}(h(U_i) - \mathsf{E}h)\Big| \le K(\alpha)\sqrt{pm_0}\,2^j V \,. \tag{3.117}$$

The idea is to use (3.111) for the functions $h\mathbf{1}_R$ where R is a suitable rectangle, and to recover (3.117) by summation. Considering j as being fixed once and for all, we define d as the largest integer for which $2^d \le 8j^{1/\alpha}$, so that $d \ge 3$. For $d \le q \le p$ we consider the partition $\mathcal{D}(q)$ of G consisting of the sets of the type

$$\{\ell_1 2^q + 1, \dots, (\ell_1 + 1)2^q\} \times \{\ell_2 2^{q-d} + 1, \dots, (\ell_2 + 1)2^{q-d}\} \,, \tag{3.118}$$

where $0 \le \ell_1 < 2^{p-q}$ and $0 \le \ell_2 < 2^{p-q+d}$. For $3 \le q \le d$, we define $\mathcal{D}(q)$ as the partition consisting of the sets of the type

$$\{\ell_1 2^q + 1, \dots, (\ell_1 + 1)2^q\} \times \{k\} \tag{3.119}$$

where $0 \le \ell_1 < 2^{p-q}$ and $1 \le k \le 2^p$.

We observe that if $q' > q$, $R' \in \mathcal{D}(q')$ and $R \in \mathcal{D}(q)$, then either $R \subset R'$ or $R \cap R' = \emptyset$.

Consider the set $C = \{(k,\ell) \;:\; h(k,\ell) \ne 0\}$ so $\operatorname{card} C \le V \le 2^{2p-1}$. We proceed to the following construction. First, we consider the set $U(p)$ that is the union of all rectangles $R \in \mathcal{D}(p)$ such that

$$\operatorname{card}(R \cap C) \ge \frac{1}{8}\operatorname{card} R \,. \tag{3.120}$$

Then we consider the union $U(p-1)$ of all the rectangles $R \in \mathcal{D}(p-1)$ that are not contained in $U(p)$ and that satisfy (3.120), and we continue in this manner until we construct $U(3)$. Since the sets $U(q)$, ..., $U(3)$ are disjoint, we have from (3.120) that

$$\sum_{3\leq q\leq p} \operatorname{card} U(p) \leq 8\operatorname{card} C \leq 8V\ . \tag{3.121}$$

Moreover

$$C \subset \sum_{3\leq q\leq p} U(q)\ . \tag{3.122}$$

This is simply because if $(k,\ell) \in C$ and $(k,\ell) \in R \in \mathcal{D}(3)$ then if $(k,\ell) \notin \bigcup_{q\geq 4} U(q)$ we have $R \subset U(3)$ since (3.120) holds because card $R = 8$. We also note that

$$R \in \mathcal{D}(q)\,,\ q \leq p-1\,,\ R \subset U(q) \Rightarrow \operatorname{card}(R\cap C) \leq \frac{1}{2}\operatorname{card} R\ . \tag{3.123}$$

Indeed if $R' \supset R$ and $R' \in \mathcal{D}(q+1)$, then $\operatorname{card} R' \leq 4\operatorname{card} R$. Since $R \in \mathcal{D}(q)$ we have $R' \not\subset U(q-1)$, so that

$$\operatorname{card}(R\cap C) \leq \operatorname{card}(R'\cap C) \leq \frac{1}{8}\operatorname{card} R' \leq \frac{1}{2}\operatorname{card} R\ .$$

From (3.122) we have

$$h = \sum h\mathbf{1}_R \tag{3.124}$$

where the summation is over $3 \leq q \leq p$, $R \in \mathcal{D}(q)$ and $R \subset U(q)$. We will apply (3.111) to each of the terms $h\mathbf{1}_R$. We start by the typical case, $R \in \mathcal{D}(q)\,,\ d \leq q < p$. Writing $R = \{k_1,\ldots,k_2\} \times \{\ell_1,\ldots,\ell_2\}$ as in Proposition 3.5.8, we have $k_2-k_1+1 = 2^q\,,\ \ell_2-\ell_1+1 = 2^{q-d}$, so that by (3.80) the function $h\mathbf{1}_R$ satisfies (3.108) and (3.109) for $S = L2^{q+j-d/2}$ (since $a(j) \leq S2^{-q}$ and $b(j) \leq S2^{-(q-d)}$). By (3.123), there exists $(k,\ell) \in R$ with $h(k,\ell) = 0$, so that (3.108) and (3.109) imply that (3.110) holds and by (3.111) we have

$$\begin{aligned}\Big|\sum_{i\leq N}(h\mathbf{1}_R(U_i) - \mathsf{E}(h\mathbf{1}_R))\Big| &\leq L\sqrt{pm_0}\,2^j 2^{2q-d} \\ &= L\sqrt{pm_0}\,2^j\operatorname{card} R\ .\end{aligned} \tag{3.125}$$

The case $3 \leq q \leq d$ being similar to the case $d \leq q < p$, we consider the case $q = p$ so that $R \in \mathcal{D}(p)$. We take $S = L2^{p+j-d/2}$. The difference with the case $R \in \mathcal{D}(q)$ for $q < p$ is that we have to find a new argument to prove (3.110). We have $R = \{1,\ldots,2^p\} \times \{\ell_1 2^{p-d}+1,\ldots,(\ell_1+1)2^{p-d}+1\}$. Given an integer r, define

$$R' = G \cap (\{1,\ldots,2^p\} \times \{\ell_1 2^{p-d}+1-r,\ldots,(\ell_1+1)2^{p-d}+1+r\})\ .$$

Then, for $r \leq 2^p$, we have $\operatorname{card} R' \geq 2^p r/L$, so that if $2^p r/L > V$, R' contains a point (k,ℓ') with $h(k,\ell') = 0$. Then R contains a point (k,ℓ) with $|\ell-\ell'| \leq r$, so that by (3.80) we have

$$|h(k,\ell)| \leq rb(j)\ .$$

Assuming that we chose r as small as possible with $2^p r/L > V$, we then have

$$|h(k,\ell)| \le LV2^{-p}b(j) \le LV2^{-p+j}j^{1/2\alpha} .$$

Since $V2^j \le K(\alpha)2^{2p}$, since $2^d \le 8j^{1/\alpha}$, and since $j^{1/\alpha} \le K(\alpha)2^j$, we have $LV2^{-p+j}j^{1/2\alpha} \le K(\alpha)S$, so that R contains a point (k,ℓ) with $|h(k,\ell)| \le K(\alpha)S$ and hence using (3.80) again

$$\sup_{(k,\ell)\in R} |h(k,\ell)| \le K(\alpha)S .$$

We can appeal to (3.111) to see that (3.125) still holds true if one replaces there L by $K(\alpha)$. Summation of the inequalities (3.125) for $R \in \mathcal{D}(q)$, $R \subset U(q)$ and $3 \le q \le p$ yields (3.117). □

Consider a number c_α depending on α only, that will be determined later, and the function θ given by

$$\theta(x) = 0 \text{ if } |x| < c_\alpha \text{ and } \theta(x) = |x|(\log(1+|x|))^{1/\alpha} \text{ for } |x| \ge c_\alpha .$$

Besides Theorem 3.5.3, the proof of Theorem 3.5.1 requires the following, where we keep the notation of Theorem 3.5.1. We consider the function

$$\varphi(x) = \exp|x|^\alpha - 1 .$$

Theorem 3.5.9. *One can choose c_α such that the following property holds. Consider numbers $u(k,\ell)$ for $(k,\ell) \in G = \{1,\ldots,2^p\}$, and define*

$$h(k,\ell) = \inf\bigl\{u(r,s) + \varphi(k-r) \ : \ (r,s) \in G\,,\ |\ell - s| \le 1\bigr\} . \tag{3.126}$$

Then

$$\begin{aligned} m_0 \sum_{k,\ell} \bigl(\theta(h(k+1,\ell) - h(k,\ell)) + |h(k,\ell+1) - h(k,\ell)|\bigr) \quad (3.127)\\ \le L \sum_{k,\ell} n(k,\ell)(u(k,\ell) - h(k,\ell)) . \end{aligned}$$

Proof of Theorem 3.5.1. Consider an integer $N \ge 2$ and an integer p, that will be determined later. For $(k,\ell) \in G$ we consider the point

$$a(k,\ell) = ((2k-1)2^{-p-1}\,, (2\ell-1)2^{-p-1}) .$$

These are the centers of 2^{2p} little squares $C(k,\ell)$ of side 2^{-p} that divide $[0,1]^2$. To lighten notation, for $w = (k,\ell) \in G$ we write $a(w) = a(k,\ell)$ and $C(w) = C(k,\ell)$.

Consider evenly spread points $(Y_i)_{i\le N}$, a map $\eta : \{1,\ldots,N\} \to G$ such that

$$Y_i \in C(\eta(i)) \ , \tag{3.128}$$

so that

$$d(Y_i, a(\eta(i))) \le 2^{-p+1} \tag{3.129}$$

and set

$$n(k,\ell) = \operatorname{card}\{i \le N \; ; \; \eta(i) = (k,\ell)\} \; ,$$

the number of points Y_i that belong to $C(k,\ell)$. To avoid trivial complications, we assume that no point Y_i belongs to the boundary of a little square $C(k,\ell)$, so that $\sum n(k,\ell) = N$. The points Y_i are evenly spread, so that these points are centers of non-overlapping rectangles of area $1/N$ and diameter at most $20/\sqrt{N}$. It should be clear that for N and $N2^{-2p}$ large enough, each square $C(k,\ell)$ contains about the same number of points Y_i, so that, for a certain integer m_0, we have

$$m_0 \le n(k,\ell) \le 2m_0 \; . \tag{3.130}$$

We consider N points (Z_i) of G such that exactly $n(k,\ell)$ of them are located at the point $n(k,\ell)$. We can assume by (3.129) that these points are labeled in a way that $d(Y_i, a(Z_i)) \le 2^{-p+1}$.

Consider points X_i independently uniformly distributed over $[0,1]^2$. We claim that we can find independently distributed points U_i of G such that $\mathsf{P}(U_i = (k,\ell)) = n(k,\ell)/N$ and $d(X_i, a(U_i)) \le L2^{-p}$. To see this we recall that by our definition, the fact that the points $(Y_i)_{i\le N}$ are uniformly spread means there exists a partition of $[0,1]^2$ into N rectangles $(R_i)_{i\le N}$ of area $1/N$, each with a width and a height of order $1/\sqrt{N}$, and each containing exactly one point Y_i. For $(k,\ell) \in G$ we define the domain $\mathcal{D}(k,\ell)$ as the union of the sets R_i for which $\eta(i) = (k,\ell)$, and we define $U_i = (k,\ell)$ when $X_i \in \mathcal{D}(k,\ell)$ to obtain the required points U_i.

Let us write $X_i = (X_i^1, X_i^2)$, $Y_i = (Y_i^1, Y_i^2)$ and, for $w \in G$, let us write $w = (w^1, w^2)$. Thus, by definition, if $w = a(k,\ell)$ we have $w^1 = 2^{-p-1}(2k-1)$ and $w^2 = 2^{-p-1}(2\ell - 1)$. Thus, for $j = 1,2$ we have

$$|a(U_i)^j - a(Z_{i'})^j| = 2^{-p}|U_i^j - Z_{i'}^j| \; . \tag{3.131}$$

For $j = 1,2$ we have

$$\begin{aligned}
|X_i^j - Y_{i'}^j| &\le |X_i^j - a(U_i)^j| + |a(U_i)^j - a(Z_{i'})^j| + |a(Z_{i'})^j - Y_{i'}^j| \\
&\le d(X_i, a(U_i)) + |a(U_i)^j - a(Z_{i'})^j| + d(a(Z_{i'}), Y_{i'}) \\
&\le L2^{-p} + 2^{-p}|U_i^j - Z_{i'}^j| \; ,
\end{aligned} \tag{3.132}$$

using (3.131) in the last line. We then see that to prove Theorem 3.5.1, it suffices to find p such that

$$2^{-p} \le L\sqrt{\frac{\log N}{N}} \tag{3.133}$$

and such that, with probability $\ge 1 - N^{-10}$, there is a permutation π of $\{1,\ldots,N\}$ for which

$$\sum_{i\le N}\big(\exp|U_i^1 - Z^1_{\pi(i)}|^\alpha - 1\big) \le K(\alpha)N \tag{3.134}$$

$$\forall i \le N\,,\ |U_i^2 - Z^2_{\pi(i)}| \le 1\ . \tag{3.135}$$

The reason why (3.134) suffices to obtain (3.70) is that the function $\varphi(x) = \exp|x|^\alpha - 1$ satisfies $\varphi(\lambda x) \le \lambda^\alpha \varphi(x)$ for $0 \le \lambda \le 1$.

We appeal to Proposition 3.2.1 (with $c_{ij} = \varphi(U_i^1 - Z_j^1)$ if $|U_i^2 - Z_j^2| \le 1$ and c_{ij} very large otherwise) to see the smallest value of the left-hand side of (3.134) among all permutations that satisfy (3.135) is given by

$$M_1 = \sup \sum_{i\le N}(w_i + w_i')\ , \tag{3.136}$$

where the supremum is taken over all families $(w_i), (w_i')$ such that

$$\forall i\,,\ j \le N\,,\ |U_i^2 - Z_j^2| \le 1 \Rightarrow w_i + w_j' \le \varphi(U_i^1 - Z_j^1)\ . \tag{3.137}$$

We fix families $(w_i)\,,\ (w_i')$ satisfying (3.137), and such that the supremum is attained in (3.136). We consider the function h' on G given by

$$h'(k,\ell) = \min_j\big\{\varphi(k - Z_j^1) - w_j'\ ;\ |\ell - Z_j^2| \le 1\big\}\ .$$

When $w = (k.\ell) \in G$ we define $h'(w) = h'(k,\ell)$. By (3.137) we have $h'(U_i) \ge w_i$ and thus by (3.136) we have

$$M_1 \le \sum_{i\le N}(h'(U_i) + w_i')\ .$$

For $(k,\ell) \in G$, we define

$$u(k,\ell) = -\frac{1}{n(k,\ell)}\sum\{w_i'\ ;\ Z_i = (k,\ell)\}\ .$$

When $Z_i = (k,\ell)$, we replace w_i' by $-u(k,\ell)$. In this manner we do not change $\sum_{i\le N} w_i' = -\sum_G n(k,\ell)u(k,\ell)$, while we can only increase h'. Thus

$$M_1 \le \sum_{i\le N} h(U_i) - \sum_{k,\ell} n(k,\ell)u(k,\ell) \tag{3.138}$$

where

$$h(k,\ell) = \inf\big\{\varphi(k-r) + u(r,s)\ ;\ |\ell - s| \le 1\big\}\ .$$

From (3.138) we get

$$M_1 \le \Big|\sum_{i\le N}(h(U_i) - \mathsf{E}h)\Big| - \sum_{k,\ell} n(k,\ell)(u(k,\ell) - h(k,\ell))\ . \tag{3.139}$$

Define

$$B = 2^{-2p} \sum |h(k,\ell+1) - h(k,\ell)| + 2^{-2p} \sum \xi(h(k+1,\ell) - h(k,\ell))\,, \tag{3.140}$$

and $B' = B+1$. For $\lambda > 0\,,\ \lambda < 1$ we have $\xi(\lambda x) \le \lambda \xi(x)$ so that $h/B' \in \mathcal{H}$. By Theorem 3.5.3, it is true with probability $\ge 1 - L\exp(-96p)$ that we have

$$\Big|\sum_{i \le N} (h(U_i) - \mathsf{E}h)\Big| \le K(\alpha)\sqrt{pm_0}\, 2^{2p} B'\ . \tag{3.141}$$

There exists a number $K(\alpha)$ such that $\xi(x) \le 2(\theta(x) + K(\alpha)$, so that $\theta(x) \ge \xi(x)/2 - K(\alpha))$, and by (3.127) and (3.140) we have

$$\begin{aligned} \sum n(k,\ell)(u(k,\ell) - h(k,\ell)) &\ge \frac{m_0}{L}\Big(\sum |h(k,\ell+1) - h(k,\ell)| \\ &\quad + \sum \theta\big(h(k+1,\ell) - h(k,\ell)\big)\Big) \\ &\ge \frac{m_0 2^{2p}}{L}(B - K(\alpha))\ . \end{aligned}$$

Combining with (3.139) and (3.141) we get, since $B' = B+1$,

$$\begin{aligned} M_1 &\le K(\alpha)\sqrt{pm_0}\, 2^{2p} B' - \frac{m_0}{L} 2^{2p}\big(B - K(\alpha)\big) \\ &\le B 2^{2p}\Big(K(\alpha)\sqrt{pm_0} - \frac{m_0}{L}\Big) + K(\alpha) 2^{2p}\Big(\frac{m_0}{L} + \sqrt{m_0 p}\Big)\ . \end{aligned} \tag{3.142}$$

Thus we see that if we have chosen p so that the first term is negative, and if $p \le m_0$, then (3.142) implies as desired that $M_1 \le K(\alpha) m_0 2^{2p} \le K(\alpha)N$, recalling (3.76). To ensure that $K(\alpha)\sqrt{pm_0} \le m_0/L$, it suffices to ensure that $p \le m_0/K(\alpha)$ (so that in particular $p \le m_0$ as required) and using (3.76) again, that $p2^{-2p} \le N/K(\alpha)$, i.e. $2^{-2p} \le N/(K(\alpha)\log N)$. Taking p as small as possible that satisfies this condition, for large N we have $96p \ge 11\log N$, and $L\exp(-96p) \le 1 - N^{-10}$. □

We turn to the proof of Theorem 3.5.9. This proof requires a significant amount of work of an elementary nature. This work is not related to the main theme of the book. We recall the function $\varphi(x) = \exp|x|^\alpha - 1$.

Lemma 3.5.10. *Consider numbers $(v_k)_{k \le 2^p}$ and for $k \le 2^p$ define*

$$g(k) = \inf\big\{v_r + \varphi(k-r)\ ;\ 1 \le r \le 2^p\big\}\,. \tag{3.143}$$

Then we have

$$\sum_{k < 2^p} \theta(g(k+1) - g(k)) \le 16 \sum_{k \le 2^p} (v_k - g(k))\ . \tag{3.144}$$

Proof. For $y \ge 0$ we write $\varphi^{-1}(y) = (\log(1+y))^{1/\alpha}$, so that we have $\varphi(\varphi^{-1}(y)) = y$. For $x \ge c_\alpha$ we have $\theta(x) = x\varphi^{-1}(x)$. Consider the set $A \subset \{1, \ldots, 2^p - 1\}$ consisting of the integers k for which $g(k+1) - g(k) \ge c_\alpha$.

For $k \in A$, we define $x_k = k+1-\varphi^{-1}(g(k+1)-g(k))$, so that $\varphi(k+1-x_k) = g(k+1)-g(k)$. Our first goal is to prove that

$$x_k \le m \le k \Rightarrow g(m) \le g(k)\ . \tag{3.145}$$

To this aim consider $1 \le r \le 2^p$ with $g(k) = v_r + \varphi(k-r)$. Thus

$$v_r + \varphi(k-r) = g(k) \le g(k+1) \le v_r + \varphi(k+1-r) \tag{3.146}$$

and hence $\varphi(k-r) \le \varphi(k+1-r)$, so that $r \le k$. Also, using (3.146) in the first inequality, we have

$$\varphi(k+1-x_k) = g(k+1) - g(k) \le \varphi(k+1-r) - \varphi(k-r) \le \varphi(k+1-r)\,,$$

so that $x_k \ge r \ge 1$. For $r \le m \le k$, we have

$$g(m) \le v_r + \varphi(m-r) \le v_r + \varphi(k-r) = g(k)\ ,$$

and this proves (3.145). Consider, for $k \in A$, the domain

$$\begin{aligned} D_k = \big\{&(m,y)\ ;\ m \in \{1,\ldots,2^p\}\,,\, y \in \mathbb{R}\,,\, x_k \le m \le k\,,\\ &g(k) \le y \le g(k+1) - \varphi(k+1-m)\big\}\ . \end{aligned}$$

We show that for $k\,,\, s \in A\,,\, k \ne s$, we have $D_k \cap D_s = \emptyset$. To see this, we can assume for definiteness that $s < k$. If $s < x_k$, we obviously have $D_k \cap D_s = \emptyset$. If $s \ge x_k$, by (3.145) and since $s+1 \le k$, we have $g(s+1) \le g(k)$. Now, if $(m,y) \in D_k$, we have $y \ge g(k)$, while if $(m,y) \in D_s$ we have $y < g(s+1) \le g(k)$.

If $(m,y) \in D_k$, and since $g(k+1) \le v_m + \varphi(k+1-m)$, we have $y \le v_m$, so that

$$D_k \subset \varDelta = \big\{(m,y)\ ;\ m \in \{1,\ldots,2^p\}\,,\, g(m) \le y \le v_m\big\}\ .$$

Let us denote by μ the measure on $\{1,\ldots,2^p\} \times \mathbb{R}$ such that its restriction to each line $\{m\} \times \mathbb{R}$ is the Lebesgue measure on that line. Since the sets $(D_k)_{k \in A}$ are disjoint, we have

$$\sum_{k \in A} \mu(D_k) \le \mu(\varDelta) = \sum_{k \le 2^p} (v_k - g(k))\ . \tag{3.147}$$

We proceed now to prove that if c_α has been suitably chosen, we have

$$\mu(D_k) \ge \frac{1}{8}\theta(g(k+1)-g(k))\ . \tag{3.148}$$

First, we observe that by definition of μ we have

$$\mu(D_k) \ge \sum \big(g(k+1) - g(k) - \varphi(k+1-m)\big)\ , \tag{3.149}$$

where the summation is over the integers m for which $x_k \le m \le k$. Calculus shows that $\varphi(z)/z$ increases for $z \ge d_\alpha := (1/\alpha)^{1/\alpha}$ so for $x_k \le m \le k+1-d_\alpha$ we have

$$\frac{\varphi(k+1-m)}{k+1-m} \le \frac{\varphi(k+1-x_k)}{k+1-x_k} = \frac{g(k+1)-g(k)}{k+1-x_k}$$

and thus

$$g(k+1)-g(k)-\varphi(k+1-m) \ge (g(k+1)-g(k))\frac{m-x_k}{k+1-x_k}$$

Now,

$$m \ge \frac{1}{2}(x_k+k+1) \Rightarrow \frac{m-x_k}{k+1-x_k} \ge \frac{1}{2}\,. \tag{3.150}$$

The number of integers contained in a closed interval of length y is greater than $y-1$. Thus the number of values of m with $(x_k+k+1)/2 \le m \le k+1-d_\alpha$ is greater than

$$\begin{aligned} k+1-d_\alpha-\frac{x_k+k+1}{2}-1 &= \frac{k+1-x_k}{2}-d_\alpha-1 \\ &\ge \frac{k+1-x_k}{4} \end{aligned}$$

because $k+1-x_k = g(k+1)-g(k) \ge \varphi^{-1}(c_\alpha)$, and provided c_α has been chosen such that $c_\alpha \ge \varphi(4(d_\alpha+1))$. Thus it follows form (3.149) that

$$\begin{aligned} \mu(D_k) &\ge \frac{1}{8}(k+1-x_k)(g(k+1)-g(k)) \\ &= \frac{1}{8}(g(k+1)-g(k))\varphi^{-1}(g(k+1)-g(k)) \end{aligned}$$

which is (3.148). Combining (3.148) and (3.147), and considering similarly the set B where $g(k)-g(k+1) \ge c_\alpha$ finishes the proof. □

We recall Definition 3.143.

Lemma 3.5.11. *Consider numbers* $(v_k)_{k\le 2^p}$, $(v'_k)_{k\le 2^p}$, *and the numbers* $g'(k)$ *defined from the sequence* (v'_k) *the way the numbers* $g(k)$ *are defined from the sequence* (v_k). *Then we have*

$$\sum_{k\le 2^p} |g(k)-g'(k)| \le \sum_{k\le 2^p} \left(v_k+v'_k-g(k)-g'(k)+|v_k-v'_k|\right)\,. \tag{3.151}$$

Proof. Since $\varphi(0)=0$ we have $g(k)\le v_k$ and $g'(k)\le v'_k$. If $g'(k)\ge g(k)$, we have

$$\begin{aligned} g'(k)-g(k) \le v'_k-g(k) &= v'_k-v_k+v_k-g(k) \\ &\le |v'_k-v_k|+v_k-g(k)+v'_k-g'(k)\,. \end{aligned}$$

A similar argument when $g(k)\ge g'(k)$ and summation finish the proof. □

We consider numbers $u(k,\ell)$ for $(k,\ell) \in G$, and $h(k,\ell)$ as in (3.126). We set

$$v(k,\ell) = \min\{u(k,s)\ ;\ |\ell - s| \le 1\}\ . \tag{3.152}$$

Thus we have

$$h(k,\ell) = \inf\bigl\{v(r,\ell) + \varphi(k-r)\ ;\ 1 \le r \le 2^p\bigr\}\ . \tag{3.153}$$

Lemma 3.5.12. *We have*

$$m_0 \sum_{k \le 2^p, \ell < 2^p} |v(k,\ell+1) - v(k,\ell)| \le 10 \sum_{k,\ell \le 2^p} n(k,\ell)(u(k,\ell) - v(k,\ell))\ . \tag{3.154}$$

Proof. We observe that $|a-b| = a + b - 2\min(a,b)$, and that

$$\begin{aligned} v(k,\ell) &\le \min(u(k,\ell+1), u(k,\ell)) \\ v(k,\ell+1) &\le \min(u(k,\ell+1), u(k,\ell))\ . \end{aligned}$$

Thus

$$\begin{aligned} |u(k,\ell+1) - u(k,\ell)| &= u(k,\ell) + u(k,\ell+1) - 2\min(u(k,\ell+1)\,,\, u(k,\ell)) \\ &\le u(k,\ell) - v(k,\ell) + u(k,\ell+1) - v(k,\ell+1)\ . \end{aligned}$$

By summation we get

$$\sum_{k \le 2^p, \ell < 2^p} |u(k,\ell+1) - u(k,\ell)| \le 2 \sum_{k,\ell \le 2^p} (u(k,\ell) - v(k,\ell))$$

and since $m_0 \le n(k,\ell)$,

$$m_0 \sum_{k \le 2^p, \ell < 2^p} |u(k,\ell+1) - u(k,\ell)| \le 2 \sum_{k,\ell \le 2^p} n(k,\ell)(u(k,\ell) - v(k,\ell))\ . \tag{3.155}$$

Now

$$|v(k,\ell) - u(k,\ell)| \le |u(k,\ell+1) - u(k,\ell)| + |u(k,\ell-1) - u(k,\ell)|$$

so that

$$\begin{aligned} |v(k,\ell+1) - v(k,\ell)| &\le |v(k,\ell+1) - u(k,\ell+1)| + |u(k,\ell+1) - u(k,\ell)| \\ &\quad + |u(k,\ell) - v(k,\ell)| \\ &\le |u(k,\ell) - u(k,\ell-1)| + 3|u(k,\ell+1) - u(k,\ell)| \\ &\quad + |u(k,\ell+2) - u(k,\ell+1)|\ . \end{aligned}$$

Then (3.154) follows by summation from (3.155). □

Proof of Theorem 3.5.9. Given $1 \le \ell < 2^p$, we use Lemma 3.5.11 for $v_k = v(k,\ell)$, and $v'_k = v(k,\ell+1)$, where $v(k,\ell)$ is given by (3.152). Thus $g(k) = h(k,\ell)$ and $g'(k) = h(k,\ell+1)$. Summing the inequalities (3.151) for $1 \le k \le 2^p$ we get

$$\begin{aligned}\sum_{k \le 2^p, \ell < 2^p} |h(k,\ell+1) - h(k,\ell)| \le 2 &\sum_{k,\ell} (v(k,\ell) - h(k,\ell)) \\ &+ \sum_{k,\ell} |v(k,\ell) - v(k,\ell+1)| \,.\end{aligned}$$

Using (3.154), and since $m_0 \le n(k,\ell)$ we get

$$\begin{aligned}m_0 \sum_{k \le 2^p, \ell < 2^p} |h(k,\ell+1) - h(k,\ell)| &\le 2 \sum_{k,\ell} n(k,\ell)(v(k,\ell) - h(k,\ell)) \\ &\quad + 10 \sum_{k,\ell} n(k,\ell)(u(k,\ell) - v(k,\ell)) \\ &\le 10 \sum_{k,\ell} n(k,\ell)(u(k,\ell) - h(k,\ell)) \,,\end{aligned}$$

using that $h(k,\ell) \le v(k,\ell) \le u(k,\ell)$ in the last line. On the other hand, summing the inequalities (3.144) for $\ell \le 2^p$, we get

$$\begin{aligned}\sum_{k < 2^p, \ell \le 2^p} \theta(h(k+1,\ell) - h(k,\ell)) &\le 16 \sum_{k,\ell} (v(k,\ell) - h(k,\ell)) \\ &\le 16 \sum_{k,\ell} (u(k,\ell) - h(k,\ell))\end{aligned}$$

and thus

$$m_0 \sum_{k < 2^p, \ell \le 2^p} \theta(h(k+1,\ell) - h(k,\ell)) \le 16 \sum_{k,\ell} n(k,\ell)(u(k,\ell) - h(k,\ell)) \,.$$

$\square$

4 The Bernoulli Conjecture

4.1 The Conjecture

Gaussian r.v. are arguably the central object of Probability theory, but Bernoulli (= coin-flipping) r.v. are also very useful. (Thus, if ϵ is a Bernoulli r.v., $P(\epsilon = \pm 1) = 1/2$.)

Consider a subset T of ℓ^2, and i.i.d. Bernoulli r.v. $(\epsilon_i)_{i \geq 1}$. We set

$$b(T) = \mathsf{E} \sup_{t \in T} \sum_{i \geq 1} t_i \epsilon_i \ . \tag{4.1}$$

We observe that $b(T) \geq 0$, that $b(T) \leq b(T')$ if $T \subset T'$, and that $b(T + t_0) = b(T)$.

We would like to understand the value of $b(T)$ from the geometry of T. We denote by $\|t\|_1 = \sum_{i \geq 1} |t_i|$ the ℓ^1 norm of t, and by B_1 the unit ball of ℓ^1. The following is trivial.

Proposition 4.1.1. *We have*

$$b(T) \leq \sup_{t \in T} \|t\|_1 \ . \tag{4.2}$$

We recall the notation $g(T) = \mathsf{E} \sup_{t \in T} \sum_{i \geq 1} t_i g_i$. Here is another way to control $b(T)$.

Proposition 4.1.2. *We have*

$$b(T) \leq \sqrt{\frac{\pi}{2}} g(T) \ . \tag{4.3}$$

Proof. If $(\epsilon_i)_{i \geq 1}$ is an i.i.d. Bernoulli sequence that is independent of the sequence $(g_i)_{i \geq 1}$, then the sequence $(\epsilon_i |g_i|)_{i \geq 1}$ is i.i.d. standard normal. Thus

$$g(T) = \mathsf{E} \sup_{t \in T} \sum_{i \geq 1} \epsilon_i |g_i| t_i \ .$$

Using Jensen's inequality to integrate in the r.v. g_i inside the supremum rather than outside, we get

$$g(T) \geq \sqrt{\frac{2}{\pi}} \mathsf{E} \sup_{t \in T} \sum_{i \geq 1} t_i \epsilon_i = \sqrt{\frac{2}{\pi}} b(T)$$

since $\mathsf{E}|g_i| = \sqrt{2/\pi}$. □

Thus, besides (4.2), another way for $b(T)$ to be small is that $g(T)$, or, equivalently, $\gamma_2(T)$ is small. The Bernoulli conjecture expresses that the only way $b(T)$ can be small is from a mixture of the two previous situations.

Conjecture 4.1.3. (The Bernoulli conjecture). There exists a universal constant L such that given any subset T of ℓ^2, we can find two subsets T_1 and T_2 of ℓ^2 with

$$\gamma_2(T_1) \leq Lb(T) \tag{4.4}$$

$$T_2 \subset Lb(T)B_1 \text{ , i.e. } t \in T_2 \Rightarrow \|t\|_1 \leq Lb(T) \tag{4.5}$$

$$T \subset T_1 + T_2 \ . \tag{4.6}$$

The decomposition (4.6) would give a very explicit reason why $b(T) < \infty$, since it implies $b(T) \leq \sqrt{\pi/2} g(T_1) + \sup_{t \in T_2} \|t\|_1 \leq L\gamma_2(T_1) + \sup_{t \in T_2} \|t\|_1$. Let us remind the reader that there is a \$ 5000 prize offered by the author for a positive solution of this conjecture.

Using Theorem 2.6.2, we see that we get an equivalent conjecture if we also request that $\gamma_1(T_1, d_\infty) \leq Lb(T)$.

One intrinsic difficulty in attacking the Bernoulli conjecture is that the decomposition (4.6), when it exists, is neither unique nor canonical.

4.2 Control in ℓ^∞ Norm

The main result of this section is as follows.

Theorem 4.2.1. *There exists a universal constant L such that for any subset T of ℓ^2 we have*

$$\gamma_2(T) \leq L\big(b(T) + \sqrt{b(T)\gamma_1(T, d_\infty)}\big) \ . \tag{4.7}$$

In particular, if $\gamma_1(T, d_\infty) \leq Lb(T)$, we have $\gamma_2(T) \leq Lb(T) \leq L'\gamma_2(T)$. Our main tool is as follows.

Proposition 4.2.2. *There exists constants L_1 and L_2 with the following properties. Consider numbers $a\,, b\,, \sigma > 0$, vectors $t_1\,, \cdots, t_m \in \ell^2$, and assume that*

$$\ell \neq \ell' \Rightarrow \|t_\ell - t_{\ell'}\|_2 \geq a \ . \tag{4.8}$$

Assume moreover that

$$\forall \ell \leq m\,, \ \|t_\ell\|_\infty \leq b \ . \tag{4.9}$$

For $\ell \le m$ consider sets H_ℓ with $H_\ell \subset B_2(t_\ell, \sigma)$. Then

$$b\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge \frac{1}{L_1}\min\Big(a\sqrt{\log m}, \frac{a^2}{b}\Big) - L_2\sigma\sqrt{\log m} + \min_{\ell\le m} b(H_\ell)\ . \tag{4.10}$$

Corollary 4.2.3. *There exists a constant L_0 such that if the points t_ℓ satisfy (4.8) and $t_\ell \in D$ with $\Delta(D, d_\infty) \le 4a/\sqrt{\log m}$, and if $H_\ell \subset B_2(t_\ell, a/L_0)$, we have*

$$b\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge \frac{a}{L_0}\sqrt{\log m} + \min_{\ell\le m} b(H_\ell)\ . \tag{4.11}$$

Of course the factor 4 in the condition $\Delta(D, d_\infty) \le 4a/\sqrt{\log m}$ can be removed. Its only purpose is that it find it convenient later to use the exact statement given here.

Proof. We observe that without loss of generality we can assume that $t_1 = 0$, so that $\|t_\ell\|_\infty \le b = 4a/\sqrt{\log m}$ for all $\ell \le m$ and (4.10) used for $\sigma = a/L_0$ gives

$$b\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge \frac{1}{4L_1}a\sqrt{\log m} - \frac{aL_2}{L_0}\sqrt{\log m} + \min_{\ell\le m} b(H_\ell)\ ,$$

so that if $L_0 \ge 8L_1L_2$ and $L_0 \ge 8L_1$ we get (4.11). □

The proof of Proposition 4.2.2 is identical to that of Proposition 2.1.4, if one replaces Lemmas 2.1.2 and 2.1.3 respectively by the following principles.

Theorem 4.2.4. *(Sudakov minoration for Bernoulli processes [53], Proposition 2.2). For $t_1, \cdots, t_m$ in ℓ^2 that satisfy (4.8) and (4.9), we have*

$$\mathsf{E}\sup_{\ell\le m}\sum_{i\ge 1} t_{\ell,i}\epsilon_i \ge \frac{1}{L}\min\Big(a\sqrt{\log m}, \frac{a^2}{b}\Big)\ . \tag{4.12}$$

Theorem 4.2.5. *([61] Theorem 8.2, or [18]) If $T \subset B(t, \sigma)$ then*

$$\forall u > 0\,,\ \mathsf{P}\Big(\Big|\sup_{t\in T}\sum_{i\ge 1} t_i\epsilon_i - b(T)\Big| \ge u\Big) \le L\exp\Big(-\frac{u^2}{L\sigma^2}\Big)\ . \tag{4.13}$$

Remark 4.2.6. In our constructions, we will not only have the information (4.8), but we will also know that for a certain s,

$$\forall \ell \le m\,,\ t_\ell \in B_2(s, ra)\ .$$

Only minor changes are required to set things in a way that (4.12) is only needed under this extra information. In that case this follows directly from the main result of [67] (combined with the Sudakov minoration for Gaussian processes), the proof of which is simpler and more elegant than that of [53] (but one can also deduce (4.12) from the result of [67], combined with iteration).

Let us note a simple fact.

Lemma 4.2.7. *For a subset T of ℓ_2 we have*

$$\Delta(T, d_2) \leq Lb(T) . \tag{4.14}$$

Proof. Assuming without loss of generality that $0 \in T$, we have

$$\begin{aligned}\forall t \in T\,,\; b(T) &\geq \mathsf{E}\max\Big(0, \sum_{i\geq 1}\epsilon_i t_i\Big) \\ &= \frac{1}{2}\mathsf{E}\Big|\sum_{i\geq 1}\epsilon_i t_i\Big| \geq \frac{1}{L}\|t\|_2 \,,\end{aligned}$$

using symmetry in the equality and Khintchine's inequality in the last inequality, and this proves (4.14). □

Proof of Theorem 4.2.1. We consider an integer $\tau \geq 1$ to be specified later, and an admissible sequence of partitions $(\mathcal{D}_n)$ of T such that

$$\sup_{t\in T}\sum_{p\geq 0} 2^p \Delta(D_p(t), d_\infty) \leq 2\gamma_1(T, d_\infty) . \tag{4.15}$$

The proof will rely on the application of Theorem 1.3.2 to the functionals

$$F_n(A) = \sup\big\{b(A\cap D) + U_n(D)\,,\; D \in \mathcal{D}_{n+\tau}\,,\; A\cap D \neq \emptyset\big\}\,,$$

where

$$U_n(D) = \sup_{t\in D}\sum_{p\geq n} 2^p \Delta(D_{p+\tau}(t), d_\infty) .$$

We now check that these functionals satisfy the growth condition of Definition 1.3.1 for a suitable value of the parameters. Consider $m = N_{n+\tau+1}$ points $t_1\,, \ldots\,, t_m$ of T such that

$$\ell \neq \ell' \Rightarrow \|t_\ell - t_{\ell'}\|_2 \geq a\,, \tag{4.16}$$

and consider sets $H_\ell \subset B_2(t_\ell, a/r)$, where $r = 8L_0\,,\ L_0 \geq 1$ being the constant of Corollary 4.2.3.

Consider $c < \min_{\ell\leq m} F_{n+1}(H_\ell)$, and for each ℓ consider $D_\ell \in \mathcal{D}_{n+\tau+1}$ such that $H_\ell \cap D_\ell \neq \emptyset$ and

$$b(H_\ell \cap D_\ell) + U_{n+1}(D_\ell) > c . \tag{4.17}$$

Each of the m sets D_ℓ is contained in one of the sets of $\mathcal{D}_{n+\tau}$. Since $m = N_{n+\tau+1} = N_{n+\tau}^2 \geq N_{n+\tau}\cdot \mathrm{card}\mathcal{D}_{n+\tau}$, by the pigeon hole principle we can find $D \in \mathcal{D}_{n+\tau}$ such that if

$$I = \{\ell \leq m\,;\; D_\ell \subset D\}$$

then $\operatorname{card} I \geq N_{n+\tau}$. We have

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq b\Big(D \cap \bigcup_{\ell \in I} H_\ell\Big) + U_n(D) \,. \tag{4.18}$$

Now, for each $\ell \in I$, we have

$$U_n(D) = 2^n \Delta(D, d_\infty) + U_{n+1}(D) \geq 2^n \Delta(D, d_\infty) + U_{n+1}(D_\ell) \,. \tag{4.19}$$

Case 1. We have $\Delta(D, d_\infty) \geq a 2^{-n/2}$. Then (4.17), (4.18) and (4.19) show that if ℓ_0 is an arbitrary element of I, we have

$$\begin{aligned} F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) &\geq 2^{n/2} a + b(D_{\ell_0} \cap H_{\ell_0}) + U_{n+1}(D_{\ell_0}) \\ &\geq 2^{n/2} a + c \,, \end{aligned}$$

using (4.17) for $\ell = \ell_0$, and thus

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq 2^{n/2} a + \inf_{\ell \leq m} F_{n+1}(H_\ell) \,. \tag{4.20}$$

Case 2. We have $\Delta(D, d_\infty) \leq a 2^{-n/2}$, and thus $\Delta(D, d_\infty) \leq a/\sqrt{\log N_n}$. We select an arbitrary subset J of I with $\operatorname{card} J = N_n$. For $\ell \in J$ we choose arbitrarily $u_\ell \in H_\ell \cap D_\ell \subset D$, so that, since $H_\ell \subset B_2(t_\ell, a/r)$, we have $H_\ell \subset B_2(u_\ell, 2a/r) = B_2(u_\ell, a/(4L_0))$ since $r = 8L_0$. We observe that, since $r \geq 4$, by (4.16) we have $d_2(u_\ell, u_{\ell'}) \geq a/4$ for $\ell \neq \ell'$.

We use Corollary 4.2.3 with $m = N_n$, $H_\ell \cap D_\ell$ instead of H_ℓ, $a/4$ instead of a and u_ℓ instead of t_ℓ to see that

$$\begin{aligned} b\Big(D \cap \bigcup_{\ell \in I} H_\ell\Big) &\geq b\Big(\bigcup_{\ell \in J} (H_\ell \cap D_\ell)\Big) \\ &\geq \frac{a}{4L_0} \sqrt{\log N_n} + \inf_{\ell \in J} b(H_\ell \cap D_\ell) \,. \end{aligned}$$

Combining with (4.17), (4.18) and (4.19) we get

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq \frac{2^{n/2} a}{L} + \inf_{\ell \in J} F_{n+1}(H_\ell) \geq \frac{2^{n/2} a}{L} + \inf_{\ell \leq m} F_{n+1}(H_\ell) \,. \tag{4.21}$$

Thus, this relation holds, whichever of the preceding cases occur. That is, we have proved that the growth condition of Definition 1.3.1 holds with $\theta(n) = 2^{n/2}/L$, $\tau + 1$ instead of τ and $\beta = 1$ and we can apply Theorem 1.3.2 for these values of the parameters. By definition we have

$$F_0(T) \leq b(T) + U_0(T)$$

and by (4.15) we have $2^\tau U_0(T) \leq 2\gamma_1(T, d_\infty)$, so that

$$F_0(T) \le b(T) + 2^{-\tau+1}\gamma_1(T, d_\infty)\,.$$

Since $\Delta(T, d_2) \le Lb(T)$ by (4.14), we deduce from Lemma 1.3.3 that

$$\gamma_2(T) \le L2^{\tau/2}\big(b(T) + 2^{-\tau}\gamma_1(T, d_\infty)\big)$$

and Theorem 4.2.1 follows by optimization over $\tau \ge 1$. □

A striking application will be given in Section 6.2.

4.3 Chopping Maps and the Weak Solution

The purpose of this section is to prove the following weak form of the Bernoulli conjecture, where B_p denotes the unit ball of ℓ^p,

$$B_p = \Big\{t\ ;\ \sum_{i\ge 1}|t_i|^p \le 1\Big\}\,.$$

Theorem 4.3.1. *Given $p > 1$, there exists a number $K(p) < \infty$ such that, given $T \subset \ell^2$, we can find two sets T_1, $T_2 \subset \ell^2$ with $T \subset T_1 + T_2$,*

$$\gamma_2(T_1) \le K(p)b(T)\,. \tag{4.22}$$

$$T_2 \subset K(p)b(T)B_p\,. \tag{4.23}$$

Besides the fact that this result provides support for the Bernoulli conjecture, it does have striking applications to Banach Space Theory (Section 6.1).

The most successful idea to date about Bernoulli processes is that of **chopping maps**.

Definition 4.3.2. *Given $c > 0$ we define the chopping map $\psi_c : \mathbb{R} \to \mathbb{R}^{\mathbb{Z}}$ as follows. We have $\psi_c(x) = (\psi_{c,j}(x))_{j\in\mathbb{Z}}$, where, if $x \ge 0$,*

$$\begin{aligned}
\psi_{c,j}(x) &= 0 && \text{if} \quad j < 0\\
\psi_{c,j}(x) &= c && \text{if} \quad c(j+1) \le x\ ,\ j \ge 0\\
\psi_{c,j}(x) &= x - cj && \text{if} \quad cj \le x \le c(j+1)\ ,\ j \ge 0\\
\psi_{c,j}(x) &= 0 && \text{if} \quad x \le cj\ ,\ j \ge 0\,.
\end{aligned}$$

If $x < 0$, then $\psi_{c,j}(x) = -\psi_{c,-j}(-x)$.

In other words, x is “chopped” in pieces of length c that are laid side to side.

We define the chopping map $\Psi_c : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}\times\mathbb{Z}}$ component-wise

$$\Psi_c((t_i)_{i\ge 1}) = (\psi_{c,j}(t_i))_{i\in\mathbb{N},j\in\mathbb{Z}}\,. \tag{4.24}$$

Thus $\|\Psi_c(t)\|_\infty \le c$.

It should be obvious that (after suitably renumbering the coordinates) we have, for $k \in \mathbb{N}$,

$$\Psi_{c/k} = \Psi_{c/k} \circ \Psi_c \,. \tag{4.25}$$

In particular, if q is an integer, we have

$$\Psi_{q^{-j-1}} = \Psi_{q^{-j-1}} \circ \Psi_{q^{-j}} \tag{4.26}$$

for all $j \in \mathbb{Z}$.

The following is obvious.

Lemma 4.3.3. *If $x\,,\, y \in \mathbb{R}$, then*

$$|x-y| = \sum_{j\in\mathbb{Z}} |\psi_{c,j}(x) - \psi_{c,j}(y)| \tag{4.27}$$

and thus

$$|x-y|^2 \ge \sum_{j\in\mathbb{Z}} (\psi_{c,j}(x) - \psi_{c,j}(y))^2 \,. \tag{4.28}$$

One of the reasons for the usefulness of chopping maps is their interplay with the ℓ^1 and ℓ^2 norms.

Lemma 4.3.4. *(a) If $|x-y| \le c$, we have*

$$|x-y|^2 \le 4\sum_{j\in\mathbb{Z}} (\psi_{c,j}(x) - \psi_{c,j}(y))^2 \le 4|x-y|^2 \,. \tag{4.29}$$

(b) If $|x-y| \ge c$, we have

$$c|x-y| \le 4\sum_{j\in\mathbb{Z}} (\psi_{c,j}(x) - \psi_{c,j}(y))^2 \le 8c|x-y| \,. \tag{4.30}$$

Proof. The right-hand side inequality of (4.29) follows from (4.28). To prove the left-hand side inequality we note that when $|x-y| \le c$, at most two of the terms $|\psi_{c,j}(x) - \psi_{c,j}(y)|$ are not zero, and by (4.27) one of them is at least $|x-y|/2$.

To prove the right-hand side inequality of (4.30) we use (4.27) and that, since $|\psi_{c,j}(x) - \psi_{c,j}(y)| \le 2c$, we have

$$(\psi_{c,j}(x) - \psi_{c,j}(y))^2 \le 2c|\psi_{c,j}(x) - \psi_{c,j}(y)| \,.$$

To prove the left-hand side inequality, we note that there are at most two indices, say j_1 and j_2, for which $u_j = |\psi_{c,j}(x) - \psi_{c,j}(y)|$ satisfies $0 < u_j < c$. Thus

$$\sum_{j\ne j_1,j_2} u_j^2 \ge c \sum_{j\ne j_1,j_2} u_j \,.$$

We are done if $\sum_{j\neq j_1,j_2} u_j \geq |x-y|/4$. Otherwise, by (4.27), we have $u_{j_1} + u_{j_2} \geq 3|x-y|/4$, so that

$$u_{j_1}^2 + u_{j_2}^2 \geq \frac{1}{2}(u_{j_1}+u_{j_2})^2 \geq \frac{9}{32}|x-y|^2 \geq \frac{1}{4}c|x-y| \ .$$

□

Corollary 4.3.5. *(a) For $x\,,\,y\in\mathbb{R}$ we have*

$$|x-y|^2\mathbf{1}_{\{|x-y|<c\}} + c|x-y|\mathbf{1}_{\{|x-y|\geq c\}} \leq 4\sum_{j\in\mathbb{Z}}|\psi_{c,j}(x)-\psi_{c,j}(y)|^2 \tag{4.31}$$

$$\sum_{j\in\mathbb{Z}}|\psi_{c,j}(x)-\psi_{c,j}(y)|^2 \leq |x-y|^2\mathbf{1}_{\{|x-y|<c\}} + 2c|x-y|\mathbf{1}_{\{|x-y|\geq c\}} \ . \tag{4.32}$$

(b) If $s\,,\,t\in\ell^2$, we have

$$\sum_{i\geq 1}(s_i-t_i)^2\mathbf{1}_{\{|s_i-t_i|<c\}} + c\sum_{i\geq 1}|s_i-t_i|\mathbf{1}_{\{|s_i-t_i|\geq c\}} \leq 4\|\Psi_c(s)-\Psi_c(t)\|_2^2 \tag{4.33}$$

$$\|\Psi_c(s)-\Psi_c(t)\|_2^2 \leq \sum_{i\geq 1}(s_i-t_i)^2\mathbf{1}_{\{|s_i-t_i|<c\}} + 2c\sum_{i\geq 1}|s_i-t_i|\mathbf{1}_{\{|s_i-t_i|\geq c\}} \ . \tag{4.34}$$

Proof. To prove (4.31) we use the left-hand side of (4.29) if $|x-y|\leq c$ and the left-hand side of (4.30) otherwise. Then (4.33) follows by summation of (4.31) over the coordinates. To prove (4.32) we use the right-hand side of (4.29) if $|x-y|\leq c$ and the right-hand side of (4.30) otherwise, and again (4.34) follows by summation. □

The chopping map Ψ_c sends $\ell^2(\mathbb{N})$ to $\ell^2(\mathbb{N}\times\mathbb{Z})$. In the following B_2 denotes either the unit ball of $\ell^2(\mathbb{N})$ or of $\ell^2(\mathbb{N}\times\mathbb{Z})$, and similarly for B_1.

Corollary 4.3.6. *If $t\in\ell^2$ we have*

$$\Psi_c\Big(t+\epsilon B_2+\frac{\epsilon^2}{c}B_1\Big) \subset \Psi_c(t)+4\epsilon B_2 \tag{4.35}$$

$$\Psi_c^{-1}(\Psi_c(t)+\epsilon B_2) \subset t+2\epsilon B_2+\frac{4\epsilon^2}{c}B_1 \ . \tag{4.36}$$

Proof. Consider $s\in\ell^2$ and $t=s+u$ where $u\in(\epsilon^2/c)B_1$. We write $u=v+w$ where

$$v_i = u_i\mathbf{1}_{\{|u_i|<c\}}\ ,\ w_i = u_i\mathbf{1}_{\{|u_i|\geq c\}} \ . \tag{4.37}$$

Thus $\sum_{i\geq 1} v_i^2 \leq c\sum_{i\geq 1}|u_i| \leq \epsilon^2$ and $\sum_{i\geq 1}|w_i| \leq \epsilon^2/c$. It then follows from (4.34) that $\|\Psi_c(t)-\Psi_c(s)\|_2^2 \leq 4\epsilon^2$ so that

$$\Psi_c\Big(s+\frac{\epsilon^2}{c}B_1\Big) \subset \Psi_c(s)+2\epsilon B_2 \ .$$

If now $u \in \epsilon B_2$, and v, w are as above, we have $\sum_{i\geq 1} v_i^2 \leq \epsilon^2$ and $\sum_{i\geq 1} |w_i| \leq \sum_{i\geq 1} u_i^2/c \leq \epsilon^2/c$, and, as before

$$\Psi_c(s + \epsilon B_2) \subset \Psi_c(s) + 2\epsilon B_2 \ .$$

This proves (4.35).

To prove (4.36), we consider $u \in \ell^2$ and the decomposition $u = v + w$ as before. If $t + u \in \Psi_c^{-1}(\Psi_c(t) + \epsilon B_2)$, then $\|\Psi_c(t+u) - \Psi_c(t)\|_2 \leq \epsilon$, and we appeal to (4.33) to see that

$$\|v\|_2 \leq 2\epsilon \, , \, \|w\|_1 \leq \frac{4\epsilon^2}{c} \, ,$$

so that $t + u \in t + 2\epsilon B_2 + (4\epsilon^2/c)B_1$. □

Besides the fact that the chopping maps have the remarkable behavior of Corollary 4.3.6, it is central for our approach that they "decrease $b(T)$".

Proposition 4.3.7. *If $T \subset \ell^2$, then*

$$b(\Psi_c(T)) \leq b(T) \ . \tag{4.38}$$

Proof. We consider i.i.d. Bernoulli r.v. $(\epsilon_i)_{i\geq 1}$, $(\epsilon_{ij})_{i\in\mathbb{N},j\in\mathbb{Z}}$. The double sequences (ϵ_{ij}) and $(\epsilon_i\epsilon_{ij})$ have the same distribution, so that

$$\begin{aligned} b(\Psi_c(T)) &= \mathsf{E} \sup_{t\in T} \sum_{i\in\mathbb{N},j\in\mathbb{Z}} \epsilon_{ij}\psi_{c,j}(t_i) \qquad (4.39) \\ &= \mathsf{E} \sup_{t\in T} \sum_{i\in\mathbb{N},j\in\mathbb{Z}} \epsilon_i\epsilon_{ij}\psi_{c,j}(t_i) \\ &= \mathsf{E}\Big(\mathsf{E}_\epsilon \sup_{t\in T} \sum_{i\geq 1} \epsilon_i\theta_i(t_i)\Big) \end{aligned}$$

where $\theta_i(x) = \sum_{j\in\mathbb{Z}} \epsilon_{ij}\psi_{c,j}(x)$, and where E_ϵ means averaging only in $(\epsilon_i)_{i\geq 1}$. We note that θ_i is a contraction, since

$$|\theta_i(x) - \theta_i(y)| \leq \sum_{j\in\mathbb{Z}} |\psi_{c,j}(x) - \psi_{c,j}(y)| \leq |x - y|$$

by (4.27). The key point is then the comparison theorem for Bernoulli processes, ([53], Theorem 2.1) that implies that we have that

$$\mathsf{E}_\epsilon \sup_{t\in T} \sum_{i\geq 1} \epsilon_i\theta_i(t_i) \leq \mathsf{E} \sup_{t\in T} \sum_{i\geq 1} \epsilon_i t_i = b(T) \ .$$

Combining with (4.39) finishes the proof. □

Chopping maps were invented to prove the following, that illustrates well their power.

Proposition 4.3.8. *There exists a constant L such that for each subset T of ℓ^2 we have, for $\epsilon > 0$*

$$\epsilon\sqrt{\log N(T, \epsilon B_2 + Lb(T)B_1)} \leq Lb(T)\,,$$

where $N(T, C)$ is the smallest number of translates of C that can cover T.

This would also follow from the Bernoulli conjecture. Indeed, by the Sudakov minoration (Lemma 2.1.2) we have $\epsilon\sqrt{\log N(T_1, \epsilon B_2)} \leq L\gamma_2(T_1)$, and if $T \subset T_1 + T_2$ we have

$$N(T, \epsilon B_2 + Lb(T)B_1) \leq N(T_1, \epsilon B_2)N(T_2, Lb(T)B_1) \leq N(T_1, \epsilon B_2)$$

whenever $T_2 \subset Lb(T)B_1$.

Proof. Considering $c > 0$, successive application of Propositions 4.3.7 and Theorem 4.2.4 yields

$$b(T) \geq b(\Psi_c(T)) \geq \frac{1}{L}\min\Big(\epsilon\sqrt{\log N(\Psi_c(T), \epsilon B_2)}, \frac{\epsilon^2}{c}\Big)\,, \tag{4.40}$$

because if $m \leq N(\Psi_c(T), \epsilon B_2)$, we can find points $(t_\ell)_{\ell \leq m}$ in $\Psi_c(T)$ with $\|t_\ell - t_{\ell'}\| \geq \epsilon/2$ for $\ell \neq \ell'$, and since $\|t\|_\infty \leq c$ for $t \in \Psi_c(T)$. Thus if we choose $c = \epsilon^2/(2Lb(T))$ where L is as in (4.40) we get

$$b(T) \geq \min\Big(\frac{1}{L}\epsilon\sqrt{\log N(\Psi_c(T), \epsilon B_2)}, 2b(T)\Big)\,,$$

so that $Lb(T) \geq \epsilon\sqrt{\log N(\Psi_c(T), \epsilon B_2)}$ and by (4.36) we have

$$N(T, 2\epsilon B_2 + \frac{4\epsilon^2}{c}B_1) \leq N(\Psi_c(T), \epsilon B_2)\ .$$

□

The proof of Theorem 4.3.1 will rely on a specialized version of the partition scheme of Section 1.3. This special scheme will not be used anywhere else in the book. In this scheme we consider a set T provided with a family of distances $(d_k)_{k \geq 0}$. This family is decreasing:

$$\forall k \geq 0\,,\ d_{k+1} \leq d_k\ . \tag{4.41}$$

We denote by $\Delta_k(A)$ the diameter of a set A for the distance d_k. We denote by $B_k(t, a)$ the ball for d_k of center t and radius a.

Consider a family of functionals $(F_k)_{k \geq 1}$ and assume that $F_{k+1} \leq F_k$. Consider $\gamma > 1$.

Definition 4.3.9. *We say that the functionals satisfy the growth condition (with parameter γ) if the following occurs. Given $n \geq 0$, and $1 \leq k \leq j$ such that*

$$k\gamma \geq j \; ; \; r^{-j}2^{n/2} \leq \frac{1}{r} \, , \tag{4.42}$$

then, setting $m = 2^{2^n}$, given any points $(t_\ell)_{\ell \leq m}$ such that

$$\exists t \in T \, , \forall \ell \leq n \, , \, t_\ell \in B_k(t, r^j)$$

$$\forall \ell \neq \ell' \, , \, d_k(t_\ell, t_{\ell'}) \geq r^{-j-1} \tag{4.43}$$

and given sets $H_\ell \subset B_k(t_\ell, r^{-j-2})$ for $\ell \leq m$, we have

$$F_k\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq r^{-j-1}2^{n/2} + \min_{\ell \leq m} F_k(H_\ell) \, . \tag{4.44}$$

The reader observes that in (4.44), the functionals depend on which distance we use rather than on n.

Theorem 4.3.10. *Under the previous conditions, and if moreover $F_1(T) \leq 1/2r^2$, $\Delta_1(T) \leq 1/r$ and $r \geq 4$, we can find an admissible sequence $(\mathcal{A}_n)$ of T and for each $A \subset \mathcal{A}_n$ an integer $j(A) \geq 1$ such that*

$$\forall A \in \mathcal{A}_n \, , \, \Delta_{j(A)}(A) \leq 2r^{-j(A)} \tag{4.45}$$

$$\forall t \in T \, , \sum_{n \geq 0} 2^{n/2} r^{-j(A_n(t))} \leq K(r, \gamma) \tag{4.46}$$

$$A \in \mathcal{A}_n \, , \, B \in \mathcal{A}_{n-1} \, , \, A \subset B \Rightarrow j(B) \leq j(A) \leq j(B) + 1 \, . \tag{4.47}$$

Proof. The proof resembles that of Theorem 1.3.2. The difference is that the construction will be performed for a distance that varies with the stage of the construction. The main difficulty is that the information we have about the behavior of certain sets for a distance d_k says little about their behavior for the distance d_{k+1}. The way around this difficulty is that (roughly speaking) we perform the construction of Theorem 1.3.2 as long as possible with the same distance d_k before we switch to another distance $d_{k'}$, where $k' > k$. When we switch distances, we lose a lot of information, but, fortunately, enough information has been gathered while we were using the distance d_k to make the construction useful.

Together with $C \in \mathcal{A}_n$ we construct a point t_C, integers $1 \leq k(C) \leq j(C)$, and numbers $a_i(C) \geq 0 \, , \, 0 \leq i \leq 2$. (The numbers $a_i(C)$ play the role of the numbers $b_i(C)$ of Theorem 1.3.2. We avoid the notation $b_i(C)$ to prevent confusion with the quantity $b(C)$.) The integer $k(C)$ indicates with which distance we work. Writing $k = k(C)$ and $j = j(C)$ we will have the following.

$$C \subset B_k(t_C, r^{-j}) \tag{4.48}$$

$$F_k(C) \le a_0(C) \le 3 - 2^{-n} \tag{4.49}$$

$$\forall t \in C\,,\ F_k(C \cap B_k(t, r^{-j-1})) \le a_1(C) \tag{4.50}$$

$$\forall t \in C\,,\ F_k(C \cap B_k(t, r^{-j-2})) \le a_2(C) \tag{4.51}$$

$$a_0(C) - 2^{(n-1)/2} r^{-j(C)-1} - 2^{-n} \le a_2(C) \le a_0(C) \tag{4.52}$$

$$a_1(C) \le a_0(C)\ . \tag{4.53}$$

We will also require the following technical conditions:

$$2^{n/2} r^{-j} \le \frac{1}{r} \tag{4.54}$$

$$U^{j-k} 2^{-n} \le a_0(C) - a_1(C)\ , \tag{4.55}$$

where $U = r^c$, for $c = 2\gamma/(\gamma - 1)$. The intuitive meaning of (4.55) is that it is a technical device to ensure that the ratio j/k stays close to one, since we want this ratio to be $< \gamma$ to be able to use (4.44). Finally, we will also ensure the following relation. If $A \in \mathcal{A}_{n+1}$, $A \subset C \in \mathcal{A}_n$, then

$$\begin{aligned} &Ua_0(A) + a_1(A) + a_2(A) + \frac{1}{4} 2^{n/2} r^{-j(A)-1} \\ &\le Ua_0(C) + a_1(C) + a_2(C) + \frac{1}{8} 2^{(n-1)/2} r^{-j(C)-1} + 2^{-n+1} U\ . \end{aligned} \tag{4.56}$$

To start the construction, we set $\mathcal{A}_0 = \{T\}$, $a_1(T) = a_2(T) = F_1(T) \le 1$, $a_0(T) = a_1(T) + 1 \le 2 = 3 - 2^0$ and $k(T) = j(T) = 1$, and we easily check all the required relations. To perform the induction, given $C \in \mathcal{A}_n$, we set $m = 2^{2^n}$, so that $mN_n \le N_{n+1}$; and we proceed to split C in at most m pieces. We set $k = k(C)$, $j = j(C)$ and $\epsilon = \min(2^{-n}\ , 2^{n/2} r^{-j-1})$.

By induction over ℓ , $1 \le \ell \le m$, we construct points $t_\ell \in C$ and sets $A_\ell \subset C$ as follows. First, we choose t_1 such that

$$F_k(C \cap B_k(t_1, r^{-j-2})) \ge \sup_{t \in C} F_k(C \cap B_k(t, r^{-j-2})) - \epsilon\ .$$

We then set $A_1 = C \cap B_k(t_1, r^{-j-1})$.

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

$$F_k(D_\ell \cap B_k(t_{\ell+1}, r^{-j-2})) \ge \sup_{t \in D_\ell} F_k(D_\ell \cap B_k(t, r^{-j-2})) - \epsilon\,, \tag{4.57}$$

we set $A_{\ell+1} = D_\ell \cap B_k(t_{\ell+1}, r^{-j-1})$ and we continue. If the construction does not stop before t_{m-1} is constructed, we define $A_m = D_{m-1} = C \setminus \bigcup_{\ell<m} A_\ell$. In this manner we have partitioned C in at most m pieces. Consider one of these pieces; call it A.

We first consider the case where $A = A_m$. We set $j(A) = j = j(C)$, $k(A) = k = k(C)$, $t_A = t_C$,

$$a_0(A) = a_0(C)\,,\ a_1(A) = a_1(C) \ \text{ and } \ a_2(A) = a_0(A) - 2^{n/2} r^{-j-1} + \epsilon\ . \quad (4.58)$$

It is obvious that (4.48) to (4.50) and (4.53), (4.55) hold for A. Since $\epsilon \le 2^{n/2} r^{-j-1}$ it is also obvious that (4.52) holds for A. From (4.55) and since $a_0(C) \le 3$ we have

$$U^{j-k} 2^{-n} \le 3$$

so that

$$\left(\frac{U}{r^2}\right)^j \le 3 \cdot U^k 2^n r^{-2j} \le U^k$$

using (4.54) and since $r \ge 4$. Since $U = r^c$, this yields

$$r^{j(c-2)} \le r^{ck}$$

so that $k \ge j(c-2)/c$, i.e. $\gamma k \ge j$. Thus using (4.54) we see that (4.42) holds, and hence, setting $D_0 = C$, and choosing an arbitrary point $t = t_m$ in A we can use (4.44) for the sets $H_\ell = D_{\ell-1} \cap B_k(t_\ell, r^{-j-2})$. We first conclude that

$$r^{-j-1} 2^{n/2} \le F_k(C) \le F_1(T) \le \frac{1}{2r^2}\ ,$$

so that $r^{-j} 2^{(n+1)/2} \le 1/r$, and (4.42) holds for A. Next, as in the proof of Theorem 1.3.2, we deduce from (4.44) and (4.57) that

$$\forall t \in A\,,\ F_k(A \cap B_k(t, r^{-j-2})) \le F_k(C) - 2^{n/2} r^{-j-1} + \epsilon$$

and since $F_k(C) \le a_0(C)$, the definition of $a_2(A)$ shows that (4.51) follows for A.

To prove (4.56), we set $w = 2^{n/2} r^{-j-1}$. By definition we have $a_2(A) = a_0(C) - w + \epsilon$, and since $-3/4 < -1/\sqrt{2}$, since $\epsilon \le 2^{-n} \le 2^{-n} U$, we have

$$\begin{aligned} U a_0(A) + a_1(A) + a_2(A) + \frac{w}{4} &\le U a_0(C) + a_1(C) + a_0(C) - \frac{3}{4} w + \epsilon \\ &\le U a_0(C) + a_1(C) + a_0(C) - \frac{w}{\sqrt{2}} + 2^{-n} U \\ &\le U a_0(C) + a_1(C) + a_2(C) + 2^{-n+1} U\ . \end{aligned}$$

using the left-hand side of (4.52) in the last inequality. This finishes the construction when $A = A_m$.

We now examine the situation where $A = A_\ell$, $\ell < m$. In this situation we will have $j(A) = j(C) + 1$. In order to maintain the ratio j/k close to 1 as is

required by (4.55) we will want to increase k often. There are two different cases.
Case 1. We have

$$(U+2)\,a_1(C) \le U a_0(C) + a_1(C) + a_2(C)\,. \tag{4.59}$$

The idea here is that this relation will suffice to prove (4.56), so that we can afford to increase k as much as possible, change distances and lose much of the information contained in (4.50) and (4.51). We set $k(A) = j(A) = j+1$, $t_A = t_\ell$ and

$$\begin{aligned} a_1(A) &= a_2(A) = a_1(C) \\ a_0(A) &= a_1(C) + 2^{-n-1}\,. \end{aligned}$$

The purpose of this latter definition is to obtain (4.55) for A. By (4.53) and (4.49) we have $a_0(A) \le a_0(C) + 2^{-n-1} \le 3 - 2^{-n-1}$. Moreover, since $k(A) = j+1 \ge k$, using (4.50) in the last inequality we have

$$F_{k(A)}(A) = F_{j+1}(A_\ell) \le F_k(A_\ell) \le F_k(C \cap B_k(t_\ell, r^{-j-1})) \le a_1(C)\,.$$

This proves (4.49) for A. It follows from (4.59) that

$$U a_0(A) + a_1(A) + a_2(A) \le U a_0(C) + a_1(C) + a_2(C) + 2^{-n-1}U\,,$$

and since we have $\sqrt{2}r^{-1} \le 1/2$ because $r \ge 4$ this implies (4.56), since $j(A) = j(C) = 1$. All conditions (4.49) to (4.56) should now be obvious.
Case 2. Relation (4.59) fails, i.e. we have

$$(U+2)a_1(C) > U a_0(C) + a_1(C) + a_2(C)\,,$$

which we rewrite as

$$a_1(C) - a_2(C) \ge U(a_0(C) - a_1(C))\,. \tag{4.60}$$

We then set $k(A) = k = k(C)$, $j(A) = j+1$, $t_A = t_\ell$,

$$a_0(A) = a_2(A) = a_1(C)\,,\ a_1(A) = \min(a_1(C), a_2(C))\,,$$

so that $a_0(A) - a_1(A) \ge a_1(C) - a_2(C)$ and (4.60) and (4.55) for C imply that (4.55) holds for A. To prove (4.56) we simply observe that $a_0(A) \le a_1(C) \le a_0(C)$ and that $a_1(A) + a_2(A) \le a_1(C) + a_2(C)$ and we recall that $\sqrt{2}r^{-1} \le 1/2$. The other relations should be obvious.

Summation of the relations (4.56) gives

$$\sum_{n\ge1} 2^{n/2} r^{-j(A_n(t))} \le K(r,\gamma)$$

and this yields (4.46), since $j(T) = 1$. As for condition (4.47), it holds by construction. □

Proof of Theorem 4.3.1. Without loss of generality we can assume that $0 \in T$, We recall the constant L_0 of Corollary 4.2.3. We consider $2 > p > 1$ and we choose γ such that $2\gamma = 1 + 1/(2-p)$ so that $1 < \gamma$ and $(2-p)\gamma < 1$. We then fix $r \geq \max(4, L_0)$ such that $q = r^{2\gamma}$ is an integer. For $k \in \mathbb{Z}$ we set $c(k) = r^{-2\gamma k} = q^{-k}$, and we consider the distance d_k on ℓ^2 given by

$$d_k(s,t) = \|\Psi_{c(k)}(s) - \Psi_{c(k)}(t)\|_2 \ .$$

It follows from (4.26) and (4.28) that $d_{k+1} \leq d_k$. For a subset A of ℓ^2, we define

$$F'_k(A) = b(\Psi_{c(k)}(A)) \ .$$

Thus by (4.26) and Proposition 4.3.7 we have $F'_{k+1} \leq F'_k$. Consider points $(t_\ell)_{\ell \leq m}$ as in (4.43), and sets $(H_\ell)_{\ell \leq m}$, with $H_\ell \subset B_k(t_\ell, r^{-j-2})$. The definition of d_k shows that we have $\Psi_{c(k)}(H_\ell) \subset B(u_\ell, r^{-j-2})$ where $u_\ell = \Psi_{c(k)}(t_\ell)$ and where the ball is for the ℓ^2 distance. Moreover $u_\ell \in D = \{u \ , \ \|u\|_\infty \leq c(k) = q^{-k}\}$, and $\Delta(D, d_\infty) \leq 2q^{-k}$. Thus, if

$$2q^{-k} \leq 4r^{-j-1}/\sqrt{\log m} \ , \tag{4.61}$$

we can apply (4.11) with $a = r^{-j-1}$ to the sets $\Psi_{c(k)}(H_\ell)$ rather than H_ℓ and the points u_ℓ rather than t_ℓ. Since $\log 2^{2^n} \leq 2^n$, we see in particular that (4.61) holds for $m = 2^{2^n}$ whenever $q^{-k} \leq r^{-j-1}2^{-n/2}$, and since $2^{n-1} \leq \log 2^{2^n}$ we then have

$$F'_k\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq \frac{1}{2L_0} r^{-j-1} 2^{n/2} + \min_{\ell \leq m} F'_k(H_\ell) \ .$$

The condition

$$q^{-k} = r^{-2\gamma k} \leq r^{-j-1}2^{-n/2}$$

is equivalent to

$$2^{n/2} r^{-j+1} \leq r^{2(\gamma k - j)}$$

so that it holds whenever $\gamma k \geq j$ and $2^{n/2} r^{-j} \leq 1/r$. Thus we see that the functionals $F_k(A) = 2L_0 F'_k(A)$ satisfy the growth condition of Definition 4.3.9. Next we observe that there is a constant $K(r)$ such that

$$b(T) \leq \frac{1}{K(r)} \Rightarrow F_1(T) \leq \frac{1}{2r^2} \ , \ \Delta(T, d_2) < \frac{1}{4r^\gamma} \ , \tag{4.62}$$

where now d_2 denote the distance induced the norm of ℓ^2. This follows from the fact that $F_1(T) \leq 2L_0 b(T)$ and from (4.14). Thus, if $b(T) \leq 1/K(r)$ all the hypothesis of Theorem 4.3.10 are satisfied, and we can find an admissible sequence $(\mathcal{A}_n)$ of T and integers $j(A)$ that satisfy (4.45) to (4.47).

From (4.33), used for $c = c(k) = r^{-2\gamma k}$ we see that if $s = (s_i)_{i \geq 1}$ and $t = (t_i)_{i \geq 1}$ we have

$$\sum_{i\geq 1}(s_i - t_i)^2 \wedge c^2 \leq 4d_k^2(s,t) ,$$

and if we set $V = r^{2\gamma}$, since $c^2 = V^{-2k}$ we get

$$\sum_{i\geq 1}(s_i - t_i)^2 \wedge V^{-2k} \leq 4d_k^2(s,t) .$$

Since $k(A) \leq j(A)$ we then see that the hypothesis of Theorem 2.6.3 are satisfied when μ is the counting measure on $\mathbb{N}$ and if $\delta(A) = 2r^{-j(A)}$. Since $0 \in T$ and $\Delta(T, d_2) \leq 1/(4r^\gamma)$, we have $\|t\|_\infty < 1/(2r^\gamma) \leq 1/(2V)$ for $t \in T$, and since $j(T) = 1$ we have $T_3 = \{0\}$ by (2.105), and thus $T \subset T_1 + T_2$. By (2.102) and (4.46) we have $\gamma_2(T_1, d_2) \leq LK(r,p)$. Also,

$$\begin{aligned} V^{2j(A_{n+1}(t))-pj(A_n(t))}\delta^2(A_{n+1}(t)) &\leq Lr^{(4\gamma-2)j(A_{n+1}(t))-2p\gamma j(A_n(t))} \\ &\leq K(r,\gamma)r^{2((2-p)\gamma-1)j(A_n(t))} \end{aligned}$$

since $j(A_{n+1}(t)) \leq j(A_n(t)) + 1$ by (4.47). Since $(2-p)\gamma - 1 < 0$, it follows from (2.104) that $\|t\|_p \leq K(r,p)$ for $t \in T_2$.

The conclusion of Theorem 4.3.1 follows by homogeneity. □

4.4 Further Thoughts

It should be stressed that the results presented in Sections 4.2 and 4.3 do not even come close to addressing the real difficulty of the Bernoulli conjecture. The method of chopping maps, by which we obtained lower bounds, is insufficient. To see this, consider an integer $k \geq 1$, and consider the class $\mathcal{C}$ of elements $t = (t_i)$ that have the following property. For each i, t_i takes one of the values $0, 2^{-1}, \cdots, 2^{-k}$, and for each $1 \leq \ell \leq k$, there are exactly 2^ℓ indices i for which $t_i = 2^{-\ell}$. Consider a set T that consists of $2^{2^{k+1}}$ elements of $\mathcal{C}$ that have disjoint support. Since for $t \in \mathcal{C}$ we have $\mathsf{P}(\sum \epsilon_i t_i = k) \geq 2^{-2^{k+1}}$, it is simple to see that $b(T) \geq k/L$. Yet it does seem possible to use chopping maps to prove this, or even to prove more than $b(T) \geq 1/L$.

The previous example shows that it could be of interest to study the special case of the Bernoulli conjecture where there is a partition $(I_k)_{k\geq 1}$ of the index set such that for each $t \in T$, we have $t_i \in \{0, 2^{-k}\}$ for $i \in I_k$.

Many of the results of this book were first discovered by following the idea that T is "large" if and only if it contains a "large tree". (What this means precisely is explained in the Appendix.) Trees do not seem to suffice in the case of the Bernoulli conjecture, as the previous example shows.

5 Families of distances

5.1 A General Partition Scheme

Not all processes of interest satisfy a condition as simple as (0.4) or even (1.21). In certain natural situations, the increments of a process cannot be controlled using only one or two distances, but can be controlled using a family of distances. Quite interestingly, once the first surprise is passed and the right setting has been found, it turns out that this is not more difficult than working with a single distance.

The goal of the present section is to generalize to this setting the partitioning scheme of Section 1.3. In Section 5.2 we will apply this tool to the study of "canonical processes" and in Section 5.3 to infinitely divisible processes. These two sections are independent of each other.

We consider a family of maps $(\varphi_j)_{j\in\mathbb{Z}}$, with the following properties:

$$\varphi_j : T \times T \to \mathbb{R}^+ \cup \{\infty\}\,,\ \varphi_j \geq 0\,,\ \varphi_j(s,t) = \varphi_j(t,s)\,,\ \varphi_{j+1} \geq \varphi_j\ .$$

These maps play the role of a family of distances (although it probably would be better to think of φ_j as the square of a distance rather than as of a distance).

We consider functionals $F_{n,j}$ on T for $n \geq 0\,,\ j \in \mathbb{Z}$. We assume

$$F_{n+1,j} \leq F_{n,j}\ ;\ F_{n,j+1} \leq F_{n,j}\ . \tag{5.1}$$

We define

$$B_j(t,c) = \{s \in T\ ;\ \varphi_j(s,t) \leq c\}\ ,$$

so that $B_{j+1}(t,c) \subset B_j(t,c)$.

We will assume that the functionals satisfy a "growth condition", that is very similar in spirit to Definition 1.3.1. This condition involves as main parameter an integer $\kappa \geq 6$. We set $r = 2^{\kappa-4}$. The role of r is as in (1.30), the bigger r, the weaker the growth condition. The reason why we take r of the type $r = 2^{\kappa-4}$ for an integer κ is purely technical convenience.

The growth condition, that also involves as secondary parameter an integer $n_0 \geq 1$, is as follows.

Definition 5.1.1. *We say that the functionals* $F_{n,j}$ *satisfy the* growth condition *(for* n_0 *and* r*) if the following occurs. Consider any* $j \in \mathbb{Z}$*, any* $n \geq n_0$ *and* $m = N_n$*. Consider any points* $t\,, t_1\,, \cdots\,, t_m$ *in* T *and assume*

$$\forall \ell \leq m\,,\ t_\ell \in B_{j-1}(t, 2^{n-1}) \tag{5.2}$$

$$\forall \ell\,,\ \ell' \leq m\,,\ \ell \neq \ell'\,,\ \varphi_j(t_\ell, t_{\ell'}) \geq 2^n\ . \tag{5.3}$$

Consider any sets $H_\ell \subset B_{j+1}(t_\ell, 2^{n+\kappa})$*. Then*

$$F_{n,j}\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq 2^n r^{-j} + \min_{\ell \leq m} F_{n+1,j+1}(H_\ell)\ . \tag{5.4}$$

Besides the rather weak requirement that $\varphi_{j+1} \geq \varphi_j$, we have not made assumptions on how φ_j relates to φ_{j+1}; but we have little chance to prove (5.4) unless $B_{j+1}(t_\ell, 2^{n+\kappa})$ is quite smaller than $B_j(t_\ell, 2^n)$.

To understand the preceding conditions we will carry out the case where

$$\varphi_j(s,t) = r^{2j} d^2(s,t) \tag{5.5}$$

for a distance d on T. The reader is encouraged to carry out the more general case where $\varphi_j(s,t) = r^{\alpha j} d^\beta(s,t)$ for $\alpha, \beta > 0$. Denoting by $B(t,b)$ the ball for d of center t and radius b, we thus have

$$B_j(t,c) = B(t, r^{-j}\sqrt{c})\,.$$

Thus in (5.3) we require that

$$\forall \ell\,,\ \ell' \leq m\,,\ \ell \neq \ell'\,,\ d(t_\ell, t_{\ell'}) \geq 2^{n/2} r^{-j} := a\ . \tag{5.6}$$

On the other hand, the condition $H_\ell \subset B_{j+1}(t_\ell, 2^{n+\kappa})$ means that

$$H_\ell \subset B(t_\ell, 2^{(n+\kappa)/2} r^{-j-1}) = B(t_\ell, \eta a)\,,$$

for $\eta = 2^{\kappa/2}/r = 2^{-\kappa/2+4} = 4/\sqrt{r}$. Thus, as r gets larger, η gets smaller, and the sets H_ℓ become better separated. Also, (5.4) reads as

$$F_{n,j}\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq 2^{n/2} a + \min_{\ell \leq m} F_{n+1,j+1}(H_\ell)\ ,$$

which strongly resembles (1.31) for $\theta(n) = 2^{n/2}$ and $\beta = 1$.

Theorem 5.1.2. *Assume that the functionals* $F_{n,j}$ *are as above, and in particular satisfy the growth condition of Definition 5.1.1, and that, for some* $j_0 \in \mathbb{Z}$ *we have*

$$\forall s\,,\ t \in T\ ,\ \varphi_{j_0-1}(s,t) \leq 2^{n_0-1}\ . \tag{5.7}$$

Then there exists an admissible sequence $(\mathcal{A}_n)$ *and for each* $A \in \mathcal{A}_n$ *an integer* $j(A) \in \mathbb{Z}$ *such that*

$$A \in \mathcal{A}_n \, , \, B \in \mathcal{A}_{n-1} \, , \, A \subset B \Rightarrow j(B) \le j(A) \le j(B) + 1 \tag{5.8}$$

$$\forall t \in T \ , \ \ \sum_{n \ge n_0} 2^n r^{-j(A_n(t))} \le L(F_{n_0,j_0}(T) + 2^{n_0} r^{-j_0}) \tag{5.9}$$

$$\forall n \ge n_0 \, , \forall A \in \mathcal{A}_n \, , \, \exists t_A \in T \, , \, A \subset B_{j(A)-1}(t_A, 2^{n-1}) \, . \tag{5.10}$$

To make sense out of this, we again carry out the case (5.5). Then (5.7) means that $\Delta(T,d) \le r^{-j_0+1} 2^{(n_0-1)/2}$, while (5.10) implies that $\Delta(A,d) \le r^{-j(A)+1} 2^{n/2+1}$, and (5.9) implies that

$$\forall t \in T \, , \sum_{n \ge n_0} 2^{n/2} \Delta(A_n(t), d) \le Lr(F_{n_0,j_0}(T) + 2^{n_0} r^{-j_0}) \, .$$

Taking for j_0 the largest integer such that $\Delta(T,d) \le r^{-j_0+1} 2^{(n_0-1)/2}$, we get

$$\forall t \in T \, , \sum_{n \ge n_0} 2^{n/2} \Delta(A_n(t), d) \le Lr\big(F_{n_0,j_0}(T) + 2^{n_0/2} \Delta(T,d)\big).$$

This relation resembles the relation one gets by combing (1.33) with Lemma 1.3.3, and the parameter n_0 plays a role similar to τ.

The proof of Theorem 5.1.2 follows closely the proof of Theorem 1.3.2, and of course the reader should master this latter result before attempting to read it.

Proof of Theorem 5.1.2. Together with $C \in \mathcal{A}_n$, $n \ge n_0$ we construct integers $j(C) \in \mathbb{Z}$, $q(C) \in \mathbb{N}$, points $t_C \in T$, and numbers $b_0(C)$, $b_1(C)$, $b_2(C) \ge 0$ satisfying the following conditions, where $j = j(C)$

$$C \subset B_{j-1}(t_C \, , \, 2^{n-1}) \tag{5.11}$$

$$F_{n,j}(C) \le b_0(C) \tag{5.12}$$

$$\forall t \in C \, , \, F_{n,j}(C \cap B_j(t, 2^{n+\kappa-q(C)})) \le b_1(C) \tag{5.13}$$

$$\forall t \in C, F_{n,j+1}(C \cap B_{j+1}(t, 2^{n+\kappa-1})) \le b_2(C) \tag{5.14}$$

$$b_0(C) \ge b_1(C) \ge b_0(C) - 2^{n+\kappa-q(C)-4} r^{-j} \tag{5.15}$$

$$b_0(C) \ge b_2(C) \ge b_0(C) - 2^{n-1} r^{-j} \, . \tag{5.16}$$

Moreover, we will arrange that if $A \subset C$, $A \in \mathcal{A}_{n+1}$ and $n \ge n_0$ then

$$\begin{aligned} &b_0(A) + b_1(A) + b_2(A) + \frac{1}{4} 2^n r^{-j(A)} \\ &\qquad \le b_0(C) + b_1(C) + b_2(C) + \frac{3}{16} 2^{n-1} r^{-j(C)} \, . \end{aligned} \tag{5.17}$$

To start the construction we pick an arbitrary point $t_T \in T$. We define $\mathcal{A}_{n_0} = \{T\}$, $j(T) = j_0$, $q(T) = 0$, $b_0(T) = b_1(T) = b_2(T) = F_{n_0,j_0}(T)$. Thus (5.11) holds by (5.7), while (5.12) to (5.16) are obvious.

To construct $\mathcal{A}_{n+1}$ once $\mathcal{A}_n$ has been constructed, we will show how to split an element C of $\mathcal{A}_n$ in at most $m = N_n$ pieces. (Thus, since $N_n^2 \le N_{n+1}$, $\mathcal{A}_{n+1}$ contains at most N_{n+1} sets.) We set $j = j(C)$ and $q = q(C)$. We consider $\epsilon > 0$ to be determined later. We set $D_0 = C$. First we choose t_1 in D_0 with

$$F_{n+1,j+1}(D_0 \cap B_{j+1}(t_1, 2^{n+\kappa})) \ge \sup_{t \in D_0} F_{n+1,j+1}(D_0 \cap B_{j+1}(t, 2^{n+\kappa})) - \epsilon \ .$$

We then set $A_1 = D_0 \cap B_j(t_1, 2^n)$ and $D_1 = D_0 \setminus A_1$. If D_1 is not empty, we choose t_2 in D_1 such that

$$F_{n+1,j+1}(D_1 \cap \ B_{j+1}(t_2, 2^{n+\kappa})) \ge \sup_{t \in D_1} F_{n+1,j+1}(D_1 \cap B_{j+1}(t, 2^{n+\kappa})) - \epsilon \ ,$$

and we set $A_2 = D_1 \cap B_j(t_2, 2^n)$ and $D_2 = D_1 \setminus A_2$. We continue in this manner until either we exhaust C or we construct D_{m-1}. In the latter case we set $A_m = D_{m-1}$ and we stop the construction.

In this manner we split C in at most $m = N_n$ pieces $A_1 \, , \cdots , \, A_m$. Consider one of these, which we call A.

We first examine the case where

$$A = A_m = D_{m-1} = C \setminus \bigcup_{\ell < m} B_j(t_\ell, 2^n) \ .$$

In that case we set $j(A) = j = j(C)$, $q(A) = q(C) + 1$, $t_A = t_C$ and

$$b_0(A) = b_0(C) \, , \, b_1(A) = b_1(C) \, , \, b_2(A) = b_0(C) - 2^n r^{-j} + \epsilon \, . \tag{5.18}$$

It is obvious that (5.11) holds for A since it holds for C. Since $F_{n+1,j} \le F_{n,j}$ and since $n+1-q(A) = n-q(C)$, the fact that (5.12) and (5.13) hold for C imply that they hold for A. It is obvious that (5.15) holds for A, and (5.16) holds for A as soon as $\epsilon \le 2^n r^{-j}$.

We turn to the proof of (5.14). Consider $1 \le \ell < m$. By construction of t_ℓ we have, since $A = D_{m-1} \subset D_{\ell-1}$,

$$\begin{aligned} \forall t \in D_{\ell-1}, \ \ & F_{n+1,j+1}(A \cap B_{j+1}(t, 2^{n+\kappa})) \\ \le \ & F_{n+1,j+1}(D_{\ell-1} \cap B_{j+1}(t, 2^{n+\kappa})) \\ \le \ & F_{n+1,j+1}(D_{\ell-1} \cap B_{j+1}(t_\ell, 2^{n+\kappa})) + \epsilon \ . \end{aligned} \tag{5.19}$$

Consider $t \in A$ and set $H_\ell = D_{\ell-1} \cap B_{j+1}(t_\ell, 2^{n+\kappa})$ for $1 \le \ell < m$ and $H_m = A \cap B_{j+1}(t, 2^{n+\kappa})$. By (5.19), for $\ell < m$ we have $F_{n+1,j+1}(H_m) \le F_{n+1,j+1}(H_\ell) + \epsilon$ and thus

$$\inf_{\ell \le m} F_{n+1,j+1}(H_\ell) \ge F_{n+1,j+1}(H_m) - \epsilon \ .$$

Define $t_m = t$. We have $\varphi_j(t_\ell, t_{\ell'}) \ge 2^n$ for $\ell \ne \ell'$, and $t_\ell \in C \subset B_{j-1}(t_C, 2^{n-1})$, so we have by (5.12) and (5.4) that

$$
\begin{aligned}
b_0(C) \ge F_{n,j}(C) \ge F_{n,j}\Big(\bigcup_{\ell\le m} H_\ell\Big) &\ge 2^n r^{-j} + F_{n+1,j+1}(H_m) - \epsilon \\
&= 2^n r^{-j} + F_{n+1,j+1}(A \cap B_{j+1}(t, 2^{n+\kappa}) - \epsilon
\end{aligned}
$$

and this proves that (5.14) holds for A.

A good choice of ϵ is now obvious. Taking $\epsilon = 2^{n-2}r^{-j}$, and since by (5.16) we have $b_2(C) \ge b_0(C) - 2^{n-1}r^{-j}$, we get by (5.18) that

$$
\begin{aligned}
b_0(A) + b_1(A) + b_2(A) + \frac{1}{4}2^n r^{-j} &\le 2b_0(C) + b_1(C) - 2^{n-1}r^{-j} \\
&\le b_0(C) + b_1(C) + b_2(C)\,,
\end{aligned}
$$

from which (5.17) follows.

We now consider the case where $A = D_{\ell-1} \cap B_j(t_\ell, 2^n)$, $1 \le \ell < m$. We define $j(A) = j+1$, $q(A) = 2$, $t_A = t_\ell$. It is obvious that (5.11) holds for A. By (5.14) for C, and since $F_{n+1,j+1} \le F_{n,j+1}$, we have

$$
\forall t \in A\,,\ F_{n+1,j(A)}(A \cap B_{j(A)}(t, 2^{n+\kappa-1})) \le b_2(C)\,. \tag{5.20}
$$

We note that, using (5.16) we have

$$
b_2(C) \ge b_0(C) - 2^{n-1}r^{-j} = b_0(C) - 2^{n+1+\kappa-q(A)-4}r^{-j(A)}\,, \tag{5.21}
$$

because $r = 2^{\kappa-4}$.

For the rest of the argument we need to distinguish cases. Let us first assume that $q(C) > \kappa$. We then take

$$
b_1(A) = b_2(C)\ ,\ b_0(A) = b_2(A) = b_0(C)\,. \tag{5.22}
$$

Since $F_{n+1,j+2} \le F_{n+1,j+1} \le F_{n,j+1} \le F_{n,j}$, it is obvious that (5.12), (5.14) and (5.16) hold for A. Moreover, since $n+1-q(A) = n-1$, (5.13) and (5.15) for A follow from (5.20) and (5.21) respectively. Using again (5.22) we have

$$
\begin{aligned}
b_0(A) + b_1(A) + b_2(A) &= 2b_0(C) + b_2(C) \\
&\le b_0(C) + b_1(C) + b_2(C) + 2^{n+\kappa-q(C)-4}r^{-j}
\end{aligned}
$$

using (5.15). Thus, since $q(C) \ge \kappa + 1$ we have

$$
b_0(A) + b_1(A) + b_2(A) \le b_0(C) + b_1(C) + b_2(C) + \frac{1}{16}2^{n-1}r^{-j}\,. \tag{5.23}
$$

Since $\kappa \ge 6$ we have $r = 2^{\kappa-4} \ge 4$, so that $2^n r^{-j(A)} \le 2^{n-1}r^{-j}/2$ and (5.23) implies (5.17).

Suppose now that $q(C) \le \kappa$. In that case we take

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

From (5.13) we see that, since $F_{n+1,j+1} \le F_{n,j}$

$$F_{n+1,j(A)}(A) \le F_{n+1,j(A)}(C \cap B_j(t_\ell, 2^n)) \le b_1(C) = b_0(A) = b_2(A) \,. \quad (5.24)$$

This implies that (5.12) and (since $F_{n+1,j+2} \le F_{n+1,j+1}$) that (5.14) hold for A. Moreover, it follows from (5.24) and (5.20) that (5.13) holds for A. Relation (5.16) for A is obvious, and since $b_1(A) \le b_0(A)$ by construction, relation (5.15) for A follows from (5.21) since $b_0(C) \ge b_1(C)$. Finally, we have

$$b_0(A) + b_1(A) + b_2(A) \le 2b_1(C) + b_2(C) \le b_0(C) + b_1(C) + b_2(C)$$

and since $2^n r^{-j(A)} \le 2^{n-1} r^{-j}/2$, (5.17) holds. This completes the construction.

Summation of the relations (5.17) for $n \ge n_0$ yields

$$\forall t \in T \,, \sum_{n \ge n_0+1} \frac{1}{16} 2^{n-1} r^{-j(A_n(t))} \le 3F_{n_0,j_0}(T) + \frac{3}{16} 2^{n_0-1} r^{-j_0} \,.$$

This concludes the proof, using that $j(A_0(t)) = j(T) = j_0$ to control the term $n = n_0$ in the summation of (5.9). □

5.2 The Structure of Certain Canonical Processes

In this section we prove a far reaching generalization of Theorem 2.1.1. We consider independent, centered, symmetric r.v. $(Y_i)_{i\ge1}$. We assume that

$$U_i(x) = -\log P(|Y_i| \ge x) \quad (5.25)$$

is convex. Since it is a matter of normalization, we assume that $U_i(1) = 1$. Since $U_i(0) = 0$ we then have $U_i'(1) \ge 1$.

Given $t = (t_i)_{i\ge1} \in \ell^2$, we define

$$X_t = \sum_{i\ge1} t_i Y_i \,.$$

The condition $t \in \ell^2$ is to ensure the convergence of the series. (Very little of the results we will present is lost if one assumes that only finitely many of the coefficients t_i are not 0). The aim of this section is to study collections of such r.v. as t varies over a set T. It is in truth a rather amazing fact that this can be done at all at the level of generality that we will achieve. The price to pay for this is the same as ever: we will have to go through a few abstract definitions. These represent the outcome of many steps of abstraction, and the ideas behind them can be understood only gradually.

We consider the function $\hat{U}_i(x)$ given by

$$\hat{U}_i(x) = x^2 \text{ if } 0 < |x| \le 1$$
$$\hat{U}_i(x) = 2U_i(|x|) - 1 \text{ if } |x| \ge 1 \,,$$

so that this function is convex.

Given $u > 0$, we define

$$\mathcal{N}_u(t) = \sup\Big\{\sum_{i\ge1} t_i a_i \ ; \ \sum_{i\ge1} \hat{U}_i(a_i) \le u\Big\} \,.$$

We define

$$B(u) = \{t \ ; \ \mathcal{N}_u(t) \le u\}$$

and, given a number r, we define

$$\varphi_j(s,t) = \inf\{u > 0 \ ; \ s - t \in r^{-j} B(u)\} \,.$$

The only way to get a feeling of what happens is to carry out the meaning of these definitions in a concrete case. The simplest case is when, for all i, we have $U_i(x) = x^2$. In that case, it is rather immediate that

$$x^2 \le \hat{U}_i(x) \le 2x^2 \ ; \ \sqrt{\frac{u}{2}}\|t\|_2 \le \mathcal{N}_u(t) \le \sqrt{u}\|t\|_2,$$

so that $B_2(0,\sqrt{u}) \subset B(u) \subset B_2(0,\sqrt{2u})$, where B_2 denotes the ball of ℓ^2, and

$$\frac{1}{2} r^{2j}\|s-t\|_2^2 \le \varphi_j(s,t) \le r^{2j}\|s-t\|_2^2 \,, \tag{5.26}$$

and we are almost in the situation of (5.5).

The second simplest example is the case where for all i we have $U_i(x) = x$ for $x \ge 0$. In that case we have $|x| \le \hat{U}_i(x) = 2|x| - 1 \le x^2$ for $|x| \ge 1$. Thus $\hat{U}_i(x) \le x^2$ and $U_i(x) \le 2|x|$ for all $x \ge 0$, and hence

$$\sum_{i\ge1} a_i^2 \le u \Rightarrow \sum_{i\ge1} \hat{U}_i(a_i) \le u$$

and

$$\sum_{i\ge1} 2|a_i| \le u \Rightarrow \sum_{i\ge1} \hat{U}_i(a_i) \le u \,.$$

Moreover, if $\sum_{i\ge1} \hat{U}_i(a_i) \le u$, writing $b_i = a_i \mathbf{1}_{\{|a_i|\ge1\}}$ and $c_i = a_i \mathbf{1}_{\{|a_i|<1\}}$ we have $\sum_{i\ge1} |b_i| \le u$ and $\sum_{i\ge1} c_i^2 \le u$. It follows that

$$\frac{1}{L}(u\|t\|_\infty + \sqrt{u}\|t\|_2) \le \mathcal{N}_u(t) \le L(u\|t\|_\infty + \sqrt{u}\|t\|_2) \,,$$

and thus

$$\frac{1}{L}\{t\,;\ \|t\|_\infty \le 1\,,\ \|t\|_2 \le \sqrt{u}\} \subset B(u) \subset L\{t\,;\ \|t\|_\infty \le 1\,,\ \|t\|_2 \le \sqrt{u}\} \,. \tag{5.27}$$

In this case the functions φ_j have a genuinely more complicated structure than in the case of (5.26).

The third simplest example is the case where for some $p \geq 1$ and for all i we have $U_i(x) = x^p$ for $x \geq 0$, and the reader who truly wants to understand what really is going on would do well to work out a version of the general result in this special case. (The cases $p > 2$ and $p < 2$ offer significant differences.) This case was treated by the author in [54] and we owe the present more general setting to a further effort by R. Latala [16].

Theorem 5.2.1. *Assume that there exists an admissible sequence $(\mathcal{A}_n)$ of $T \subset \ell^2$, and for $A \in \mathcal{A}_n$ an integer $j(A) \in \mathbb{Z}$ such that*

$$\forall A \in \mathcal{A}_n \,, \forall s \,,\ s' \in A \,,\ \varphi_{j(A)-1}(s, s') \leq 2^n \,. \tag{5.28}$$

Then

$$\mathsf{E} \sup_{t \in T} X_t \leq Lr \sup_{t \in T} \sum_{n \geq 0} 2^n r^{-j(A_n(t))} \,. \tag{5.29}$$

Let us first interpret this statement in the case where for each i we have $U_i(x) = x^2$. In that case (and more generally when $U_i(x) \geq x^2/L$ for $x \geq 1$) we have $\varphi_j(s,t) \leq Lr^{2j}\|s-t\|_2^2$, so that (5.28) holds as soon as $r^{2j(A)-2}\Delta(A, d_2)^2 \leq 2^n/L$, where of course d_2 denotes the distance induced by the norm of ℓ^2. Taking for $j(A)$ the largest integer that satisfies this inequality, we see that the right-hand side of (5.29) is bounded by $Lr \sup_{t\in T} \sum_{n\geq 0} 2^{n/2}\Delta(A_n(t), d_2)$, so by taking the infimum over the admissible sequences $(\mathcal{A}_n)$ we see that (5.29) implies

$$\mathsf{E} \sup_{t \in T} X_t \leq Lr\gamma_2(T, d_2) \,.$$

Let us now interpret Theorem 5.2.1 when $U_i(x) = x$ for each i. Using (5.27) we see that if $\|s-t\|_\infty \leq r^{-j}/L$ we have $\varphi_j(s,t) \leq Lr^{2j}\|s-t\|_2^2$, so that (5.28) holds whenever $r^{j(A)-1}\Delta(A, d_\infty) \leq 1/L$ and $r^{2j(A)-2}\Delta(A, d_2)^2 \leq 2^n/L$, where of course d_∞ denotes the distance induced by the norm of ℓ^∞. Taking for $j(A)$ the largest integer that satisfies both conditions we see that

$$r^{-j(A)} \leq L\big(\Delta(A, d_\infty) + 2^{-n/2}\Delta(A, d_2)\big) \,,$$

so that (5.29) implies that

$$\mathsf{E} \sup_{t \in T} X_t \leq Lr \sup_{t \in T} \sum_{n \geq 0} \big(2^n \Delta(A_n(t), d_\infty) + 2^{n/2}\Delta(A_n(t), d_2)\big) \,.$$

Using the argument at the beginning of the proof of Theorem 1.2.7, we then see that

$$\mathsf{E} \sup_{t \in T} X_t \leq Lr\big(\gamma_2(T, d_2) + \gamma_1(T, d_1)\big) \,. \tag{5.30}$$

This resembles Theorem 1.2.7, and could actually be deduced from this theorem and an appropriate version of Bernstein's inequality.

It will be a simple adaptation of the proof of Theorem 1.2.6 to deduce Theorem 5.2.1 from the following.

Proposition 5.2.2. *If $u > 0$, $v \geq 1$, we have*

$$\mathsf{P}(X_t \geq Lv\mathcal{N}_u(t)) \leq \exp(-uv) \,. \tag{5.31}$$

Proof of Theorem 5.2.1. We consider an arbitrary element t_0 of T and we set $T_0 = \{t_0\}$. For $n \geq 1$ we consider a set T_n such that

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

For $t \in T$ we define $\pi_n(t)$ by $\{\pi_n(t)\} = A_n(t) \cap T_n$. For any integer k and any t in T_k we have

$$X_t - X_{t_0} = \sum_{1 \leq n \leq k} X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \,. \tag{5.32}$$

For $v \geq 1$ consider the event Ω_v defined by

$$\forall n \geq 1 \,,\ \forall s \in T_n \,,\ \forall s' \in T_{n-1} \,,\ |X_s - X_{s'}| \leq Lv\mathcal{N}_{2^{n-1}}(s - s') \,, \tag{5.33}$$

where L is as in (5.31). Then, using (5.31) and the fact that $\mathrm{card} T_n \cdot \mathrm{card} T_{n-1} \leq N_n N_{n-1} \leq 2^{2^{n+1}}$, we see that

$$\mathsf{P}(\Omega_v^c) \leq p(v) := \sum_{n \geq 1} 2^{2^{n+1}} \exp(-v2^{n-1}) \,. \tag{5.34}$$

From (5.28) and the definition of φ_j we have

$$\forall s \,,\ s' \in A \in \mathcal{A}_n \,, s - s' \in r^{-j(A)+1} B(2^n) \,. \tag{5.35}$$

Since $\pi_n(t), \pi_{n-1}(t) \in A_{n-1}(t)$, by (5.35) we have

$$\pi_n(t) - \pi_{n-1}(t) \in r^{-j(A_{n-1}(t))+1} B(2^{n-1}) \,,$$

so that, by definition of $B(u)$ we have

$$\mathcal{N}_{2^{n-1}}(\pi_n(t) - \pi_{n-1}(t)) \leq 2^{n-1} r^{-j(A_{n-1}(t))+1} \,.$$

By definition of Ω_v, when this event occurs, we see using (5.33) for $s = \pi_n(t)$ and $s' = \pi_{n-1}(t)$ that we have

$$|X_{\pi_n(t)} - X_{\pi_{n-1}}(t)| \leq Lv2^n r^{-j(A_{n-1}(t))+1} \,,$$

and by (5.32) for $t \in T_k$ we have

$$|X_t - X_{t_0}| \leq Lv \sum_{1 \leq n \leq k} 2^n r^{-j(A_{n-1}(t))+1} \,,$$

and thus

$$\sup_{t\in T_k} |X_t - X_{t_0}| \le Lv \sup_{t\in T} \sum_{1\le n\le k} 2^n r^{-j(A_{n-1}(t))+1} ,$$

so that

$$\mathsf{P}\Big(\sup_{t\in T_k} |X_t - X_{t_0}| > Lv \sup_{t\in T} \sum_{1\le n\le k} 2^n r^{-j(A_{n-1}(t))+1}\Big) \le \mathsf{P}(\Omega_v^c) ,$$

and using (5.34) we get

$$\mathsf{E} \sup_{t\in T_k} |X_t - X_{t_0}| \le L \sup_{t\in T} \sum_{1\le n\le k} 2^n r^{-j(A_{n-1}(t))+1} ,$$

which implies the conclusion since k is arbitrary. □

The proof of Proposition 5.2.2 requires several lemmas. For $\lambda \ge 0$ we define $V_i(\lambda) = \sup_x(\lambda x - \hat{U}_i(x))$, so that $V_i(\lambda) < \infty$ for $\lambda < \lambda_i$, where $\lambda_i = \lim_{x\to\infty} \hat{U}_i(x)/x \ge 1$ (that exists by convexity). Taking $x = \lambda/2$ we observe that

$$\lambda \le 2 \Rightarrow V_i(\lambda) \ge \frac{\lambda^2}{4} \tag{5.36}$$

and taking $x = 1$ we get

$$V_i(\lambda) \ge \lambda - 1 . \tag{5.37}$$

Lemma 5.2.3. *For $\lambda \ge 0$ we have*

$$\mathsf{E} \exp \lambda Y_i \le \exp V_i(L\lambda) .$$

Proof. Since $U_i'(1) \ge 1$, for $x \ge 1$ we have $U_i(x) \ge x$, so that by (5.25) we have $\mathsf{P}(|Y_i| \ge x) \le e^{-x}$ and hence

$$\mathsf{E}Y_i^2 \exp \frac{|Y_i|}{2} \le L .$$

We have $e^x \le 1 + x + x^2 e^{|x|}$ so that, if $\lambda \le 1/2$,

$$\begin{aligned} \mathsf{E} \exp \lambda Y_i &\le 1 + \lambda^2 \mathsf{E} X_i^2 \exp \lambda |Y_i| \le 1 + L\lambda^2 \\ &\le \exp L\lambda^2 \le \exp V_i(L'\lambda) \end{aligned}$$

using (5.36).

If $\lambda \ge 1/2$, we have

$$\begin{aligned} \mathsf{E} \exp \lambda |Y_i| &= 1 + \lambda \int_0^\infty \exp \lambda x \, \mathsf{P}(|Y_i| \ge x) \mathrm{d}x \\ &\le 1 + \lambda \int_0^\infty \exp(\lambda x - U_i(x)) \mathrm{d}x . \end{aligned}$$

If $x \le 1$, we have $4\lambda x \le 4\lambda \le 6\lambda - 1 \le V_i(6\lambda)$, using that $\lambda \ge 1/2$ and (5.37). Thus we have

$$\lambda x - U_i(x) \le \lambda x \le \frac{V_i(6\lambda)}{2} - \lambda x \, .$$

If $x \ge 1$ we have $U_i(x) \ge \hat{U}_i(x)/2$ and

$$\lambda x - U_i(x) \le \lambda x - \frac{\hat{U}_i(x)}{2} \le \frac{V_i(4\lambda)}{2} - \lambda x$$

by definition of V_i. Thus, since $V_i(4\lambda) \le V_i(6\lambda)$ we have

$$\begin{aligned} \mathsf{E} \exp \lambda |Y_i| &\le 1 + \lambda \int_0^\infty \exp(\frac{V_i(6\lambda)}{2} - \lambda x) \mathrm{d}x \\ &= 1 + \exp \frac{V_i(6\lambda)}{2} \le 2 \exp \frac{V_i(6\lambda)}{2} \\ &\le \exp V_i(6\lambda) \end{aligned}$$

because $V_i(6\lambda) \ge V_i(3) \ge 2$. □

Lemma 5.2.4. *We have*

$$\sum_{i \ge 1} V_i(\frac{u|t_i|}{\mathcal{N}_u(t)}) \le u \, .$$

Proof. It suffices to show that given numbers $x_i \ge 0$, we have

$$\sum_{i \ge 1} \frac{u t_i x_i}{\mathcal{N}_u(t)} - \sum_{i \ge 1} \hat{U}_i(x_i) \le u \, . \tag{5.38}$$

If $\sum_{i \ge 1} \hat{U}_i(x_i) \le u$, then by definition of $\mathcal{N}_u(t)$ we have $\sum_{i \ge 1} t_i x_i \le \mathcal{N}_u(t)$ so we are done. If $\sum_{i \ge 1} \hat{U}_i(x_i) = \theta u$ with $\theta > 1$, then $\sum_{i \ge 1} \hat{U}_i(x_i/\theta) \le u$, so that $\sum_{i \ge 1} t_i x_i \le \theta \mathcal{N}_u(t)$ and the left-hand side of (5.38) is in fact ≤ 0. □

Lemma 5.2.5. *If $v \ge 1$ we have*

$$\mathcal{N}_{uv}(t) \le v \mathcal{N}_u(t) \, . \tag{5.39}$$

Proof. For $v \ge 1$ we have $\hat{U}_i(a_i/v) \le \hat{U}_i(a_i)/v$. If $\sum_{i \ge 1} \hat{U}_i(a_i) \le uv$ we then have $\sum_{i \ge 1} \hat{U}_i(a_i/v) \le u$ so that by definition of $\mathcal{N}_u$ we have $\sum_{i \ge 1} t_i a_i / v \le \mathcal{N}_u(t)$ and thus $\sum_{i \ge 1} t_i a_i \le v \mathcal{N}_u(t)$. □

Proof of Proposition 5.2.2. Since by Lemma 5.2.5 we have $v\mathcal{N}_u(t) \ge \mathcal{N}_{vu}(t)$, we can assume $v = 1$. Using Lemma 5.2.3 we have

$$\begin{aligned} \mathsf{P}(X_t \ge y) &\le \exp(-\lambda y) \mathsf{E} \exp \lambda X_t \\ &\le \exp\bigl(-\lambda y + \sum_{i \ge 1} V_i(L_0 \lambda |t_i|)\bigr) \, . \end{aligned}$$

We choose $y = 2L_0 \mathcal{N}_u(t)$, $\lambda = 2u/y$, and we apply Lemma 5.2.4 to see that

$$-\lambda y + \sum_{i \geq 1} V_i(L_0 \lambda t_i) \leq -2u + u = -u \, .$$

□

Let us now turn to the converse of Theorem 5.2.1. We assume the following regularity conditions. For some constant C_0, we have

$$\forall i \geq 1 \, , \, \forall s \geq 1 \, , \, U_i(2s) \leq C_0 U_i(s) \, . \tag{5.40}$$

$$\forall i \geq 1 \, , \, U_i'(0) \geq 1/C_0 \, . \tag{5.41}$$

Here, $U_i'(0)$ is the right derivative at 0 of the function $U_i(x)$. Condition (5.40) is often called "the Δ_2 condition".

Theorem 5.2.6. *Under conditions (5.40) and (5.41) we can find r (depending on C_0 only) and a number $K = K(C_0)$ such that for each subset T of ℓ^2 there exists an admissible sequence $(\mathcal{A}_n)$ of T and for $A \in \mathcal{A}_n$ an integer $j(A) \in \mathbb{Z}$ such that (5.28) holds together with*

$$\sup_{t \in T} \sum_{n \geq 0} 2^n r^{-j(A_n(t))} \leq K(C_0) \mathsf{E} \sup_{t \in T} X_t \, . \tag{5.42}$$

Together with Theorem 5.2.1, this essentially allows the computation of $\mathsf{E} \sup_{t \in T} X_t$ as a function of the geometry of T.

Let us interpret this statement in the case where for $x \geq 1$ we have $U_i(x) = x^2$. In that case, and more generally when $U_i(x) \leq x^2/L$ for $x \geq 1$, we have

$$\varphi_j(s,t) \geq r^{2j} \|s-t\|_2^2 / L \, , \tag{5.43}$$

so that (5.28) implies that $\Delta(A, d_2) \leq L 2^{n/2} r^{-j(A)+1}$ and (5.42) shows that

$$\sup_{t \in T} \sum_{n \geq 0} 2^{n/2} \Delta(A_n(t), d_2) \leq Lr \mathsf{E} \sup_{t \in T} X_t \, ,$$

and hence

$$\gamma_2(T, d_2) \leq Lr \mathsf{E} \sup_{t \in T} X_t \, . \tag{5.44}$$

Thus, we have proved (an extension of) Theorem 2.1.1.

Consider now the case where $U_i(x) = x$ for all x. From (5.27) we see that we not only have (5.43), and thus (5.44), but also $\varphi_j(s,t) = \infty$ if $\|s-t\|_\infty \geq Lr^{-j}$. Thus (5.28) implies that $\Delta(A, d_\infty) \leq Lr^{-j(A)+1}$, and (5.42) implies that

$$\gamma_1(T, d_\infty) \leq Lr \mathsf{E} \sup_{t \in T} X_t \, .$$

Recalling (5.30) (and since here r is a universal constant) we thus have proved the following very pretty fact.

Theorem 5.2.7. *Assume that the r.v. Y_i are independent, symmetric and satisfy $\mathsf{P}(|Y_i| \ge x) = \exp(-x)$. Then we have*

$$\frac{1}{L}(\gamma_2(T,d_2)+\gamma_1(T,d_\infty)) \le \mathsf{E}\sup_{t\in T} X_t \le L(\gamma_2(T,d_2)+\gamma_1(T,d_\infty))\,.$$

We turn to the proof of Theorem 5.2.6.

Lemma 5.2.8. *Under (5.40), given $\rho > 0$ we can find r_0, depending on C_0 and ρ only, such that if $r \ge r_0$, for $u \in \mathbb{R}^+$ we have*

$$B_{j+1}(0,16ru) \subset \rho B_j(0,u)\,, \tag{5.45}$$

where $B_j(t,u) = \{s\,;\ \varphi_j(s,t) \le u\} = s + r^{-j}B(u)$.

Proof. It suffices to prove that

$$B(16ru) \subset \rho r B(u)\,. \tag{5.46}$$

Indeed if $t \in B_{j+1}(0,16ru)$, then we have $t \in r^{-j-1}B(16ru) \subset \rho r^{-j}B(u) = \rho B_j(0,u)$ by definition of φ_j.

To prove (5.46), consider $t \in B(16ru)$. Then $\mathcal{N}_{16ru}(t) \le 16ru$ by definition, so that for any numbers $(a_i)_{i\ge1}$ we have

$$\sum_{i\ge1}\hat{U}_i(a_i) \le 16ru \Rightarrow \sum_{i\ge1} a_i t_i \le 16ru\,. \tag{5.47}$$

From (5.40) we see that for some constant C_1, depending only on C_0 we have

$$\forall u>0\,,\ \hat{U}_i(2u) \le C_1\hat{U}_i(u)\,. \tag{5.48}$$

Consider an integer k large enough that $2^{-k+4} \le \rho$ and let $r_0 = C_1^k$. Assume that $r \ge r_0$ and consider numbers b_i with $\sum_{i\ge1}\hat{U}_i(b_i) \le u$. Then by (5.48) we have $\hat{U}_i(2^k b_i) \le C_1^k\hat{U}_i(b_i) \le r\hat{U}(b_i)$, so that $\sum_{i\ge1}\hat{U}_i(2^k b_i) \le ru \le 16ru$, and by (5.47) we have $\sum_{i\ge1} 2^k b_i t_i \le 16ru$ i.e. we have shown that

$$\sum_{i\ge1}\hat{U}_i(b_i) \le u \Rightarrow \sum_{i\ge1} b_i\frac{t_i}{\rho r} \le u$$

so that $\mathcal{N}_u(t/\rho r) \le u$ and thus $t/\rho r \in B(u)$. □

Theorem 5.2.9. *Under Condition (5.41) we can find a number $\rho > 0$ with the following property. Given any points $t_1,\cdots,t_m$ in ℓ^2 such that*

$$\ell \ne \ell' \Rightarrow t_\ell - t_{\ell'} \notin B(u) \tag{5.49}$$

and given any sets $H_\ell \subset t_\ell + \rho B(u)$, we have

$$\mathsf{E}\sup_{t\in\bigcup H_\ell} X_t \ge \frac{1}{L}\min(u,\log m) + \min_{\ell\le m}\mathsf{E}\sup_{t\in H_\ell} X_t\,. \tag{5.50}$$

The proof of this statement is very similar to the proof of (2.8). The first ingredient is a suitable version of Sudakov minoration, asserting that, under (5.49)

$$\mathsf{E} \sup_{\ell \le m} X_{t_\ell} \ge \frac{1}{L} \min(u, \log m) \tag{5.51}$$

and the second is a "concentration of measure" result that quantifies the deviation of $\sup_{t \in H_\ell} X_t$ from its mean. Condition (5.41) is used there, to assert that the law of Y_i is the image of the probability ν of density $e^{-2|x|}$ with respect to Lebesgue measure by a Lipschitz map. This allows to apply the result of concentration of measure concerning ν first proved in [47]. Since neither of these arguments is closely related to our main topic, we refer the reader to [54] and [16].

Proof of Theorem 5.2.6. Consider the functionals $F_{n,j}(A) = 2L\mathsf{E}\sup_{t\in A} X_t$, where L is the constant of (5.50). Consider $n \ge 1$. We use (5.50) with $u = 2^n$, $m = N_n$ and homogeneity to see that if

$$\ell \ne \ell' \Rightarrow t_\ell - t_{\ell'} \notin r^{-j} B(2^n) \tag{5.52}$$

then given any sets $H_\ell \subset t_\ell + r^{-j}\rho B(2^n)$ we have

$$F_{n,j}\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge 2^n r^{-j} + \min_{\ell \le m} F_{n+1,j+1}(H_\ell),$$

because $\log m = 2^n \log 2 \ge 2^{n-1}$. The definition of φ_j shows that (5.3) coincides with (5.52). If $r = 2^{\kappa-4}$, where κ is large enough (depending on C_0 only), Lemma 5.2.8 used for $u = 2^n$ shows that $B(2^{\kappa+n}) \subset \rho r B(2^n)$ and the condition $H_\ell \subset t_\ell + r^{-j}\rho B(2^n)$ follows from the condition $H_\ell \subset B_{j+1}(t_\ell, 2^{\kappa+n}) = t_\ell + r^{-j-1}B(2^{\kappa+n})$. So we have proved that (5.4) holds true for $n \ge n_0 = 1$ under (5.2), i.e. we have proved that the growth condition of Definition 5.1.1 holds true (for n_0 and r).

From (5.51) we see that if $s\,,\, t \in T\,,\; s - t \notin aB(1)$, then

$$\frac{a}{L} \le \mathsf{E}\max(X_s, X_t) \le \mathsf{E}\sup_{t \in T} X_t\,.$$

Thus if j_0 denotes the largest integer such that $r^{-j_0+1} > L\mathsf{E}\sup_{t\in T} X_t$, for $s\,,\, t \in T$ we have $s - t \in r^{-j_0+1}B(1)$ and thus $\varphi_{j_0-1}(s,t) \le 1$, that is (5.7) holds for $n_0 = 1$ and this value of j_0. Thus we are in a position to apply Theorem 5.1.2. Setting $j(T) = j_0$, and recalling that $\mathcal{A}_0 = \{T\}$, we then see that (5.42) follows from (5.9), and we prove (5.28). By definition of $B_j(t,u)$ and of φ_j, we have

$$s \in B_{j-1}(t,u) \Rightarrow \varphi_{j-1}(s,t) \le u \Rightarrow s - t \in r^{-j+1}B(u)\,.$$

Thus, from (5.10) we have

$$\forall n \ge 1\,,\; \forall A \in \mathcal{A}_n\,,\; \forall s \in A\,,\; s - t_A \in r^{-j(A)+1}B(2^n)\,.$$

Since $B(u)$ is a convex symmetric set, we have

$$s - t_A \in r^{-j(A)+1}B(2^n),\ s' - t_A \in r^{-j(A)+1}B(2^n) \Rightarrow \frac{s-s'}{2} \in r^{-j(A)+1}B(2^n)$$
$$\Rightarrow \varphi_{j(A)-1}\Big(\frac{s}{2}, \frac{s'}{2}\Big) \le 2^n\,,$$

and finally

$$\forall n \ge 1\,, \forall A \in \mathcal{A}_n\,, \forall s, s' \in A\,, \varphi_{j(A)-1}\Big(\frac{s}{2}, \frac{s'}{2}\Big) \le 2^n\,.$$

This is not exactly (5.28), but of course to get rid of the factor $1/2$ it would have sufficed to apply the above proof to $2T = \{2t\,;\, t \in T\}$ instead of T. □

As a consequence of Theorems 5.2.1 and 5.2.6, we have the following geometrical result. Consider a set $T \subset \ell^2$, an admissible sequence $(\mathcal{A}_n)$ of T and for $A \in \mathcal{A}_n$ an integer $j(A)$ such that (5.28) holds true. Then there is an admissible sequence $(\mathcal{B}_n)$ of conv T and for $B \in \mathcal{B}_n$ an integer $j(B)$ that satisfies (5.28) and

$$\sup_{t \in \mathrm{conv} T} \sum_{n \ge 0} 2^n r^{-j(B_n(t))} \le K(C_0) \sup_{t \in T} \sum_{n \ge 0} 2^n r^{-j(A_n(t))}\,.$$

Research problem 5.2.10. Give a geometrical proof of this fact.

This is a far-reaching generalization of Research Problem 2.1.9.

The following generalizes Theorem 2.1.8.

Theorem 5.2.11. *Assume (5.40) and (5.41). Consider a countable subset T of ℓ^2, with $0 \in T$. Then we can find a sequence (x_n) of vectors of ℓ^2 such that*

$$T \subset \mathrm{conv}\{x_n\,,\, n \ge 2\} \cup \{0\}$$

and, for each n,

$$\mathcal{N}_{\log n}(x_n) \le K(C_0) \mathsf{E} \sup_{t \in T} X_t\,.$$

To appreciate this result, one should note that, by (5.31), if the sequence $(x_n)_{n\ge 2}$ satisfies $\mathcal{N}_{\log n}(x_n) \le 1$, then $\mathsf{E}\sup_{n \ge 2} X_{x_n} \le L$.

Proof. We consider a sequence of partitions of T as provided by Theorem 5.2.6. We choose $t_T = 0$, and for $A \in \mathcal{A}_n$, $n \ge 1$ we select $t_A \in \mathcal{A}_n$, making sure (as in the proof of Theorem 2.1.1) that each point of T is of the form t_A for a certain A. For $A \in \mathcal{A}_n$, $n \ge 1$, we denote by A' the unique element of $\mathcal{A}_{n-1}$ that contains A.

We define

$$u_A = \frac{t_A - t_{A'}}{2^n r^{-j(A')+1}}$$

and $U = \{u_A \,,\, A \in \mathcal{A}_n \,,\, n \geq 1\}$. Consider $t \in T$, so that $t = t_A$ for some n and some $A \in \mathcal{A}_n$, and, since $A_0(t) = T$ and $t_T = 0$,

$$t = t_A = \sum_{1 \leq k \leq n} t_{A_k(t)} - t_{A_{k-1}(t)} = \sum_{1 \leq k \leq n} 2^k r^{-j(A_{k-1}(t))+1} u_{A_k(t)} \,.$$

Since $\sum_{k \geq 0} 2^k r^{-j(A_k(t))} \leq K(C_0) \mathsf{E} \sup_{t \in T} X_t$ by (5.42), this shows that

$$T \subset (K(C_0) \mathsf{E} \sup_{t \in T} X_t) U \,.$$

Next, we prove that $\mathcal{N}_{2^{n+1}}(u_A) \leq 2$ whenever $A \in \mathcal{A}_n$. By (5.28) and the definition of φ_j we have by homogeneity of $\mathcal{N}_u$,

$$\forall s \,,\, s' \in A \,,\, s - s' \in r^{-j(A)+1} B(2^n) \,.$$

Since $t_A, t_{A'} \in A'$, using this for A' instead of A and using the definition of $B(u)$ we have

$$\mathcal{N}_{2^{n-1}}(t_A - t_{A'}) \leq 2^{n-1} r^{-j(A')+1} \,,$$

and thus $\mathcal{N}_{2^{n-1}}(u_A) \leq 1/2$, so that $\mathcal{N}_{2^{n+1}}(u_A) \leq 2$ using (5.39).

Let us enumerate $U = (y_n)_{n \geq 2}$ in such a manner that the points of the type u_A for $A \in \mathcal{A}_1$ are enumerated before the points of the type u_A for $A \in \mathcal{A}_2$, etc. Then if $y_n = u_A$ for $A \in \mathcal{A}_k$, we have $n \leq N_0 + N_1 + \ldots + N_k \leq N_k^2$ and $\log n \leq 2^{k+1}$. Thus $\mathcal{N}_{\log n}(y_n) \leq \mathcal{N}_{2^{k+1}}(y_n) = \mathcal{N}_{2^{k+1}}(u_A) \leq 2$.We then set $x_n = y_n K(C_0) \mathsf{E} \sup_{t \in T} X_t$ $\square$

It is not very difficult to prove that Theorem 5.2.6 still holds true without condition (5.41), and this is done in [16]. But it is an entirely different matter to remove condition (5.40). In fact, it is not difficult to see that if one could obtain (5.42) without condition (5.40) (and a universal constant instead of $K(C_0)$), one would have a positive answer to the Bernoulli conjecture.

5.3 Lower Bounds for Infinitely Divisible Processes

If T is a finite set, a stochastic process $(X_t)_{t \in T}$ is called (real, symmetric, without Gaussian component) infinitely divisible if there exists a positive measure ν on $\mathbb{R}^T$ such that $\int_{\mathbb{R}^T} (\beta(t)^2 \wedge 1) \mathrm{d}\nu(\beta) < \infty$ for all t in T, and such that for all families $(\alpha_t)_{t \in T}$ of real numbers we have

$$E \exp i \sum_{t \in T} \alpha_t X_t = \exp\Big(- \int_{\mathbb{R}^T} \Big(1 - \cos\big(\sum_{t \in T} \alpha_t \beta(t) \big) \Big) \mathrm{d}\nu(\beta) \Big) \,. \qquad (5.53)$$

Here, as in the rest of this section, $x \wedge 1 = \min(x, 1)$. The positive measure ν is called the Lévy measure of the process. To get a feeling for this formula, consider the case where ν consists of a mass a at a point $\beta \in \mathbb{R}^T$. Then, in distribution, we have $(X_t)_{t\in T} = (\beta(t)(Y - Y'))_{t\in T}$ where Y and Y' are independent Poisson r.v. of expectation $a/2$. One can then view the formula (5.53) as saying that the general case is obtained by taking a (kind of infinite) sum of independent processes of the previous type.

The reason for which we exclude Gaussian components is that they are well understood, as was seen in Chapter 2.

The reason why we consider only the symmetric case is that this is essentially not a restriction, using the symmetrization procedure that we have met several times, i.e. replacing the process $(X_t)_{t\in T}$ by the process $(X_t - X'_t)_{t\in T}$ where $(X'_t)_{t\in T}$ is an independent copy of $(X_t)_{t\in T}$.

For the type of inequalities we wish to prove, it is not a restriction to assume that T is finite, which we will do for simplicity. But for the purpose of this introduction it will be useful to consider also the case where T is infinite. In that case, we still say that the process $(X_t)_{t\in T}$ is infinitely divisible if (5.53) holds for each family $(\alpha_t)_{t\in T}$ such that only finitely many coefficients are not 0. Now ν is a "cylindrical measure" that is known through its projections on $\mathbb{R}^S$ for S finite subset of T, projections that are positive measures (and satisfy the obvious compatibility conditions).

An infinitely divisible process indexed by T is thus parameterized by a cylindrical measure on $\mathbb{R}^T$ (with the sole restriction that $\int (\beta(t)^2 \wedge 1)\mathrm{d}\nu(\beta) < \infty$ for each $t \in T$). This is a *huge* class, and only some extremely special subclasses have yet been studied in any detail. The best known class is that of infinitely divisible process with *stationary increments*. Then $T = \mathbb{R}^+$ and ν is the image of $\mu \otimes \lambda$ under the map $(x, u) \mapsto (x\mathbf{1}_{\{t\geq u\}})_{t\in\mathbb{R}^+}$, where μ is a positive measure on $\mathbb{R}$ such that $\int (x^2 \wedge 1)\mathrm{d}\mu(x) < \infty$ and where λ is Lebesgue measure. More likely than not a probabilist selected at random (!) will think that infinitely divisible process are intrinsically discontinuous. This is simply because he has this extremely special case as a mental picture. As will be apparent later (through Rosinski's representation) discontinuity in this case is created by the fact that ν is supported by the discontinuous functions $t \mapsto x\mathbf{1}_{\{t\geq u\}}$ and is certainly not intrinsic to infinitely divisible processes. In fact, some lesser known classes of infinitely divisible processes studied in the literature, such as moving averages (see e.g. [24]) are often continuous. They are still very much more special than the structures we consider.

Continuity will not be studied here, and was mentioned simply to stress that we deal here with hugely general and complicated structures, and it is almost surprising that so much can be said about them.

Unfortunately to prove lower bounds on infinitely divisible processes, we need the following regularity condition. (On the other hand, the upper bound of Theorem 5.4.5 requires no special conditions.)

Definition 5.3.1. *Consider $\delta > 0$ and $C_0 > 0$. We say that condition $H(C_0, \delta)$ holds if for all $s, t \in T$, and all $u > 0$, $v > 1$ we have*

$$\nu(\{\beta\ ;\ |\beta(s) - \beta(t)| \geq uv\}) \leq C_0 v^{-1-\delta} \nu(\{\beta\,;\,|\beta(s) - \beta(t)| \geq u\})\ .$$

It should be observed that without loss of generality one can assume that $\delta < 1$. This will be assumed henceforth.

Condition $H(C_0, \delta)$ is certainly annoying, since it rules out important cases, such as the case where the image of ν under the map $\beta \mapsto \beta(t)$ charges only one point (i.e. X_t is a symmetrized Poisson r.v.).

It is probably an interesting and difficult program to investigate what happens when condition $H(C_0, \delta)$ does not hold.

A large class of measures ν that satisfy condition $H(C_0, \delta)$ can be constructed as follows. Consider a measure μ on $\mathbb{R}$, and assume that

$$\forall u > 0, \forall v > 1\,,\ \mu(\{x\,;\,|x| \geq uv\}) \leq C_0 v^{-1-\delta} \mu(\{x\,;\,|x| \geq u\})\,. \tag{5.54}$$

Consider a probability measure m on $\mathbb{R}^T$. Assume that

$$\nu \text{ is the image of } \mu \otimes m \text{ under the map } (x, \gamma) \mapsto x\gamma\,. \tag{5.55}$$

Then ν satisfies condition $H(C_0, \delta)$. This follows from (5.54) using the formula

$$\nu(\{\beta\,;\,|\beta(s) - \beta(t)| \geq u\}) = \int \mu(\{x\,;\,|x||\gamma(s) - \gamma(t)| \geq u\}) \mathrm{d}m(\gamma)\,.$$

An important special class of measures that satisfy (5.54) is obtained when μ has density x^{-p-1} with respect to Lebesgue measure on $\mathbb{R}^+$, and $1 < p < 2$. In that case, if the Lévy measure is given by (5.55), the process in $(X_t)_{t\in T}$ is p-stable. To see this, we observe the formula

$$\int_{\mathbb{R}^+} (1 - \cos(ax)) x^{-p-1} \mathrm{d}x = C(p)|a|^p\,,$$

that is obvious through change of variable. Then, for each real λ we have

$$\begin{aligned}
&\int_{\mathbb{R}^T} \Big(1 - \cos\big(\lambda \sum_{t\in T} \alpha_t \beta(t)\big)\Big) \mathrm{d}\nu(t) \\
&= \int_{\mathbb{R}^T} \int_{\mathbb{R}^+} \Big(1 - \cos\big(\lambda x \sum_{t\in T} \alpha_t \gamma(t)\big)\Big) x^{-p-1} \mathrm{d}x \mathrm{d}m(\gamma) \\
&= |\lambda|^p \frac{\sigma^p}{2}\,,
\end{aligned} \tag{5.56}$$

where

$$\sigma^p = 2C(p) \int_{\mathbb{R}^T} \big|\sum_{t\in T} \alpha_t \gamma(t)\big|^p \mathrm{d}m(\gamma)\,, \tag{5.57}$$

and (5.53) and (2.31) show that the r.v. $\sum_{t\in T} \alpha_t X(t)$ is p-stable.

The functions

$$\varphi(s,t,u) = \int_{\mathbb{R}^T} \big((u^2(\beta(s) - \beta(t))^2) \wedge 1\big) \mathrm{d}\nu(\beta)$$

will help measure the size of T.

For any fixed number u, $\varphi(s,t,u)$ is the square of a distance on T, so that for elements s, s', t of T we have

$$\varphi(s,s',u)^{1/2} \le \varphi(s,t,u)^{1/2} + \varphi(s',t,u)^{1/2},$$

and, using the inequality $(a+b)^2 \le 2(a^2+b^2)$ we have

$$\varphi(s,s',u) \le 2(\varphi(s,t,u) + \varphi(s',t,u)). \tag{5.58}$$

Theorem 5.3.2. *Under Condition $H(C_0,\delta)$, there exists a number $r \ge 4$ (depending only on C_0 and δ), an admissible sequence of partitions $\mathcal{A}_n$ and for $A \in \mathcal{A}_n$ a number $j(A) \in \mathbb{Z}$ such that (5.8) holds together with*

$$\forall n \ge 0\,,\ \forall A \in \mathcal{A}_n\,,\ \forall s\,,\ s' \in A\,,\ \varphi(s,s',r^{j(A)-1}) \le 2^{n+1} \tag{5.59}$$

$$\forall t \in T\,,\ \sum_{n\ge 0} 2^n r^{-j(A_n(t))} \le K \mathsf{E} \sup_{t\in T} X_t\,. \tag{5.60}$$

Here, and everywhere in this section, K denotes a number that depends on C_0 and δ only and that need not be the same at each occurrence.

Of course the level of abstraction reached here might make it hard for the reader to understand the content of Theorem 5.3.2. As we will see in the next section, this theorem is extremely precise. It exactly captures a certain aspect of the process, in the sense that the necessary condition of this theorem is sufficient to imply the boundedness of an "important part" of the process.

To gain a first understanding of Theorem 5.3.2, let us prove that in the case where ν is obtained as in (5.55) and where μ has density x^{-p-1} on $\mathbb{R}^+$ with respect to Lebesgue measure, we recover Theorem 2.3.1. By change of variable, it is obvious that

$$\int_{\mathbb{R}^+} ((ax)^2 \wedge 1) x^{-p-1} \mathrm{d}x = C_1(p)|a|^p\,,$$

so that

$$\begin{aligned}\varphi(s,t,u) &= \int_{\mathbb{R}^T}\int_{\mathbb{R}^+} \Big(\big(xu(\gamma(s)-\gamma(t))\big)^2 \wedge 1 \Big) x^{-p-1} \mathrm{d}x \mathrm{d}m(\gamma) \\ &= C_1(p) u^p \int_{\mathbb{R}^T} |\gamma(s)-\gamma(t)|^p \mathrm{d}m(\gamma)\,.\end{aligned}$$

Using (5.56) and (5.57) when $\sum_{t\in T}\alpha_t\beta(t) = \beta(t)-\beta(s)$ and comparing with (2.31) and (2.33) we see that

$$\varphi(s,t,u) = C_2(p) u^p d^p(s,t)\,,$$

so that (5.59) implies $\Delta(A,d) \le K 2^{n/p} r^{-j(A)+1}$, and (5.60) implies

$$\sum_{n\geq 0} 2^{n/q}\Delta(A_n(t),d) \leq K\mathsf{E}\sup_{t\in T} X_t\,,$$

where $1/q = 1-1/p$ and thus $\gamma_q(T,d) \leq K\mathsf{E}\sup_{t\in T} X_t$.

It is possible to prove a result that is conceivably stronger than Theorem 5.3.2. We know how to replace in (5.60) the upper bound $K\mathsf{E}\sup_{t\in T} X_t$ by KM where M is such that $\mathsf{P}(\sup_{t\in T} X_t \leq M) \geq 1-\varepsilon_0$, where ε_0 is a certain number (depending on C_0 and δ). Carrying out this improvement is straightforward, but cumbersome, and, for clarity, we will refrain to do it. Moreover, it is believable (although we do not know how to prove it) that when Condition $H(C_0,\delta)$ holds, there always exists a number ϵ_1 (depending on C_0 and δ only) such that if $\mathsf{P}(\sup_{t\in T} X_t \leq M) \geq 1-\epsilon_1$, then $\mathsf{E}\sup_{t\in T} X_t \leq KM$.

The key to Theorem 5.3.2 is that infinitely divisible processes can be represented as a mixture of Bernoulli processes. Since we do not understand Bernoulli processes as well as we understand Gaussian processes we can expect that the proof of Theorem 5.3.2 will be more difficult than the proof of Theorem 2.3.1. We can also expect that there should be a strong correlation between the possibility of extending Theorem 5.3.2 beyond condition $H(C_0,\delta)$ and a better understanding of Bernoulli processes. While our proof of Theorem 5.3.2 is very much simpler than the original proof of [53], it certainly remains one of the hardest of this work. The reader should probably learn first the remarkable consequences of this theorem explained in Section 5.4 to gather the energy required to penetrate this proof.

We now start the description of this representation of infinitely divisible processes as a mixture of Bernoulli processes. We denote by $(\tau_i)_{i\geq 1}$ the sequence of arrival times of a Poisson process of parameter 1. Equivalently, $\tau_i = \Gamma_1 + \cdots + \Gamma_i$, where the sequence $(\Gamma_k)_{k\geq 1}$ is i.i.d. and $\mathsf{P}(\Gamma_k \geq u) = e^{-u}$. Consider a measurable function $G : \mathbb{R}^+ \times \mathbb{R}^T \to \mathbb{R}^T$. Consider a probability measure m on $\mathbb{R}^T$, and denote Lebesgue's measure on $\mathbb{R}^+$ by λ. We denote by $(Y_i)_{i\geq 1}$ an i.i.d. sequence of $\mathbb{R}^T$-valued r.v, distributed like m, and by $(\epsilon_i)_{i\geq 1}$ a Bernoulli sequence. We assume that the sequences (τ_i), (ϵ_i), (Y_i) are independent of each other.

Theorem 5.3.3. *(Rosinski [35]) Denote by ν the image measure of $\lambda\otimes m$ under G, and assume that it is a Lévy measure, i.e. that $\int_{\mathbb{R}^T}(\beta(t)^2\wedge 1)\mathrm{d}\nu(\beta) < \infty$ for each t in T. Then the series $\sum_{i\geq 1}\epsilon_i G(\tau_i, Y_i)$ converges a.e. in $\mathbb{R}^T$ and its law is the law of the symmetric infinitely divisible process of Lévy measure ν.*

In practice, we are not given λ and m, but ν. There are many ways to represent ν as the image of a product $\lambda\otimes m$ under a measurable transformation. One particular method is very fruitful (and is also brought to light in [35]). Consider a probability measure m such that ν is absolutely continuous with respect to m. (There are of course many possible choices, but, remarkably enough, the particular choice is irrelevant.) Consider a Radon-Nikodym derivative g of ν with respect to m and define $G(u,\beta) = R(u,\beta)\beta$ where

$$R(u,\beta) = \mathbf{1}_{[0,g(\beta)]}(u) \ . \tag{5.61}$$

For simplicity we write $R_i = R(\tau_i, Y_i)$. It follows from Theorem 5.3.3 that

$$\sum_{i\geq 1} \epsilon_i R_i Y_i \tag{5.62}$$

is distributed like the infinitely divisible process of Lévy measure ν. This representation of the process will be called *Rosinski's representation.* Let us note that

$$R_i \in \{0,1\} \ ; \ R_i \text{ is a non-increasing function of } \ \tau_i \ . \tag{5.63}$$

Conditionally on the sequence $(\tau_i)_{i\geq 1}$, the sequence $(R_i Y_i)_{i\geq 1}$ is independent. The influence of the sequence $(\tau_i)_{i\geq 1}$ will be felt only through the two quantities (that exist from the law of large numbers)

$$\alpha_- = \min_{i\geq 1} \frac{\tau_i}{i} \ ; \ \alpha_+ = \max_{i\geq 1} \frac{\tau_i}{i} \ . \tag{5.64}$$

Conditionally in the sequences $(\tau_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$, the process (5.62) is a Bernoulli process. Before we study it, however, we must understand better the behavior of the sequence $(R_i Y_i)_{i\geq 1}$. We start a series of simple lemmas to this effect. The first one is obvious.

Lemma 5.3.4. *Consider $\alpha > 0$ and a non-increasing function θ on $\mathbb{R}^+$. Then*

$$\alpha \sum_{i\geq 1} \theta(\alpha i) \leq \int_0^\infty \theta(x)\mathrm{d}\lambda(x) \leq \alpha \sum_{i\geq 0} \theta(\alpha i) \ . \tag{5.65}$$

Since $R(x,\beta) \in \{0,1\}$, the following is also obvious.

Lemma 5.3.5. *If $h(0) = 0$ then*

$$h(R(x,\beta)\beta) = R(x,\beta)h(\beta) \ . \tag{5.66}$$

We denote by E_τ expectation given the sequence $(\tau_i)_{i\geq 1}$.

Lemma 5.3.6. *Consider a non-negative measurable function h on $\mathbb{R}^T$, with $h(0) = 0$. Then*

$$\frac{1}{\alpha_+}\int_{\mathbb{R}^T} h(\beta)\mathrm{d}\nu(\beta) - \sup h \leq \sum_{i\geq 1} \mathsf{E}_\tau h(R_i Y_i) \leq \frac{1}{\alpha_-}\int_{\mathbb{R}^T} h(\beta)\mathrm{d}\nu(\beta) \ . \tag{5.67}$$

Proof. Given β, the function $\theta(x) = h(R(x,\beta)\beta)$ is non-increasing since its value is $h(\beta) \geq 0$ for $u \leq g(\beta)$ and $h(0) = 0$ for $u > g(\beta)$. Thus by (5.65) we have

$$\begin{aligned}\sum_{i\geq 1} h(R(\tau_i,\beta)\beta) + \sup h &\geq \sum_{i\geq 0} h(R(\alpha_+ i,\beta)\beta)\\ &\geq \frac{1}{\alpha^+}\int_0^\infty h(R(x,\beta)\beta)\mathrm{d}\lambda(x)\ .\end{aligned}$$

Since Y_i is distributed like m and since ν is the law of $\lambda \otimes m$ under the map $(x,\beta) \mapsto R(x,\beta)\beta$, integrating both sides in β with respect to m yields the left-hand side of (5.67). The right-hand side is similar. $\square$

The following simple fact can be deduced from Bernstein's inequality, but it is amusing to give a direct proof.

Lemma 5.3.7. *Consider independent r.v.* $(W_i)_{i\geq 1}$, *with* $0 \leq W_i \leq 1$. *Then*
(a) if $4A \leq \sum_{i\geq 1} \mathsf{E} W_i$, *then*

$$\mathsf{P}\Big(\sum_{i\geq 1} W_i \leq A\Big) \leq \exp(-A)$$

(b) if $A \geq 4\sum_{i\geq 1} \mathsf{E} W_i$, *then*

$$\mathsf{P}\Big(\sum_{i\geq 1} W_i \geq A\Big) \leq \exp\Big(-\frac{A}{2}\Big)\ .$$

Proof. (a) Since $1 - x \leq e^{-x} \leq 1 - x/2$ for $0 \leq x \leq 1$ we have

$$\mathsf{E}\exp(-W_i) \leq 1 - \frac{\mathsf{E} W_i}{2} \leq \exp\Big(-\frac{\mathsf{E} W_i}{2}\Big)$$

and thus

$$\mathsf{E}\exp\Big(-\sum_{i\geq 1} W_i\Big) \leq \exp\Big(-\frac{1}{2}\sum_{i\geq 1}\mathsf{E} W_i\Big) \leq \exp(-2A)$$

and we use the inequality $\mathsf{P}(Z \leq A) \leq \exp A\ \mathsf{E}\exp(-Z)$.

(b) Observe that $1 + x \leq e^x \leq 1 + 2x$ for $0 \leq x \leq 1$, so, as before,

$$\mathsf{E}\exp\sum_{i\geq 1} W_i \leq \exp 2\sum_{i\geq 1}\mathsf{E} W_i \leq \exp\frac{A}{2}$$

and we use now that $\mathsf{P}(Z \geq A) \leq \exp(-A)\mathsf{E}\exp Z$. $\square$

We denote by P_τ conditional probability given the sequence $(\tau_i)_{i\geq 1}$.

Lemma 5.3.8. *(a) Assume that* $2\alpha_+ \le \varphi(s,t,u)$*. Then*

$$\mathsf{P}_\tau\Big(\sum_{i\ge 1}\big(R_i u^2 (Y_i(s)-Y_i(t))^2\big)\wedge 1 \le \frac{\varphi(s,t,u)}{8\alpha_+}\Big) \le \exp\Big(-\frac{\varphi(s,t,u)}{8\alpha_+}\Big)\ . \quad (5.68)$$

(b) If $A \ge 4\varphi(s,t,u)/\alpha_-$ *then*

$$\mathsf{P}_\tau\Big(\sum_{i\ge 1}\big(R_i u^2 (Y_i(s)-Y_i(t))^2\big)\wedge 1 \ge A\Big) \le \exp\Big(-\frac{A}{2}\Big)\ . \quad (5.69)$$

Proof. We set $W_i = (R_i u^2 (Y_i(s)-Y_i(t))^2)\wedge 1$. We use Lemma 5.3.6 with $h(\beta) = u^2(\beta(s)-\beta(t))^2 \wedge 1$ and the definition of φ to get that, under the assumptions of (a),

$$\frac{1}{2\alpha_+}\varphi(s,t,u) \le \frac{1}{\alpha_+}\varphi(s,t,u) - 1 \le \sum_{i\ge 1}\mathsf{E}_\tau W_i \le \frac{1}{\alpha_-}\varphi(s,t,u)$$

and we use Lemma 5.3.7 to conclude. □

We now explore some consequences of Condition $H(C_0,\delta)$. We recall that K denotes a quantity depending on C_0 and δ only.

Lemma 5.3.9. *Under condition* $H(C_0,\delta)$*, for* s , $t\in T$ *and* $u>0$ *, we have, for* $v\ge 1$*,*

$$\varphi(s,t,uv) \ge \frac{v^{1+\delta}}{K}\varphi(s,t,u)\ .$$

Proof. We write $f(\beta) = |\beta(s)-\beta(t)|$, so that

$$\begin{aligned}\varphi(s,t,u) &= \int \big((u^2 f^2(\beta))\wedge 1\big)\mathrm{d}\nu(\beta)\\ &= \int_0^1 \nu(\{\beta\,;\, u^2 f^2(\beta)\ge x\})\mathrm{d}x\\ &= \int_0^{1/v^2} \nu(\{\beta\,;\, u^2 f^2(\beta)\ge x\})\mathrm{d}x + \int_{1/v^2}^1 \nu(\{\beta\,;\, u^2 f^2(\beta)\ge x\})\mathrm{d}x\ .\end{aligned}$$

Setting $x = y/v^2$ in the first term and $x = y^2/v^2$ in the second term we get

$$\varphi(s,t,u) = \frac{1}{v^2}\varphi(s,t,uv) + \frac{1}{v^2}\int_1^v 2y\nu\Big(\{\beta\,;\, f(\beta)\ge \frac{y}{uv}\}\Big)\mathrm{d}y\ . \quad (5.70)$$

Now, by condition $H(C_0,\delta)$, for $y\ge 1$ we have

$$\begin{aligned}\nu(\{\beta\,;\, f(\beta)\ge \frac{y}{uv}\}) &\le C_0 y^{-1-\delta}\nu(\{\beta\,;\, uvf(\beta)\ge 1\})\\ &\le C_0 y^{-1-\delta}\varphi(s,t,uv)\ .\end{aligned}$$

Since we assume $\delta < 1$, substitution in (5.70) yields

$$\varphi(s,t,u) \le \frac{K}{v^2}\varphi(s,t,uv)(1+\int_1^v y^{-\delta}\mathrm{d}y) \le Kv^{-1-\delta}\varphi(s,t,uv)\ .$$

□

Lemma 5.3.10. *Assume condition $H(C_0,\delta)$. Consider s, $t \in T$ and $u > 0$. Set*

$$W_i = R_i|Y_i(s) - Y_i(t)|\ .$$

Then

$$u^{1+\delta/2}\sum_{i\ge 1}\mathsf{E}_\tau W_i^{1+\delta/2}\mathbf{1}_{\{uW_i\ge 1\}} \le \frac{K}{\alpha_-}\varphi(s,t,u)\ . \tag{5.71}$$

Proof. We set $f(\beta) = |\beta(s) - \beta(t)|$. We use Lemma 5.3.6 with the function $h(\beta) = (uf(\beta))^{1+\delta/2}\mathbf{1}_{\{uf(\beta)\ge 1\}}$ to see that the left hand side of (5.71) is bounded by

$$\frac{1}{\alpha_-}\int_{\{uf(\beta)\ge 1\}}(uf(\beta))^{1+\delta/2}\mathrm{d}\nu(\beta) \le \mathrm{I} + \mathrm{II}$$

where

$$\mathrm{I} = \frac{1}{\alpha_-}\nu(\{\beta\,;\,uf(\beta)\ge 1\}) \le \frac{1}{\alpha_-}\varphi(s,t,u)$$
$$\mathrm{II} = \frac{1}{\alpha_-}\int_1^\infty \nu(\{\beta\,;\,(uf(\beta))^{1+\delta/2}\ge x\})\mathrm{d}x\ .$$

Now, by condition $H(C_0,\delta)$, for $x \ge 1$ we have

$$\begin{aligned}\nu(\{\beta\,;\,(uf(\beta))^{1+\delta/2}\ge x\}) &= \nu(\{\beta\,;\,uf(\beta)\ge x^{1/(1+\delta/2)}\})\\ &\le C_0 x^{-\frac{1+\delta}{1+\delta/2}}\nu(\{\beta\,;\,uf(\beta)\ge 1\})\end{aligned}$$

and since $\nu(\{\beta\,;\,uf(\beta)\ge 1\}) \le \varphi(s,t,u)$ the result follows. □

Lemma 5.3.11. *Consider independent r.v. $(W_i)_{i\ge 1}$ such that $W_i \ge 0$, and consider $0 < \delta < 2$. Assume that for certain numbers u, $S > 0$ we have*

$$u^{1+\delta/2}\sum_{i\ge 1}\mathsf{E}W_i^{1+\delta/2} \le S\ . \tag{5.72}$$

Then we have

$$\mathsf{P}\Big(u\sum_{i\ge 1}W_i\mathbf{1}_{\{uW_i\ge 1\}} \ge 4S\Big) \le LS^{-\delta/2}\ . \tag{5.73}$$

Proof. Replacing W_i by uW_i we can assume $u = 1$. We can assume $S \geq 1$, for there is nothing to prove otherwise. We set

$$U_i = W_i \mathbf{1}_{\{1 \leq W_i \leq S\}} - \mathsf{E}(W_i \mathbf{1}_{\{1 \leq W_i \leq S\}}) \,.$$

We have

$$\sum_{i \geq 1} \mathsf{E}(W_i \mathbf{1}_{\{1 \leq W_i \leq S\}}) \leq \sum_{i \geq 1} \mathsf{E} W_i^{1+\delta/2} \leq S,$$

so that

$$\mathsf{P}\Big(\sum_{i \geq 1} W_i \mathbf{1}_{\{1 \leq W_i \leq S\}} \geq 4S\Big) \leq \mathsf{P}\Big(\sum_{i \geq 1} U_i \geq 3S\Big)$$

and thus

$$\mathsf{P}\Big(\sum_{i \geq 1} W_i \mathbf{1}_{\{W_i \geq 1\}} \geq 4S\Big) \leq \mathsf{P}\Big(\sum_{i \geq 1} U_i \geq 3S\Big) + \sum_{i \geq 1} \mathsf{P}(W_i \geq S) \,.$$

Now

$$\sum_{i \geq 1} \mathsf{P}(W_i \geq S) \leq \frac{1}{S^{1+\delta/2}} \sum_{i \geq 1} \mathsf{E} W_i^{1+\delta/2} \leq S^{-\delta/2} \,.$$

Also, we note that $|U_i| \leq S$, so that since $1 - \delta/2 \geq 0$ we have

$$\mathsf{E} U_i^2 \leq S^{1-\delta/2} \mathsf{E}|U_i|^{1+\delta/2} \leq L S^{1-\delta/2} \mathsf{E} W_i^{1+\delta/2} \,, \tag{5.74}$$

using that $(a+b)^c \leq L(a^c + b^c)$ for $c \leq 2$. Thus we have

$$\mathsf{P}\Big(\sum_{i \geq 1} U_i \geq 3S\Big) \leq \frac{1}{9S^2} \mathsf{E}\Big(\sum_{i \geq 1} U_i\Big)^2 = \frac{1}{9S^2} \sum_{i \geq 1} \mathsf{E} U_i^2 \leq L S^{-\delta/2}$$

using (5.74) and (5.72). □

Lemma 5.3.12. *Consider a measure μ on $\mathbb{R}$ such that $\int (x^2 \wedge 1) \mathrm{d}\mu(x) < \infty$, and a r.v. X that satisfies*

$$\forall \alpha \in \mathbb{R} \,,\ \mathsf{E} \exp i\alpha X = \exp\Big(-\int_{\mathbb{R}} (1 - \cos \alpha x) \mathrm{d}\mu(x)\Big) \,. \tag{5.75}$$

Then we have

$$\int_{\mathbb{R}} \Big(\Big(\frac{x}{2\mathsf{E}|X|}\Big)^2 \wedge 1\Big) \mathrm{d}\mu(x) \leq L \,. \tag{5.76}$$

Proof. Since $\cos x \geq 1 - |x|$ we have

$$\mathsf{E} \cos \alpha X \geq 1 - \alpha \mathsf{E}|X| \geq \frac{1}{2}$$

if $0 \leq \alpha \leq 1/2\mathsf{E}|X|$. By (5.75) we have

$$\frac{1}{2} \leq \mathsf{E} \cos \alpha X = \exp\Big(-\int_{\mathbb{R}} (1-\cos \alpha x) \mathrm{d}\mu(x)\Big)$$

and hence

$$\int_{\mathbb{R}} (1-\cos \alpha x) \mathrm{d}\mu(x) \leq \log 2 \ .$$

Averaging the previous inequality over $0 \leq \alpha \leq 1/2\mathsf{E}|X|$, we get

$$\int_{\mathbb{R}} \left(1 - \frac{\sin(x/(2\mathsf{E}|X|))}{x/(2\mathsf{E}|X|)}\right) \mathrm{d}\mu(x) \leq \log 2$$

and the result since $y^2 \wedge 1 \leq L(1-(\sin y)/y)$. □

The proof of Theorem 5.3.2 will rely upon the application of Theorem 5.1.2 to suitable functionals, and we turn to the task of defining these. Consider an integer $\kappa \geq 6$, to be determined later, and $r = 2^{\kappa-4}$. Consider the maps

$$\varphi_j(s,t) = \varphi(s,t,r^j)$$

for $j \in \mathbb{Z}$. From Lemma 5.3.9, we note that

$$\varphi_{j+1}(s,t) \geq \frac{r^{1+\delta}}{K} \varphi_j(s,t) \ . \tag{5.77}$$

Given the sequences $(\tau_i)_{i\geq 1}$, $(Y_i)_{i\geq 1}$, to each $t \in T$ we can associate the sequence $\mathcal{S}(t) = (R_i Y_i(t))_{i\geq 1}$. To a subset A of T we can associate $\mathcal{S}(A) = \{\mathcal{S}(t)\,;\, t \in A\}$. This is a random set of sequences. Since the process (5.62) is distributed like (X_t), we have the identity

$$\mathsf{E} \sup_{t \in T} X_t = \mathsf{E} b(\mathcal{S}(T)) \ , \tag{5.78}$$

where the notation b is defined in (4.1). We recall the chopping maps Ψ_c of Chapter 4, and for a set of sequences U, we use the notation $b_j(U) = b(\Psi_{r^{-j}}(U))$.

Let us fix once and for all a number α such that $\mathsf{P}(\Omega_0) \geq 3/4$, where

$$\Omega_0 = \left\{\frac{1}{\alpha} \leq \alpha_- \leq \alpha_+ \leq \alpha\right\} \ . \tag{5.79}$$

A *random subset* Z of T is a subset of T that depends on the sequences $(\tau_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ and is such that for each t the set $\{t \in Z\}$ is measurable. It might help to think to a random subset of T as a (small) set of badly behaved points.

Consider two decreasing sequences $c(n), d(n) > 0$, tending to 0, that will be determined later. Given a probability measure μ on T, we first define the functional $F_{n,j}(\mu)$ as the supremum of the numbers c that have the property that there exists a random subset Z of T with $\mathsf{E}\mu(Z) \leq d(n)$, with the following property:

$$\forall U \subset T\,,\ U \cap Z = \emptyset\,,\ \mu(U) \geq c(n) \Rightarrow b_j(\mathcal{S}(U)) \geq c\mathbf{1}_{\Omega_0}\ . \tag{5.80}$$

In words, when Ω_0 occurs, for any subset U of T that is not too small and contains no badly behaved points, we have $b_j(S(U)) \geq c$.

Given a subset A of T, we then define

$$F_{n,j}(A) = \sup\{F_{n,j}(\mu)\,;\ \mu(A) = 1\}\ .$$

Since $c(n+1) \leq c(n)$ and $d(n+1) \leq d(n)$, it is obvious that $F_{n+1,j} \leq F_{n,j}$; and it follows from Proposition 4.3.7 that $F_{n,j+1} \leq F_{n,j}$.

Lemma 5.3.13. *If $d(n) \leq 1/8$ and $c(n) \leq 1/2$, then $F_{n,j}(T) \leq 2\mathsf{E}\sup_{t\in T} X_t$ for all $j \in \mathbb{Z}$.*

Proof. Consider a probability measure μ on T with $F_{n,j}(\mu) > c$ and Z a random subset of T with $\mathsf{E}\mu(Z) \leq d(n)$, that satisfies (5.80). If $d(n) \leq 1/8$, then $\mathsf{P}(\mu(Z) \leq 1/2) \geq 3/4$ and thus $\mathsf{P}(\Omega_1) \geq 1/2$ where $\Omega_1 = \Omega_0 \cap \{\mu(Z) \leq 1/2\}$. Since

$$\mu(T\backslash Z) \geq \frac{1}{2}\mathbf{1}_{\{\mu(Z)\leq 1/2\}}$$

we see that by (5.80) we have

$$b_j(\mathcal{S}(T\backslash Z)) \geq c\mathbf{1}_{\Omega_1}\ ,$$

so that $b(\mathcal{S}(T)) \geq c\mathbf{1}_{\Omega_1}$. Since $b(\mathcal{S}(T)) \geq 0$, taking expectation we get $c \leq 2\mathsf{E}b(\mathcal{S}(T))$, and the conclusion by (5.78). □

Lemma 5.3.14. *Under condition $H(C_0,\delta)$ and if $r \geq K$, there exists $j = j_0 \in \mathbb{Z}$ such that*

$$\forall s\,,\ t \in T\,,\ \varphi_{j-1}(s,t) \leq 1 \tag{5.81}$$

$$r^{-j} \leq 4r\mathsf{E}\sup_{t\in T} X_t\ . \tag{5.82}$$

Proof. Consider $s\,,t \in T$. By Lemma 1.2.8 we have

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

Using (5.53) we see that the r.v. $X = X_s - X_t$ satisfies (5.75) where μ is the image of ν under the map $\beta \mapsto \beta(s) - \beta(t)$. From (5.76) we get that $\varphi(s,t,1/(4S)) \leq L$. Consider the largest integer j such that $r^{-j} \geq 2S$, so that $\varphi_j(s,t) \leq L$ and $r^{-j} \leq 2rS$. It follows from (5.77) that if $r \geq K$ we have $\varphi_{j-1}(s,t) \leq 1$. □

Proof of Theorem 5.3.2. The main step is to prove the existence of a number K_1, depending on C_0 and δ only, and of sequences $(d(n))_{n\geq 0}$ and $(c(n))_{n\geq 0}$, also depending on C_0 and δ only, tending to 0 as $n \to \infty$, such that if $n \geq K_1$ and $\kappa \geq K_1$, the functionals $K_1 F_{n,j}$ satisfy the growth condition of Definition 5.1.1.

Once this is done, we complete the argument as follows. We choose $n_0 \geq K_1$ large enough that $d(n_0) \leq 1/8$ and $c(n_0) \leq 1/2$, and depending only on C_0 and δ. Consider the value of j_0 constructed in Lemma 5.3.14, so that $2^{n_0} r^{-j_0} \leq K \mathsf{E} \sup_{t\in T} X_t$ by (5.82) and $F_{n_0,j_0}(T) \leq L \mathsf{E} \sup_{t\in T} X_t$ by Lemma 5.3.13. We then apply Theorem 5.1.2 with these values of j_0 and n_0. We observe that for $n \geq n_0$ (5.59) follows from (5.10) and (5.58). Finally for $A \in \mathcal{A}_n, n \leq n_0$ we define $j(A) = j_0$, and (5.81) shows that (5.59) remains true for $n \leq n_0$. This completes the proof of Theorem 5.3.2.

We now turn to the proof of the growth condition. Let us assume that we are given points $(t_\ell)_{\ell\leq m}$, as in (5.3), where $m = N_n$, and consider sets $H_\ell \subset B_{j+1}(t_\ell, 2^{n+\kappa})$.

The basic idea is that we want to apply (4.11) to the sets $H'_\ell = \Psi_{r^{-j}}(\mathcal{S}(H_\ell))$. We however face two problems.

(a) It is not always true that the points $u_\ell := \Psi_{r^{-j}}(\mathcal{S}(t_\ell))$ for $\ell \leq m$ are far from each other.

(b) It is not true that all the points of H'_ℓ are close to u_ℓ.

The route around these problems is to prove that

(c) it is very likely that sufficiently many of the points u_ℓ are far from each other,

(d) for most of the points t of H_ℓ it is true that $\Psi_{r^{-j}}(\mathcal{S}(t))$ is close to u_ℓ.

The exceptional points are eliminated by putting them into a suitable random set. This is the main purpose of introducing random sets.

Consider $c < \min_{\ell\leq m} F_{n+1,j+1}(H_\ell)$. Since $F_{n+1,j+1} \leq F_{n+1,j}$, we can find for each $\ell \leq m$ a probability measure μ_ℓ with $\mu_\ell(H_\ell) = 1$, a random subset Z_ℓ of H_ℓ with $\mathsf{E}(\mu_\ell(Z_\ell)) \leq d(n+1)$, with the property that

$$U \subset H_\ell \backslash Z_\ell \, , \, \mu_\ell(U) \geq c(n+1) \Rightarrow b_j(\mathcal{S}(U)) \geq c\mathbf{1}_{\Omega_0} \, . \tag{5.83}$$

Consider an integer p that will be determined later, and, assuming $n \geq p$, set

$$\mu = \frac{1}{N_{n-p}} \sum_{\ell\leq N_{n-p}} \mu_\ell \, ; \, Z' = \bigcup_{\ell\leq N_{n-p}} Z_\ell \, .$$

The sets H_ℓ are disjoint, and $\mu_\ell(H_\ell) = 1$. This implies that we have $\mu(Z') = N_{n-p}^{-1} \sum_{\ell\leq N_{n-p}} \mu_\ell(Z_\ell)$ and thus $\mathsf{E}\mu(Z') \leq d(n+1)$.

We turn to the realization of (c) above. Since $\varphi_j(t_\ell, t_{\ell'}) \geq 2^n$ for $\ell \neq \ell'$, we see from (5.68), used for $u = r^j$, that

$$\mathsf{P}_\tau\Big(\sum_{i\geq 1} \big(R_i(Y_i(t_\ell) - Y_i(t_{\ell'}))^2\big) \wedge r^{-2j} \leq \frac{2^n r^{-2j}}{8\alpha_+}\Big) \leq \exp\Big(-\frac{2^n}{8\alpha_+}\Big) \, .$$

Recalling (5.79), we have

$$\mathsf{P}\Big(\Omega_0\cap\{\sum_{i\geq 1}\big(R_i(Y_i(t_\ell)-Y_i(t_{\ell'}))^2\big)\wedge r^{-2j}\leq\frac{2^n r^{-2j}}{8\alpha}\}\Big)\leq\exp\Big(-\frac{2^n}{8\alpha}\Big)\,. \quad (5.84)$$

It follows from (4.33) that

$$\sum_{i\geq 1}\big(R_i(Y_i(t_\ell)-Y_i(t_{\ell'}))^2\big)\wedge r^{-2j}\leq 4\|\Psi_{r^{-j}}(\mathcal{S}(t_\ell))-\Psi_{r^{-j}}(\mathcal{S}(t_{\ell'}))\|_2^2\,.$$

We choose p once and for all such that $2^{-p}\leq 1/(32\alpha)$. Then we get from (5.84) that

$$\begin{aligned}\mathsf{P}(\Omega_0\cap\{\|\Psi_{r^{-j}}(\mathcal{S}(t_\ell))-\Psi_{r^{-j}}(\mathcal{S}(t_{\ell'}))\|_2\leq 2^{(n-p)/2}r^{-j}\})&\\ \leq\exp(-2^{n-p+2})\,.&\end{aligned}$$

Consider the event Ω_1 defined as

$$\Omega_0\cap\{\exists\ell<\ell'\leq N_{n-p}\,;\,\|\Psi_{r^{-j}}(\mathcal{S}(t_\ell))-\Psi_{r^{-j}}(\mathcal{S}(t_{\ell'}))\|_2<2^{(n-p)/2}r^{-j}\}\,.$$

Then

$$\mathsf{P}(\Omega_1)\leq N_{n-p}^2\exp(-2^{n-p+2})\leq\exp(-2^{n-p+1})\,.$$

Consider the random subset Z'' of T given by $Z''=T\mathbf{1}_{\Omega_1}$. Thus $\mathsf{E}\mu(Z'')\leq\exp(-2^{n-p+1})$.

We turn to the realization of (d) above. We observe that for $t\in H_\ell\subset B_{j+1}(t_\ell,2^{n+\kappa})$, we have $\varphi_{j+1}(t,t_\ell)\leq 2^{n+\kappa}$. By (5.77), and since $r=2^{\kappa-4}$, we have $\varphi_j(t,t_\ell)\leq K2^n r^{-\delta}$. Thus by (5.69) we have

$$\mathsf{P}\Big(\mathbf{1}_{\Omega_0}\sum_{i\geq 1}R_i(Y_i(t)-Y_i(t_\ell))^2\wedge r^{-2j}\geq K2^n r^{-\delta}r^{-2j}\Big)\leq\exp(-2^n r^{-\delta})\,. \quad (5.85)$$

Using (5.71) and (5.73) with $S=K2^n r^{-\delta}$ and $u=r^j$ we get

$$\begin{aligned}&\mathsf{P}\Big(\mathbf{1}_{\Omega_0}\sum_{i\geq 1}R_i|Y_i(t)-Y_i(t_\ell)|\mathbf{1}_{\{|Y_i(t)-Y_i(t_\ell)|\geq r^{-j}\}}\geq K2^n r^{-\delta}r^{-j}\Big) \quad (5.86)\\ &\leq K2^{-n\delta/2}r^{\delta^2/2}\,.\end{aligned}$$

Let us say that a point $t\in\bigcup_{\ell\leq m}H_\ell$ is *regular* if, when ℓ is such that $t\in H_\ell$, we have

$$\sum_{i\geq 1}R_i(Y_i(t)-Y_i(t_\ell))^2\wedge r^{-2j}\leq K_0 2^n r^{-\delta}r^{-2j} \quad (5.87)$$

$$\sum_{i\geq 1}R_i|Y_i(t)-Y_i(t_\ell)|\mathbf{1}_{\{|Y_i(t)-Y_i(t_\ell)|\geq r^{-j}\}}\leq K_0 2^n r^{-\delta}r^{-j}\,. \quad (5.88)$$

Thus, (5.85) and (5.86) imply that if K_0 is large enough, the probability that Ω_0 occurs and t is not regular is at most $\exp(-2^n r^{-\delta})+K2^{-n\delta/2}r^{\delta^2/2}$.

Consider the random set Z''' defined as

$$Z''' = \Big\{t \in \bigcup_{\ell \le m} H_\ell\,;\, t \text{ is not regular}\Big\}\,,$$

if Ω_0 occurs and $Z''' = \emptyset$ otherwise. Thus $\mathsf{E}\mu(Z''') \le \exp(-2^n r^{-\delta}) + K2^{-n\delta/2} r^{\delta^2/2}$.

We consider the random set $Z = Z' \cup Z'' \cup Z'''$, so that

$$\mathsf{E}\mu(Z) \le d(n+1) + \exp(-2^{n-p+1}) + \exp(-2^n r^{-\delta}) + K2^{-n\delta/2} r^{\delta^2/2}\,. \quad (5.89)$$

This finishes the construction of the appropriate random set, and we turn to the question of bounding $F_{n,j}(\bigcup_{\ell\le m} H_\ell)$ from below. Consider a set $U \subset \bigcup_{\ell \le m} H_\ell$, with $U \cap Z = \emptyset$, and assume that

$$\mu(U) \ge c(n+1) + \frac{1}{N_{n-p-1}}\,. \quad (5.90)$$

Using the definition of μ we see that

$$\mu(U) \le \frac{1}{N_{n-p}} \operatorname{card}\{\ell \le N_{n-p}\,;\, \mu_\ell(U) \ge c(n+1)\} + c(n+1)\,,$$

so that $\operatorname{card} I \ge N_{n-p}/N_{n-p-1} \ge N_{n-p-1}$, where

$$I = \{\ell \le N_{n-p}\,;\, \mu_\ell(U) \ge c(n+1)\}\,.$$

Let $U_\ell = U \cap H_\ell$, so that $\mu_\ell(U) = \mu_\ell(U_\ell)$, and since $U_\ell \cap Z_\ell \subset U \cap Z = \emptyset$, it follows from (5.83) that we have

$$\forall \ell \in I\,,\ b_j(\mathcal{S}(U \cap H_\ell)) \ge c\mathbf{1}_{\Omega_0}\,.$$

We define $H'_\ell = \Psi_{r^{-j}}(\mathcal{S}(U_\ell))$, so that $b(H'_\ell) = b_j(\mathcal{S}(U_\ell))$. Defining $u_\ell = \Psi_{r^{-j}}(\mathcal{S}(t_\ell))$, since U consists only of regular points, (5.87), (5.88) and (4.34) show that

$$H'_\ell \subset B(u_\ell, Kr^{-\delta/2} 2^{n/2} r^{-j})\,. \quad (5.91)$$

Moreover, since $U \ne \emptyset$, $U \cap Z'' = \emptyset$, and since $Z'' = T\mathbf{1}_{\Omega_1}$, the event Ω_1 does not occur, i.e.

$$\forall \ell < \ell' \le N_{n-p}\,,\ \|u_\ell - u_{\ell'}\|_2 \ge 2^{(n-p)/2} r^{-j}\,.$$

We appeal to (4.11) with $m = N_{n-p-1}$ and $a = r^{-j} 2^{(n-p)/2}$. We then see that if we choose $r = 2^{\kappa-4}$ where κ is the smallest integer such that $Kr^{-\delta/2} \le 2^{-p/2}/L_0$, where L_0 is the constant of (4.11), then

$$b\Big(\bigcup_{\ell \in I} H'_\ell\Big) \ge \big(c + \frac{1}{K_2} 2^n r^{-j}\big)\mathbf{1}_{\Omega_0}\,.$$

Since

$$b\Big(\bigcup_{\ell\in I} H'_\ell\Big) = b\Big(\Psi_{r^{-j}}\big(\mathcal{S}(\bigcup_{\ell\in I} U_\ell)\big)\Big) \le b\big(\Psi_{r^{-j}}(\mathcal{S}(U)\big) = b_j(\mathcal{S}(U))\,,$$

this implies that

$$b_j(\mathcal{S}(U)) \ge b\Big(\bigcup_{\ell\in I} H'_\ell\Big) \ge (c + \frac{1}{K_2}2^n r^{-j})\mathbf{1}_{\Omega_0}\,. \tag{5.92}$$

From (5.90), we see that it is appropriate to define

$$c(n) = \sum_{q\ge n} \frac{1}{N_{q-p-1}}$$

and from (5.89)

$$d(n) = \sum_{q\ge n}\Big(\exp(-2^{q-p+1}) + \exp(-2^q r^{-\delta}) + K2^{-q\delta/2} r^{\delta^2/2}\Big)\,.$$

(Where the value of r is the one previously fixed.)

With these choices (5.89) implies that $\mathsf{E}\mu(Z) \le d(n)$, and (5.90) means that $\mu(U) \ge c(n)$. We have proved that these conditions, together with $U \subset \bigcup_{\ell\le m} H_\ell$ and $Z \cap U = \emptyset$, imply (5.92). By definition of the functionals $F_{n,j}$, this implies that

$$F_{n,j}\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge c + \frac{1}{K_2}2^n r^{-j}\,,$$

so that (5.92) implies that (5.4) holds for the functionals $K_2F_{n,j}$, and this completes the proof of Theorem 5.3.2. □

5.4 The Decomposition Theorem for Infinitely Divisible Processes

This section is closely connected to the previous one. The reader needs in particular to keep in mind Rosinski's representation (5.62) of an infinitely divisible process of Lévy measure ν. We consider a Borel subset Ω of $\mathbb{R}^T$ with $m(\Omega^c) = 0$ and we assume without loss of generality that Y_i is valued in Ω. On T we consider the distance $d_\infty(s,t)$ given by $d_\infty(s,t) = \sup_{\beta\in\Omega} |\beta(s) - \beta(t)|$, and the distance $d_2(s,t)$ given by $d_2^2(s,t) = \int_\Omega (\beta(s) - \beta(t))^2 \mathrm{d}\nu(\beta)$.

Theorem 5.4.1. *We have*

$$\mathsf{E}\sup_{t\in T} X_t \le L\big(\gamma_2(T,d_2) + \gamma_1(T,d_\infty)\big)\,. \tag{5.93}$$

Proof. Let us denote by E_τ and P_τ expectation and probability given the sequence $(\tau_i)_{i\geq 1}$. We will prove that

$$\mathsf{E}_\tau \sup_{t\in T} X_t \leq L\Big(\frac{1}{\sqrt{\alpha_-}}\gamma_2(T,d_2) + \gamma_1(T,d_\infty)\Big)\,. \tag{5.94}$$

Writing $\mathsf{P}(\alpha_- \leq \epsilon) \leq \sum_{i\geq 1}\mathsf{P}(\tau_i \leq i\epsilon)$, simple estimates show that $\mathsf{P}(\alpha_- \leq \epsilon) \leq L\epsilon$, so that $\mathsf{E}(1/\sqrt{\alpha_-}) < \infty$ and taking expectation in (5.94) finishes the proof.

Consider $s\,,\,t \in T$, and $G_i = \epsilon_i R_i(Y_i(s) - Y_i(t))$. Thus $|G_i| \leq d_\infty(s,t)$, and, by the right-hand side of (5.67), used for $h(\beta) = (\beta(s)-\beta(t))^2$, we have $\sum_{i\geq 1}\mathsf{E}_\tau G_i^2 \leq d_2(s,t)^2/\alpha_-$. Thus (5.94) follows from Theorem 1.2.7 since, by Bernstein's inequality (Lemma 2.7.1), we have

$$\mathsf{P}_\tau\Big(\big|\sum_{i\geq 1} G_i\big| \geq v\Big) \leq \exp\Big(-\frac{1}{L}\min\big(\frac{v^2\alpha_-}{d_2(s,t)^2}\,,\,\frac{v}{d_\infty(s,t)}\big)\Big)\,.$$

□

We have thus described a class of processes that we can certify are bounded, and we turn to the description of another class of processes that we can also certify are bounded, but for a very different reason.

Given a finite set T, we say that the process $(X_t)_{t\in T}$ is *positive infinitely divisible* if there exists a positive measure ν on $(\mathbb{R}^+)^T$, such that

$$\int (\beta(t)\wedge 1)\mathrm{d}\nu(\beta) < \infty$$

for each t in T, and that for each family $(\alpha_t)_{t\in T}$ of real numbers we have

$$\mathsf{E}\exp i\sum_{t\in T}\alpha_t X_t = \exp\Big(-\int\big(1-\exp(i\sum_{t\in T}\alpha_t\beta(t))\big)\mathrm{d}\nu(\beta)\Big)\,.$$

We will call ν the Lévy measure of the process. While by "infinitely divisible process" we understand that the process is symmetric, a *positive* infinitely divisible process is certainly not symmetric. It is not obvious that this process is positive; but another version of Rosinski's representation shows that (with the notation of Section 5.3) the process

$$\sum_{i\geq 1} R_i Y_i(t) \tag{5.95}$$

has the same law as $(X_t)_{t\in T}$. This is also proved in [35]. (The representation (5.95) will also be called the Rosinski representation of the positive infinitely divisible process.) The important feature here is that all terms in (5.95) are non-negative. There is no cancelation in this sum.

Consider now a (symmetric) infinitely divisible process $(X_t)_{t\in T}$ with Lévy measure ν and assume that

$$\forall t \in T\,,\ \int (|\beta(t)| \wedge 1)\mathrm{d}\nu(\beta) < \infty\,.$$

Consider the positive measure ν' on $(\mathbb{R}^+)^T$ that is the image of ν under the map $\beta \mapsto |\beta|$, where $|\beta|(t) = |\beta(t)|$. Then

$$\forall t \in T\,,\ \int (\beta(t) \wedge 1)\mathrm{d}\nu'(\beta) < \infty$$

so ν' is the Lévy measure of a positive infinitely divisible process that we denote by $(|X|_t)$. If ν is the image of $\lambda \otimes m$ under the map $(x, \beta) \mapsto R(x, \beta)\beta$, then, since $R(x, \beta) \geq 0$, ν' is the image of $\lambda \otimes m$ under the map

$$(x, \beta) \mapsto |R(x, \beta)\beta| = R(x, \beta)|\beta|\,.$$

Thus if $\sum_{i\geq 1} \epsilon_i R_i Y_i$ is a Rosinski representation of the process (X_t), then $\sum_{i\geq 1} R_i |Y_i|$ is a Rosinski representation of the process $(|X|_t)$. Hence

$$\mathsf{E} \sup_{t\in T} X_t = \mathsf{E} \sup_{t\in T} \sum_{i\geq 1} \epsilon_i R_i Y_i(t) \leq \mathsf{E} \sup_{t\in T} \sum_{i\geq 1} R_i |Y_i(t)| = \mathsf{E} \sup_{t\in T} |X|_t\,.$$

This shows that to control $\mathsf{E} \sup_{t\in T} X_t$ it suffices to control $\mathsf{E} \sup_{t\in T} |X|_t$, a control that does not involve any cancelation.

We now come to the main result of this chapter. In the following definition as usual ν is the image of $\lambda \otimes m$ under the map $(x, \beta) \mapsto R(x, \beta)\beta$.

Definition 5.4.2. *We say that an infinitely divisible process is S-certified if $\gamma_1(T, d_\infty) \leq S$ and $\gamma_2(T, d_2) \leq S$, where, for a certain set $\Omega \subset \mathbb{R}^T$ with $m(\Omega^c) = 0$, we have*

$$d_\infty(s, t) = \sup_{\beta\in\Omega} |\beta(s) - \beta(t)|\,,$$

and

$$d_2(s, t) = \Big(\int_\Omega (\beta(s) - \beta(t))^2 \mathrm{d}\nu(\beta)\Big)^{1/2}\,.$$

Theorem 5.4.3. *Consider an infinitely divisible process $(X_t)_{t\in T}$, and assume that condition $H(C_0, \delta)$ of Definition 5.3.1 holds. Let $S = \mathsf{E} \sup_{t\in T} X_t$. Then we can write in distribution*

$$X_t = X'_t + X''_t$$

where both processes $(X'_t)_{t\in T}$ and $(X''_t)_{t\in T}$ are infinitely divisible with the following properties: (X'_t) is KS-certified, and $\mathsf{E} \sup_{t\in T} |X''|_t \leq KS$.

In other words, we know two ways to control $\mathsf{E}\sup_{t\in T} X_t$. One way is that the process is S-certified. The other way is that we already control $\mathsf{E}\sup_{t\in T} |X|_t$. Under condition $H(C_0,\delta)$ there is *no* other method: every situation is a combination of these.

To prove this theorem, it is convenient to adopt a different point of view. This will also bring to light the fact that the material of this section is closely connected to the material of Section 2.7. To make this more apparent, rather than considering $\beta \in \Omega \subset \mathbb{R}^T$ as a function of $t \in T$, we will think of $t \in T$ as a function of β, by the formula $t(\beta) = \beta(t)$. Since ν is a Lévy measure, we have

$$\forall t \in T\,, \int_\Omega (t(\beta)^2 \wedge 1)\mathrm{d}\nu(\beta) < \infty\,. \tag{5.96}$$

Conversely, assume that we are given a (σ-finite) positive measure space (Ω,ν) and a (countable) set T of measurable functions on Ω such that (5.96) holds. Consider a probability measure m such that $\nu \ll m$ and a function g such that $\nu = gm$. Consider an i.i.d. sequence Y_i distributed like m, and set $R_i = \mathbf{1}_{[0,g(Y_i)]}(\tau_i)$. Then Rosinski's representation

$$X_t = \sum_{i\ge 1} \epsilon_i R_i t(Y_i)$$

defines an infinitely divisible process $(X_t)_{t\in T}$. Its Lévy measure $\bar\nu$ is the image of ν under the map $\omega \mapsto (t(\omega))_{t\in T}$. If, moreover,

$$\forall t \in T\,, \int_\Omega (|t(\beta)| \wedge 1)\mathrm{d}\nu(\beta) < \infty\,, \tag{5.97}$$

we can define a positive infinitely divisible process $(|X|_t)_{t\in T}$ by

$$|X|_t = \sum_{i\ge 1} R_i |t(Y_i)|\,.$$

The distances d_2 and d_∞ of Theorem 5.4.1 are simply the distances induced by the norms of $L^2(\nu)$ and $L^\infty(\nu)$ respectively.

Let us repeat: for the purpose of studying boundedness, an infinitely divisible process is essentially *a class of functions on a measure space.*

Theorem 5.4.4. *We have*

$$\mathsf{E}\sup_{t\in T} |X|_t \le L\Big(\mathsf{E}\sup_{t\in T} |X_t| + \sup_{t\in T} \int |t(\beta)|\mathrm{d}\nu(\beta)\Big)\,.$$

Proof. As explained, if a Rosinski representation of X_t is $\sum_{i\ge 1} \epsilon_i R_i t(Y_i)$, a Rosinski representation of $|X|_t$ is $\sum_{i\ge 1} R_i |t(Y_i)|$. We will need to use (5.67) and a minor technical difficulty arises because $1/\alpha_-$ is not integrable. This is why below we consider the first term separately. We write

$$\mathsf{E}\sup_{t\in T}|X|_t = \mathsf{E}\sup_{t\in T}\sum_{i\geq 1} R_i|t(Y_i)| \leq \mathrm{I}+\mathrm{II} \tag{5.98}$$

$$\mathrm{I} = \mathsf{E}\sup_{t\in T} R_1|t(Y_1)| \quad ; \quad \mathrm{II} = \mathsf{E}\sup_{t\in T}\sum_{i\geq 2} R_i|t(Y_i)| \,. \tag{5.99}$$

We have

$$\mathsf{E}\sup_{t\in T} R_1|t(Y_1)| = \mathsf{E}\sup_{t\in T}|\epsilon_1 R_1 t(Y_1)| \leq \mathsf{E}\sup_{t\in T}\Big|\sum_{i\geq 1}\epsilon_i R_i t(Y_i)\Big| = \mathsf{E}\sup_{t\in T}|X_t| \,. \tag{5.100}$$

To control the term II, we denote by E_τ expectation at $(\tau_i)_{i\geq 1}$ given. Given τ_i, the pairs of r.v. (R_i, Y_i) are independent. We appeal to (2.153) to get

$$\begin{aligned}\mathsf{E}_\tau\sup_{t\in T}\sum_{i\geq 2} R_i|t(Y_i)| &\leq \sup_{t\in T}\sum_{i\geq 2}\mathsf{E}_\tau R_i|t(Y_i)| + 2\mathsf{E}_\tau\sup_{t\in T}\Big|\sum_{i\geq 2}\epsilon_i R_i t(Y_i)\Big| \\ &\leq \sup_{t\in T}\sum_{i\geq 2}\mathsf{E}_\tau R_i|t(Y_i)| + 2\mathsf{E}_\tau\sup_{t\in T}|X_t| \,. \end{aligned}\tag{5.101}$$

An obvious extension of the right-hand side inequality of (5.67) gives

$$\sum_{i\geq 2}\mathsf{E}_\tau R_i|t(Y_i)| \leq \frac{1}{\alpha'}\int |t(\beta)|\mathrm{d}\nu(\beta)\,, \tag{5.102}$$

where $\alpha' = \inf_{i\geq 2}\tau_i/i$. Writing

$$\mathsf{P}(\alpha' \leq \epsilon) \leq \sum_{i\geq 2}\mathsf{P}(\tau_i \leq \epsilon i)\,,$$

one sees through simple estimates that $\mathsf{E}(1/\alpha') < \infty$. We plug (5.102) in (5.101), we take expectation, and we combine with (5.98), (5.99) and (5.100) to conclude the proof. □

Proof of Theorem 5.4.3. We still think of T as a set of functions on (Ω, m). Without loss of generality we can assume that $0 \in T$, so that $\mathsf{E}\sup_{t\in T}|X_t| \leq 2S$ by Lemma 1.2.8. The main argument consists in decomposing $T \subset T_1 + T_4$, where $\gamma_1(T_1, d_\infty) \leq KS$, $\gamma_2(T_1, d_2) \leq KS$, $0 \in T_1$ and $\sup_{t\in T_4}\int |t(\beta)|d\nu(\beta) \leq KS$. Once this is done, it follows from Theorem 5.4.1 and Lemma 1.2.8 that $\mathsf{E}\sup_{t\in T_1}|X_t| \leq KS$, and, since we can assume that $T_4 \subset T - T_1$, that $\mathsf{E}\sup_{t\in T_4}|X_t| \leq KS$. It then follows from Theorem 5.4.4 that $\mathsf{E}\sup_{t\in T_4}|X|_t \leq KS$. Finally, the decomposition $X_t = X'_t + X''_t$ is obtained by fixing a decomposition $t = t_1 + t_2$ for each t in T with $t_1 \in T_1$, $t_2 \in T_4$, and setting $X'_t = X_{t_1}$, $X''_t = X_{t_2}$.

To decompose T we first use Theorem 5.3.2 to find a number r (depending only on C_0 and δ), an admissible sequence $(\mathcal{A}_n)$ of T and for $A \in \mathcal{A}_n$ an integer $j(A) \in \mathbb{Z}$ that satisfies (5.8) and

$$\forall s\,,\, t \in A\,,\, \varphi(s,t,r^{j(A)-1}) \le 2^{n+1} \tag{5.103}$$

$$\forall t \in T\,,\, \sum_{n\ge 0} 2^n r^{-j(A_n(t))} \le KS\,. \tag{5.104}$$

We then use Theorem 2.6.3, choosing $\mu = \nu\,, V = r$ and $\delta(A) = 2^{(n+1)/2}r^{-j(A)+1}$, so that (5.103) implies (2.101). Condition (2.99) follows from (5.8) and Condition (2.100) is automatically satisfied since $n \ge n'$.

We consider the decomposition $T \subset T_1 + T_2 + T_3$ provided by Theorem 2.6.3. We set $T_4 = T_2 + T_3$, so that $T \subset T_1 + T_4$. By (2.102), (2.103) and (5.104) we have $\gamma_2(T_1, d_2) \le KS$ and $\gamma_1(T_1, d_\infty) \le KS$. By (2.104), used for $p = 1$, we have $\|t\|_1 \le KS$ for $t \in T_2$, as is obvious since

$$V^{2j(A_{n+1}(t))-j(A_n(t))}\delta^2(A_{n+1}(t)) = 2^{n+1}r^{-j(A_n(t))+2}\,.$$

Thus all we have to do is to show that $\|t\|_1 \le KS$ for $t \in T_3$, and, from (2.105), we see that it suffices to show that $\|t\|_1 \le KS$ for $t = |s|\mathbf{1}_{\{2|s|\ge r^{-j(T)}\}}$, $s \in T$. Now, since $0 \in T$, using (5.103) for $n = 0$ and $A = T$ we have

$$\begin{aligned}\nu(\{\beta\,;\,|s(\beta)| \ge r^{-j(T)}/2\}) &\le 4r^2 \int \big((sr^{j(T)-1})^2 \wedge 1\big)\mathrm{d}\nu \\ &\le 4r^2\varphi(s,0,r^{j(T)-1}) \le 8r^2\,.\end{aligned}$$

It follows from condition $H(C_0,\delta)$ and integration by parts that $\|t\|_1 \le Kr^{-j(T)}$, and since $r^{-j(T)} \le KS$ by (5.104), that $\|t\|_1 \le KS$. The proof is complete. □

We conclude this section by a "bracketing theorem" in the spirit of Ossiander's Theorem (Theorem 2.7.10). In this theorem, we still think of T as a set of measurable functions on (Ω, m).

Theorem 5.4.5. *Consider an admissible sequence $(\mathcal{A}_n)$ of T, and for $A \in \mathcal{A}_n$ consider h_A given by $h_A(\omega) = \sup_{s,t\in A}|t(\omega) - s(\omega)|$. Assume that for $A \in \mathcal{A}_n$ we are given $j(A) \in \mathbb{Z}$ satisfying $A \subset B \Rightarrow j(A) \ge j(B)$. Assume that for some numbers $r \ge 2$ and $S > 0$ we have*

$$\forall A \in \mathcal{A}_n\,,\, \int \big(r^{2j(A)}h_A^2 \wedge 1\big)\mathrm{d}\nu \le 2^n$$

$$\forall t \in T\,,\, \sum_{n\ge 0} 2^n r^{-j(A_n(t))} \le S\,,$$

and that $h_T < r^{-j(T)}$. Then we have $\mathsf{E}\sup_{t\in T}|X_t| \le LS$.

This result follows from Theorem 2.7.12 (with $\delta(A) = 2^n r^{-j(A)}$) just as Theorem 2.7.10, using that $T_3 = \{0\}$ since $h_T < r^{-j(T)}$. The details are left to the reader.

5.5 Further Thoughts

The content of Theorem 5.4.3 can be viewed as the statement that (at least under the condition $H(C_0, \delta)$), to understand boundedness of infinitely divisible processes, it suffices to understand the boundedness of *positive* infinitely divisible processes. As seen is the previous section, this basically amounts to the following question. Given a probability space (Ω, m), a non-negative function g on Ω, the positive measure $\nu = gm$, and a class $\mathcal{F}$ of measurable non-negative functions on Ω such that

$$\forall f \in \mathcal{F}\,, \int (f \wedge 1) \mathrm{d}\nu < \infty\,,$$

how can we control

$$\sup_{f \in \mathcal{F}} \sum_{i \geq 1} R_i f(Y_i)\,, \tag{5.105}$$

where $(Y_i)_{i \geq 1}$ are i.i.d. of law m and where $R_i = \mathbf{1}_{\{g(Y_i) \leq \tau_i\}}$? More precisely, how could we compute this quantity from the "geometry" of $\mathcal{F}$? This is a rather difficult question, in particular because it is not clear in what direction one should look. The author has made several attempts on problems of a somewhat similar nature, and in particular [63], [44] and [68] (see also [27]), by considering combinatorial quantities.

We would like to explain a new direction of investigation that became apparent during the writing of this book. Let us first revisit our results on Gaussian processes. Theorem 2.1.1 gives a complete description of the quantity $\mathsf{E} \sup_{t \in T} X_t$ as a function of the geometry of the metric space (T, d). This is the kind of result one wishes to prove, as it provides a full understanding of the situation. But is there a way to gather some understanding even if we do not yet have the hope to fully understand the situation? Let us look back at Theorem 2.1.8. A consequence of this result is that for any Gaussian process we can find a jointly Gaussian sequence (u_k) such that

$$\Big\{\sup_{t \in T} |X_t| \geq K \mathsf{E} \sup_{t \in T} |X_t|\Big\} \subset \bigcup_{k \geq 1} \{u_k \geq 1\} \tag{5.106}$$

and moreover

$$\sum_{k \geq 1} \mathsf{P}(u_k \geq 1) \leq \frac{1}{2}\,.$$

In words, we have found a *concrete witness* that the set on the left-hand side of (5.106) has a probability at most 1/2. This is a non-trivial information, even though it is not as good as the information provided by Theorem 2.1.1. (Let us observe in particular that this information is rather easy to deduce from Theorem 2.1.1, but that it does not seem easy to go the other way around.)

Going back to the quantity (5.105), if M denotes a median of this quantity, one can hope that, for some universal constant K, there always exists a "concrete witness" that the set

$$\Big\{\sup_{f\in\mathcal{F}}\sum_{i\geq 1} R_i f(Y_i) \geq KM\Big\} \tag{5.107}$$

has a probability at most 1/2. What should these concrete witness be? We will not discuss this question, because it seems more fruitful at this stage to turn towards a simpler problem of the same nature. We consider a number $0 < \delta < 1$ and we consider independent r.v. $(\delta_i)_{i\geq 0}$ such that $\mathsf{P}(\delta_i = 1) = \delta$ and $\mathsf{P}(\delta_i = 0) = 1-\delta$. (These random variables are often called selectors, and have been used in a variety of situations to select at random a "small proportion" of a given set.) Consider a (finite if one wishes) class $\mathcal{F}$ of functions f on $\mathbb{N}$, such that (to avoid summability problems) each function has a finite support. Assume moreover that for each i and each $f \in \mathcal{F}$ we have $f(i) \geq 0$. We are interested in the quantity

$$\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\geq 0}\delta_i f(i)\ .$$

According to the philosophy previously explained, we would like that for some universal constant K there always exists a simple witness that the set

$$\sup_{f\in\mathcal{F}}\sum_{i\geq 0}\delta_i f(i) \geq K\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\geq 0}\delta_i f(i) \tag{5.108}$$

has a probability at most 1/2.

There is a simple and natural choice for these witnesses. For a finite subset I of $\mathbb{N}$, let us consider the event $B(I)$ defined by

$$B(I) = \{\forall i \in I, \delta_i = 1\}\,,$$

so that $\mathsf{P}(B(I)) = \delta^{\operatorname{card} I}$.

Research problem 5.5.1. Is it true that we can find a universal constant K such that given a class of functions $\mathcal{F}$ as in (5.108), we can find a family $\mathcal{G}$ of subsets I of $\mathbb{N}$ with

$$\sum_{I\in\mathcal{G}}\delta^{\operatorname{card} I} \leq 1/2 \tag{5.109}$$

$$\Big\{\sup_{f\in\mathcal{F}}\sum_{i\geq 0}\delta_i f(i) \geq K\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\geq 0}\delta_i f(i)\Big\} \subset \bigcup_{I\in\mathcal{G}} B(I)\,?$$

In this problem, the sets $B(I)$ play the role that the half-spaces play for Gaussian processes in (5.106).

Part of the beauty of Problem 5.5.1 is that possibly the best way to approach it is through a natural question of a more general nature. To formulate this more general question, we need to consider the law P of the sequence $(\delta_i)_{i\geq 0}$ in $\{0,1\}^{\mathbb{N}}$. With some abuse of notation, we will denote by $(\delta_i)_{i\geq 0}$ the generic point of $\{0,1\}^{\mathbb{N}}$. We define an abstract operation as follows. Given a set $D \subset \{0,1\}^{\mathbb{N}}$ and an integer q, let us define the set $D^{(q)}$ as the subset of $\{0,1\}^{\mathbb{N}}$ consisting of the sequences $(\delta_i)_{i\geq 1}$ such that

$$\forall (\delta_i^1)_{i\geq 0}, \cdots, (\delta_i^q)_{i\geq 0} \in D, \exists i \in \mathbb{N}\,,\ \delta_i = 1, \forall \ell \leq q, \delta_i^\ell = 0\,.$$

In words, $D^{(q)}$ consist of the sequences (δ_i) such that the set $\{i \in \mathbb{N}\,;\, \delta_i = 1\}$ cannot be covered by q sets of the type $\{i \in \mathbb{N}\,;\, \delta_i = 1\}$ for $\delta \in D$. To understand the link with Problem 5.5.1, we observe that if D is the set consisting of the sequences $(\delta_i)_{i\in\mathbb{N}}$ for which $\sup_{f\in\mathcal{F}} \sum_{i\geq 0} \delta_i f(i) \leq M$, where M is a median of the left-hand side, then $\mathsf{P}(D) \geq 1/2$, while, due to positivity, we have

$$\Big\{\sup_{f\in\mathcal{F}} \sum_{i\geq 0} \delta_i f(i) > qM\Big\} \subset D^{(q)}.$$

Research problem 5.5.2. Prove (or disprove) that there exist an integer q with the following property. Consider any value of δ and any subset D of $\{0,1\}^{\mathbb{N}}$ with $\mathsf{P}(D) \geq 1 - 1/q$. Then we can find a family $\mathcal{G}$ of sets I as in Problem 5.5.1, that satisfies (5.109), and such that

$$\mathsf{P}\Big(D^{(q)} \setminus \bigcup_{I\in\mathcal{G}} B(I)\Big) = 0\,.$$

Maybe one can even get $D^{(q)} \subset \bigcup_{I\in\mathcal{G}} B(I)$. A positive solution of this problem will be rewarded by a \$1000 prize, even if it applies only to sufficiently small values of δ. It seems probable that progress on this question requires methods unrelated to those of this book.

Material on "selector processes" related to this line of thought and to Problem 5.5.2 can be found in the paper [69]. The methods of [69] being unrelated to those of this book, we do not reproduce the results of this paper. We will meet again selector processes in Chapter 6.

6 Applications to Banach Space Theory

6.1 Cotype of Operators from $C(K)$

We start by recalling some basic definitions. More background can be found in classical books such as [9] or [71].

Given an operator U (i.e. a continuous linear map) from a Banach space X to a Banach space Y and a number $q \geq 2$, we denote by $C_q^g(U)$ its Gaussian cotype-q constant, that is, the smallest number A (possibly infinite) for which, given any integer n, any elements $x_1, \cdots, x_n$ of X, we have

$$\Big(\sum_{i \leq n} \|U(x_i)\|^q\Big)^{1/q} \leq A \mathsf{E} \Big\| \sum_{i \leq n} g_i x_i \Big\| \,.$$

Here, $(g_i)_{i \leq n}$ are i.i.d. standard normal, the norm of $U(x_i)$ is in Y and the norm of $\sum_{i \leq n} g_i x_i$ is in X.

Given a number $q \geq 2$, we define the Rademacher cotype-q constant $C_q^r(U)$ as the smallest number A (possibly infinite) such that, given any integer n, any elements $(x_i)_{i \leq n}$ of X, we have

$$\Big(\sum_{i \leq n} \|U(x_i)\|^q\Big)^{1/q} \leq A \mathsf{E} \Big\| \sum_{i \leq n} \epsilon_i x_i \Big\| \,, \tag{6.1}$$

where $(\epsilon_i)_{i \leq n}$ are i.i.d. Bernoulli r.v. The name Rademacher cotype stems from the fact that Bernoulli r.v. are usually (but inappropriately) called Rademacher r.v. in Banach space theory.

Given $1 \leq p \leq q$, we define the (q, p)-summing norm $\|U\|_{q,p}$ of U as the smallest number A (possibly infinite) such that, for any integer n, any vectors $x_1, \cdots, x_n$ of X we have

$$\Big(\sum_{i \leq n} \|U(x_i)\|^q\Big)^{1/q} \leq A \sup\Big\{ \Big(\sum_{i \leq n} |x^*(x_i)|^p\Big)^{1/p} \; ; \; x^* \in X^* \,, \; \|x^*\| \leq 1 \Big\} \,. \tag{6.2}$$

This is an ideal norm, and in particular for an operator W from Y to another Banach space we have

$$\|W \circ U\|_{q,p} \leq \|W\| \|U\|_{q,p} \,. \tag{6.3}$$

The proof is immediate. These quantities are related as follows.

Proposition 6.1.1. *We have*

$$C_q^g(U) \le \sqrt{\frac{\pi}{2}} C_q^r(U) \tag{6.4}$$

$$\|U\|_{q,1} \le C_q^r(U) \ . \tag{6.5}$$

Proof. To prove (6.4) we simply observe that by Proposition 4.1.2 we have $\mathsf{E}\|\sum_{i\le n} \epsilon_i x_i\| \le \sqrt{\pi/2}\mathsf{E}\|\sum_{i\le n} g_i x_i\|$. To prove (6.5) we observe that

$$\begin{aligned} \Big\|\sum_{i\le n} \epsilon_i x_i\Big\| &= \sup\Big\{\sum_{i\le n} \epsilon_i x^*(x_i) \ ; \ x^* \in X^* \, , \, \|x^*\| \le 1\Big\} \\ &\le \sup\Big\{\sum_{i\le n} |x^*(x_i)| \ ; \ x^* \in X^* \, , \, \|x^*\| \le 1\Big\} \ . \end{aligned}$$

□

In the rest of this section we specialize to the case where X is the space ℓ_N^∞ of sequences $x = (x_j)_{j\le N}$ provided with the norm

$$\|x\| = \sup_{j\le N} |x_j| \ .$$

It is however possible to show that these results also hold in the case where $X = C(W)$, the space of continuous functions over a compact topological space W. The reduction technique that allows this is unrelated to the methods of this book, see [26].

Theorem 6.1.2. *Given $q \ge 2$, there exists a number $K(q)$ depending on q only, such that, given an operator U from ℓ_N^∞ to a Banach space Y, we have*

$$\begin{aligned} \sqrt{\frac{2}{\pi}} \max(C_q^g(U), \|U\|_{q,1}) &\le C_q^r(U) \qquad (6.6)\\ &\le K(q) \max(C_q^g(U), \|U\|_{q,1}) \ . \end{aligned}$$

We observe right away that the left-hand side inequality is a consequence of Proposition 6.1.1.

One of the ingredients of the proof of Theorem 6.1.2 is the following result of B. Maurey, that will be proved just after Theorem 6.1.7.

Proposition 6.1.3. *If $1 \le p < q$ there exists a constant $K(p,q)$ depending on p, q only, such that, for any operator U from ℓ_N^∞ to Y, we have*

$$\|U\|_{q,p} \le K(p,q)\|U\|_{q,1} \ . \tag{6.7}$$

Proof of Theorem 6.1.2. In this proof we fix a value of p with $1 < p < 2$, e.g. $p = 3/2$. We will prove that

$$C_q^r(U) \le L(C_q^g(U) + \|U\|_{q,p}), \tag{6.8}$$

from which the right-hand side inequality of (6.6) will follow using (6.7).

For $i \le n$ we consider elements x_i of ℓ_N^∞, and we write $x_i = (x_{ij})_{j \le N}$. We want to prove that

$$\Big(\sum_{i \le n} \|U(x_i)\|^q\Big)^{1/q} \le L\big(C_q^r(U) + \|U\|_{p,q}\big)\mathsf{E}\Big\|\sum_{i \le n} \epsilon_i x_i\Big\| . \tag{6.9}$$

Clearly, it suffices to consider the case where $x_{1i} = 0$ for each $i \le n$. For $j \le N$, consider $t_j \in \mathbb{R}^n$ given by $t_j = (x_{ij})_{i \le n}$, so that $t_1 = 0$. Consider $T = \{t_1, \cdots, t_N\}$, so that

$$b(T) = \mathsf{E} \sup_{j \le N} \sum_{i \le n} \epsilon_i x_{ij} \le \mathsf{E}\Big\|\sum_{i \le n} \epsilon_i x_i\Big\| . \tag{6.10}$$

We appeal to the weak solution of the Bernoulli problem (Theorem 4.3.1) for this value of p. We can write $t_j = t'_j + t''_j$, where $t'_j = (x'_{ij})_{i \le n}, t''_j = (x''_{ij})_{i \le n}$, and

$$\mathsf{E} \sup_{j \le N} \sum_{i \le n} g_i x'_{ij} \le Lb(T) \tag{6.11}$$

$$\forall j \le N \,,\ \Big(\sum_{i \le n} |x''_{ij}|^p\Big)^{1/p} \le Lb(T) . \tag{6.12}$$

(Since we have fixed $p = 3/2$ we do get a universal constant L in the right-hand sides of these inequalities.) Since $t_1 = 0 = t'_1 + t''_1$, we can replace t'_j by $t'_j - t'_1$ and t''_j by $t''_j - t''_1$, so that we can assume that $t'_1 = 0$. For $i \le n$, we consider the elements $x'_i = (x'_{ij})_{j \le N}$ and $x''_i = (x''_{ij})_{j \le N}$ of ℓ_N^∞. Thus $x_i = x'_i + x''_i$. We will prove that

$$\Big(\sum_{i \le n} \|U(x'_i)\|^q\Big)^{1/q} \le LC_q^g(U)b(T) \tag{6.13}$$

$$\Big(\sum_{i \le n} \|U(x''_i)\|^q\Big)^{1/q} \le L\|U\|_{q,p}b(T) . \tag{6.14}$$

Since $\|U(x_i)\| \le \|U(x'_i)\| + \|U(x''_i)\|$, by the triangle inequality in ℓ_n^q we have

$$\Big(\sum_{i \le n} \|U(x_i)\|^q\Big)^{1/q} \le \Big(\sum_{i \le n} \|U(x'_i)\|^q\Big)^{1/q} + \Big(\sum_{i \le n} \|U(x''_i)\|^q\Big)^{1/q} ,$$

and combining with (6.10), (6.13) and (6.14), this proves (6.9) and hence (6.8) and (6.6).

To prove (6.14) we observe that in the quantity

$$\sup\Big\{\sum_{i\le n}|x^*(x_i'')|^p\ ;\ \|x^*\|_1\le 1\Big\}\,,$$

by convexity the supremum is attained at an extreme point of the unit ball of $(\ell_N^\infty)^*=\ell_N^1$. These extreme points are the canonical basis vectors, and so by (6.12) we have

$$\sup\Big\{\big(\sum_{i\le n}|x^*(x_i'')|^p\big)^{1/p}\ ;\ \|x^*\|_1\le 1\Big\}\le Lb(T)\,, \tag{6.15}$$

and this implies (6.14) by definition of the norm $\|U\|_{q,p}$.

To prove (6.13) we observe that since $t_1'=(x_{i1}')_{i\le n}=0$, using Lemma 1.2.8, we have

$$\begin{aligned}\mathsf{E}\Big\|\sum_{i\le n}g_ix_i'\Big\| &= \mathsf{E}\sup_{j\le N}\Big|\sum_{i\le n}g_ix_{ij}'\Big| \\ &\le 2\mathsf{E}\sup_{j\le N}\sum_{i\le n}g_ix_{ij}'\ .\end{aligned}$$

Thus, using (6.11), we see that (6.13) follows from the definition of $C_q^g(U)$.□

We now turn to the computation of $C_q^g(U)$. We denote by $H_q(U)$ the quantity

$$H_q(U)=\sup\Big\{\big(\sum_{i\le n}\|U(x_i)\|^q\big)^{1/q}\Big\}\,,$$

where the supremum is taken over all n and all families $(x_i)_{i\le n}$ with $x_i=\sum_{k\ge 2}a_{ik}u_k$, where $u_k\in\ell_N^\infty$, the elements $(u_k)_{k\ge 2}$ have disjoint supports, $\|u_k\|_\infty\le 1$, and the numbers a_{ik} satisfy

$$\forall k\ge 2\,,\ \sum_{i=1}^n a_{ik}^2\le\frac{1}{\log k}\ . \tag{6.16}$$

The following result is due to S. Montgomery-Smith.

Theorem 6.1.4. *[28] For all $U:\ell_\infty^N\to Y$ we have*

$$\frac{1}{L}H_q(U)\le C_q^g(U)\le LH_q(U)\,.$$

Proof. Suppose first that for $i\le n$ the elements x_i satisfy $x_i=\sum_{k\ge 2}a_{ik}u_k$, where the elements $(u_k)_{k\ge 2}$ of ℓ_N^∞ have disjoint support, $\|u_k\|_\infty\le 1$, and (6.16) holds. Then, if $u_k=(u_{kj})_{j\le N}$, we have

$$\begin{aligned}\Big\|\sum_{i\le n}g_ix_i\Big\| &= \sup_{j\le N}\Big|\sum_{i\le n}g_i\sum_{k\ge 2}a_{ik}u_{kj}\Big| \\ &= \sup_{j\le N}\Big|\sum_{k\ge 2}X_ku_{kj}\Big|\,,\end{aligned}$$

where $X_k = \sum_{i\le n} g_i a_{ik}$. Since the elements u_k have disjoint support and $\|u_k\|_\infty \le 1$, for each j we have $\sum_{k\ge 2}|u_{kj}| \le 1$, and hence we have $\sup_{j\le N}|\sum_{k\ge 2} X_k u_{kj}| \le \sup_k |X_k|$. Now the r.v. X_k are Gaussian and by (6.16) we have $\mathsf{E}X_k^2 \le 1/\log k$. Thus $\mathsf{E}\sup_{k\ge 2}|X_k| \le L$ by Proposition 2.1.7, and thus $\mathsf{E}\|\sum_{i\le n} g_i x_i\| \le L$. Hence

$$\Big(\sum_{i\le n}\|U(x_i)\|^q\Big)^{1/q} \le LC_q^g(U)\,,$$

and thus $H_q(U) \le LC_q^g(U)$.

We now turn to the proof of the converse inequality. Consider for $i \le n$ elements x_i in ℓ_N^∞ , $x_i = (x_{ij})_{j\le N}$ and

$$D := \mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| = \mathsf{E}\sup_{j\le N}\Big|\sum_{i\le n} g_i x_{ij}\Big| \ge \mathsf{E}\sup_{j\le N}\Big(\sum_{i\le n} g_i x_{ij}\Big)_+ = \mathsf{E}\sup_T \sum_{i\le n} g_i t_i$$

where

$$T = \{0\} \cup \{t_j = (x_{ij})_{i\le n}\,,\, j \le N\}\,.$$

Since $0 \in T$, by Theorem 2.1.8 we can find a sequence $(a_k)_{k\ge 2}$ of points of ℓ_n^2, with $\|a_k\|_2 \le 1/\sqrt{\log k}$, such that

$$T \subset LD\ \mathrm{conv}(\{a_k\,,\, k\ge 2\} \cup \{0\})\,.$$

Thus for each $j \le N$, we can find numbers $(u_{jk})_{k\ge 2}$ with $t_j = \sum_{k\ge 2} u_{jk} a_k$ and

$$\forall j \le N\,,\ \sum_{k\ge 2}|u_{jk}| \le LD\,.$$

Writing $a_k = (a_{ik})_{i\le n}$, this means that

$$\forall j \le N\,,\, \forall i \le n\,,\, x_{ij} = \sum_{k\ge 2} u_{jk} a_{ik}\,,$$

so that

$$\forall i \le n\,,\, x_i = \sum_{k\ge 2} a_{ik} u_k, \tag{6.17}$$

where $u_k = (u_{jk})_{j\le N}$. We observe that

$$\sum_{i\le n} a_{ik}^2 = \|a_k\|_2^2 \le \frac{1}{\log k}\,.$$

When we fix the numbers a_{ik}, the quantity $\sum_{i\le n}\|U(\sum_{k\ge 2} a_{ik}u_k)\|^q$ is a convex function of the numbers $(u_{jk})_{j\le N,k\ge 2}$. On the set

$$\Big\{\forall j \le N\,,\, \sum_{k\ge 2}|u_{jk}| \le LD\Big\}\,,$$

the maximum of this function is attained at an extreme point $(v_{jk})_{j\le N,k\ge 2}$. By extremality, for each j, there is at most one value of k for which $v_{jk}\neq 0$, and of course $|v_{jk}|\le LD$. Thus if we define $v_k=(v_{jk})_{j\le N}$, this means that the elements $(v_k)_{k\ge 2}$ have disjoint supports and satisfy $\|v_k\|\le LD$. Hence

$$\begin{aligned}\Big(\sum_{i\le n}\|U(x_i)\|^q\Big)^{1/q} &= \Big(\sum_{i\le n}\Big\|U\Big(\sum_{k\ge 2}a_{ik}u_k\Big)\Big\|^q\Big)^{1/q}\\ &\le \Big(\sum_{i\le n}\Big\|U\Big(\sum_{k\ge 2}a_{ik}v_k\Big)\Big\|^q\Big)^{1/q}\\ &\le LDH_q(U)\,,\end{aligned}$$

where the first inequality follows from the choice of the numbers $(v_{jk})_{j\le N,k\ge 2}$ and the second inequality from the definition of $H_q(U)$. This completes the proof. □

Theorem 6.1.4 is the starting point of a rather complete theory for the cotype of operators from ℓ_N^∞. We will refer the reader to [48] for a full development. We will only illustrate how sharp results can be with a precise example. This example involves the spaces $L_{q,1}(\mu)$, $q\ge 1$, where μ is a positive measure. (These spaces will again be used in Section 6.4.) The norm is given by

$$\|f\|_{q,1}=\int_0^\infty\big(\mu(\{|f|\ge t\})\big)^{1/q}\mathrm{d}t\,. \tag{6.18}$$

(This quantity is not really a norm, but can be shown to be equivalent to a norm.)

We note that

$$p<q\;\Rightarrow\;\|f\|_{q,1}\le K(p,q)\|f\|_\infty^{1-p/q}\|f\|_p^{p/q}\,, \tag{6.19}$$

where $K(p,q)$ depends only on p and q. Indeed, if $q'=q/(q-1)$ denotes the conjugate exponent of q, by Hölder's inequality we have

$$\begin{aligned}\|f\|_{q,1}&=\int_0^{\|f\|_\infty}\big(\mu(\{|f|\ge t\})\big)^{1/q}\mathrm{d}t\\ &\le\Big(\int_0^{\|f\|_\infty}t^{(1-p)\frac{q'}{q}}\mathrm{d}t\Big)^{1/q'}\Big(\int_0^\infty t^{p-1}\mu(\{|f|\ge t\})\mathrm{d}t\Big)^{1/q}\\ &\le K(p,q)\|f\|_\infty^{1-p/q}\|f\|_p^{p/q}\,,\end{aligned}$$

since $(p-1)q'/q=(p-1)/(q-1)<1$ and $(1+(1-p)q'/q)/q'=1-p/q$.

Also, we have, for all probability measures μ

$$q<p\;\Rightarrow\;\|f\|_{q,1}\le K(p,q)\|f\|_p\,. \tag{6.20}$$

Indeed, assuming without loss of generality that $\|f\|_p=1$, we have $\mu(\{|f|\ge t\})\le\min(1,t^{-p})$ and $\|f\|_{q,1}\le K(p,q)$ by (6.18).

Finally, for all probability measures μ we have

$$\|f\|_q \leq K(q)\|f\|_{q,1} \,. \tag{6.21}$$

Indeed, assuming without loss of generality that $\|f\|_{q,1} = 1$, by (6.18) we have $\mu(\{|f| \geq t\})^{1/q} \leq t^{-1}$, so that we have

$$\begin{aligned} t^{q-1}\mu(\{|f| \geq t\}) &\leq (t\mu(\{|f| \geq t\})^{1/q})^{q-1}(\mu(\{|f| \geq t\}))^{1/q} \\ &\leq (\mu(\{|f| \geq t\}))^{1/q} \,, \end{aligned}$$

and thus

$$\|f\|_q^q = \int_0^\infty qt^{q-1}\mu(\{|f| \geq t\})\mathrm{d}t \leq K(q) \,,$$

using (6.18) again.

Proposition 6.1.5. *Consider a probability measure μ on $\{1, \cdots, N\}$. Then if $p < q$ the canonical injection* $\mathrm{Id} : \ell_N^\infty \hookrightarrow L_{q,1}(\mu)$ *satisfies*

$$\|\mathrm{Id}\|_{q,p} \leq K(p,q) \,,$$

where $K(p,q)$ depends on p and q only.

Proof. Consider elements $(x_i)_{i \leq n}$ of ℓ_N^∞, $x_i = (x_{ij})_{j \leq N}$, and assume that

$$\forall x^* \in \ell_N^1 = (\ell_N^\infty)^* \,, \sum_{i \leq n} |x^*(x_i)|^p \leq \|x^*\|^p \,.$$

Thus we have

$$\forall j \leq N \,, \sum_{i \leq n} |x_{ij}|^p \leq 1 \,, \tag{6.22}$$

which we simply write as

$$\sum_{i \leq n} |x_i|^p \leq 1 \,. \tag{6.23}$$

Here and in the next few pages, we view an element of ℓ_N^∞ as a function on $\{1, \cdots, N\}$, so that for $x = (x_j)_{j \leq N} \in \ell_N^\infty$, $|x|^p$ is the element $(|x_j|^p)_{j \leq N}$ of ℓ_N^∞. From (6.23) we have $\|x_i\|_\infty \leq 1$, so that, still viewing x_i as a function on $\{1, \cdots, N\}$, by (6.19) we have

$$\|x_i\|_{q,1}^q \leq K(p,q) \int |x_i|^p \mathrm{d}\mu \,,$$

so that $\sum_{i \leq n} \|x_i\|_{q,1}^q \leq K(p,q)$ by (6.23) and since μ is a probability. □

Remark 6.1.6. When $p = 1$, we can take $K(p,q) = 1$.

The importance of the previous example stems from the fact that it is essentially "generic" as the following factorization theorem, due to G. Pisier, shows.

Theorem 6.1.7. *Given an operator* $U : \ell_N^\infty \to Y$, *there is a probability measure* μ *on* $\{1, \cdots, N\}$ *such that if we denote by* V *the operator* U *as seen operating from* $L_{q,1}(\mu)$ *to* Y *we have*

$$\|V\| \le L\|U\|_{q,1} \,. \tag{6.24}$$

We refer the reader to [32] for a proof.

This result witnesses the value of $\|U\|_{q,1}$ (within the multiplicative constant L). Indeed, by (6.3), Proposition 6.1.5 and Remark 6.1.6 we have

$$\|U\|_{q,1} = \|V \circ \mathrm{Id}\|_{q,1} \le \|V\|\|\mathrm{Id}\|_{q,1} \le \|V\| \,.$$

Proof of Proposition 6.1.3. Consider the positive measure provided by Theorem 6.1.7, and V as in this theorem. Then, using Proposition 6.1.5 and (6.3) we have

$$\|U\|_{q,p} = \|V \circ \mathrm{Id}\|_{q,p} \le \|V\|\|\mathrm{Id}\|_{q,p} \le K(q,p)\|V\| \le LK(q,p)\|U\|_{q,1} \,.$$

□

Here is a simple fact.

Lemma 6.1.8. *If* $M \ge 2$, *for a positive measure* μ *on* $\{1, \cdots, M\}$ *and elements* $(x_i)_{i\le n}$ *of* ℓ_M^∞, *we have*

$$\sum_{i\le n} |x_i|^2 \le 1 \Rightarrow \sum_{i\le n} \|x_i\|_{2,1}^2 \le L \log M \mu(\{1, \cdots, M\}) \,.$$

Proof. By homogeneity we can and do assume that μ is a probability measure. There exists a probability measure μ' on $\{1, \cdots, M\}$ such that $\mu' \ge \mu/2$ and μ' gives mass $\ge 1/(2M)$ to each point of $\{1, \cdots, M\}$. With obvious notation we have $\|x\|_{2,1,\mu} \le \sqrt{2}\|x\|_{2,1,\mu'}$. Thus we can assume without loss of generality that μ gives mass $\ge 1/(2M)$ to each point of $\{1, \cdots, M\}$. We will prove that this implies that

$$\forall x \ , \ \|x\|_{2,1}^2 \le L \log M \|x\|_2^2 \,.$$

This will conclude the proof since $\sum_{i\le n} \|x_i\|_2^2 \le 1$. We set $t_0 = 0$ and for $\ell \ge 1$, we define

$$t_\ell = \sup\{t \ ; \ \mu(\{|x| \ge t\}) \ge 2^{-\ell}\} \,,$$

so that

$$t_\ell < t < t_{\ell+1} \Rightarrow 2^{-\ell-1} \le \mu(\{|x| \ge t\}) \le 2^{-\ell} \tag{6.25}$$

and thus

$$\|x\|_{2,1} = \int_0^\infty \sqrt{\mu(\{|x| \ge t\})} \, \mathrm{d}t \le \sum_{\ell \ge 0} 2^{-\ell/2}(t_{\ell+1} - t_\ell) \,. \tag{6.26}$$

If ℓ_0 is the smallest integer with $2^{-\ell_0} < 1/2M$, for $\ell \geq \ell_0$ we have $t_\ell = t_{\ell_0} = \|x\|_\infty$, so that the sum in (6.26) has in fact at most $(\ell_0 + 1)$ terms. Since $(t_{\ell+1} - t_\ell)^2 \leq t_{\ell+1}^2 - t_\ell^2$, using (6.26), the Cauchy-Schwarz inequality and (6.25) we get

$$\begin{aligned}\|x\|_{2,1}^2 &\leq (\ell_0 + 1) \sum_{\ell \geq 0} (t_{\ell+1}^2 - t_\ell^2) 2^{-\ell} \\ &\leq 4(\ell_0 + 1) \sum_{\ell \geq 0} \int_{t_\ell}^{t_{\ell+1}} t\mu(\{|x| \geq t\}) \,\mathrm{d}t \\ &= 2(\ell_0 + 1)\|x\|_2^2 \,.\end{aligned}$$

□

Theorem 6.1.9. *For an operator U from ℓ_N^∞ to any Banach space Y, we have, for $N \geq 3$*

$$C_q^r(U) \leq L\sqrt{\log\log N}\|U\|_{2,1} \,.$$

Proof. Consider the probability measure provided by Theorem 6.1.7, and the operator V as in that theorem. We have

$$C_q^r(U) = C_q^r(V \circ \mathrm{Id}) \leq \|V\| C_q^r(\mathrm{Id}) \,.$$

Thus it suffices to prove that $C_q^r(U) \leq L\sqrt{\log\log N}$ when $U(= \mathrm{Id})$ is the canonical injection from ℓ_N^∞ to $L_{2,1}(\mu)$, where μ is any probability measure on $\{1, \cdots, N\}$. Using Theorem 6.1.2 and Proposition 6.1.5, it suffices to show that $C_2^g(U) \leq L\sqrt{\log\log N}$, and, using Theorem 6.1.4, that $H_2(U) \leq L\sqrt{\log\log N}$. Consider then elements $(u_k)_{k\geq 2}$ of ℓ_N^∞ with disjoint supports, $\|u_k\|_\infty \leq 1$, and numbers $(a_{ik})_{i\leq n, k\geq 2}$ such that

$$\forall k \geq 2 \,, \sum_{i\leq n} a_{ik}^2 \leq \frac{1}{\log k} \,. \tag{6.27}$$

Set $x_i = \sum_{k\geq 2} a_{ik} u_k$. We want to prove that

$$\sum_{i\leq n} \|x_i\|_{2,1}^2 \leq L \log\log N \,. \tag{6.28}$$

We observe that there are at most N of the elements u_k that are not zero (since they have disjoint support). By renumbering them, we can assume that $k \geq N + 2 \Rightarrow u_k = 0$. For $\ell \geq 0$, we set

$$x_{i,\ell} = \sum_{M_\ell \leq k < M_{\ell+1}} a_{ik} u_k \tag{6.29}$$

where $M_\ell = 2^{2^\ell}$ (so that $M_0 = 2$). Consider the smallest integer ℓ_0 such that $M_{\ell_0} \geq N + 2$. Then

$$x_i = \sum_{0 \le \ell \le \ell_0} x_{i,\ell} \, ,$$

so that, since $\| \cdot \|_{2,1}$ is equivalent to a norm, we have

$$\|x_i\|_{2,1} \le L \sum_{0 \le \ell \le \ell_0} \|x_{i,\ell}\|_{2,1} \, ,$$

and, by the Cauchy-Schwarz inequality,

$$\sum_{i \le n} \|x_i\|_{2,1}^2 \le L(\ell_0 + 1) \sum_{0 \le \ell \le \ell_0, i \le n} \|x_{i,\ell}\|_{2,1}^2 \, .$$

Thus, it suffices to prove that $\sum_{\ell \le \ell_0, i \le n} \|x_{i,\ell}\|_{2,1}^2 \le L$. Denoting by S_ℓ the union of the supports of the vectors u_k for $M_\ell \le k < M_{\ell+1}$, we observe that the sets S_ℓ are disjoint, so that it suffices to prove that

$$\sum_{i \le n} \|x_{i,\ell}\|_{2,1}^2 \le L\mu(S_\ell) \, . \tag{6.30}$$

First we prove that for each ℓ we have $\sum_{i \le n} |x_{i,\ell}|^2 \le L2^{-\ell}$. We set $x_{i,\ell} = (x_{i,\ell,j})_{j \le N}$ and $u_k = (u_{kj})_{j \le N}$. Consider $j \le N$. By (6.29), if j does not belong to the support of any u_k, the numbers $x_{i,\ell,j}$ are 0 for each i. Otherwise, since the supports of the elements u_k are disjoint, j belongs to the support of a unique element u_{k_0}. If either $k_0 < M_\ell$ or $k_0 \ge M_{\ell+1}$, by (6.29) the numbers $x_{i,\ell,j}$ are again 0 for each i. If $M_\ell \le k_0 < M_{\ell+1}$ then by (6.29) for each $i \le n$ we have $|x_{i,\ell,j}| \le a_{ik_0}$, so that $\sum_{i \le n} x_{i,\ell,j}^2 \le 1/\log k_0 \le L2^{-\ell}$ by (6.27), and since $k_0 \ge M_\ell$. This proves the claim that $\sum_{i \le n} |x_{i,\ell}|^2 \le L2^{-\ell}$. Since $\|u_k\|_\infty \le 1$ we have $|u_{kj}| \le 1$. Since the vectors u_k have disjoint support, increasing $|u_{kj}|$ increases $\| \sum_{M_\ell \le k < M_{\ell+1}} a_{ik} u_k \|_{2,1}$, and we can assume without loss of generality that $|u_{kj}| \in \{0, 1\}$. The span of the elements $|x_{i,\ell}|$, $i \le n$ in ℓ_N^∞ consists of functions on $\{1, \cdots, N\}$ that are constants on the sets $\{|u_k| = 1\}$ for $M_\ell \le k < M_{\ell+1}$, and that are zero outside the union of these sets. If we identify each of these sets to a point, we are in a situation where the underlying measured space has at most $M_{\ell+1}$ points, and since $\log M_{\ell+1} \le 2^{\ell+1}$, we see that (6.30) follows from Lemma 6.1.8. □

Theorem 6.1.10. *Consider the uniform probability measure μ on $\{1,\cdots,N\}$. Then for $N \ge 3$ we have*

$$C_2^g(U) \ge \frac{1}{L} \sqrt{\log \log N} \, ,$$

where U is the canonical injection from ℓ_N^∞ into $L_{2,1}(\mu)$.

Thus, we have shown that $\|U\|_{2,1} \le 1$, and that both $C_2^g(U)$ and $C_2^r(U)$ are of order $\sqrt{\log \log N}$.

Proof. To avoid messy details we will assume that N is of the type $N = (p-3)2^{2^p}$ for some $p \geq 4$.

For $2 \leq j \leq p-2$ we consider disjoint sets S_j with card $S_j = 2^{2^{j+2}}$, and $S = \bigcup_{2\leq j\leq p-2} S_j$.

Consider the probability measure ν on S that gives mass $1/((p-3)\,\mathrm{card}\, S_j)$ to each point of S_j. The mass of each point of S is a multiple of N^{-1}, so that $L_{2,1}(\nu)$ is isometric to a subspace of $L_{2,1}(\mu)$, and it suffices to prove that if V is the canonical injection from $\ell^\infty(S)$ into $L_{2,1}(\nu)$ we have

$$H_2(V) \geq \frac{\sqrt{p}}{L}\,. \tag{6.31}$$

We consider the family $\mathcal{X}$ consisting of all the elements x of $\ell^\infty(S)$ of the following type. The element x takes only the values 0 and 2^k for $3 \leq k \leq 2^{p-2}$. If $2^{j-1} < k \leq 2^j$ where $2 \leq j \leq p-2$, then the set $\{x = 2^k\}$ consists of exactly $2^{-2k-2j}\mathrm{card}\, S_j$ points of S_j. This is possible because this number is an integer since $2^{j+2} - 2j - 2k \geq 2^{j+2} - 2j - 2^{j+1} \geq 0$. Thus $\nu(\{x = 2^k\}) = 2^{-2k-2j}/(p-3)$, and

$$\begin{aligned} \|x\|_{2,1} &\geq \sum_k \int_{2^{k-1}}^{2^k} \sqrt{\nu(\{|x| \geq t\})}\,\mathrm{d}t \geq \sum_k 2^{k-1}\sqrt{\nu(\{|x| = 2^k\})} \\ &\geq \sum_{2\leq j\leq p-2}\ \sum_{2^{j-1}<k\leq 2^j} 2^{k-1}\frac{2^{-k-j}}{\sqrt{p-3}} = \frac{p-3}{4\sqrt{p-3}} \geq \frac{\sqrt{p}}{L}\,. \end{aligned} \tag{6.32}$$

Let us consider the family $\mathcal{F}$ consisting of the elements of $\ell^\infty(S)$ of the type $x/\sqrt{M}$, where $x \in \mathcal{X}$ and $M = \mathrm{card}\mathcal{X}$. Then, by (6.32), we have

$$\sum_{y\in\mathcal{F}} \|y\|_{2,1}^2 \geq \frac{p}{L}\,. \tag{6.33}$$

For $x \in \mathcal{X}$, the average value of x^2 on the set S_j is

$$\sum_{2^{j-1}<k\leq 2^j} 2^{2k}2^{-2k-2j} = 2^{-j-1}$$

and thus, writing $y = (y_k)_{k\in S}$, we have

$$\forall k \in S_j\,,\ \sum_{y\in\mathcal{F}} y_k^2 = 2^{-j-1}\,, \tag{6.34}$$

since by symmetry $\sum_{y\in\mathcal{F}} y_k^2$ is independent of $k \in S_j$. We can and do assume that the sets S_j are consecutive intervals. In that case, for $k \in S_j$ we have $\log k \leq L2^j$, and (6.34) implies that

$$\forall k \in S_j\,,\ \sum_{y\in\mathcal{F}} y_k^2 \leq \frac{L}{\log k}\,. \tag{6.35}$$

If we denote by (u_k) the canonical basis of $\ell^\infty(S)$ then

$$y = \sum_{k\in S} y_k u_k$$

and the elements u_k have disjoint supports. Combining with (6.33) and (6.35) we have indeed shown that $H_2(V) \geq \sqrt{p}/L$. $\square$

6.2 Computing the Rademacher Cotype-2 Constant

When U is an operator between two finite dimensional Banach spaces X and Y, one may ask "how many vectors of X are needed in general to compute the Rademacher cotype-2 constant $C_2^r(U)$ of U" within a constant L, that is, how large should n be so that one can find $(x_1, \cdots, x_n)$ in X with

$$\Big(\sum_{i\leq n} \|U(x_i)\|^2\Big)^{1/2} > \frac{1}{L} C_2^r(U) \mathsf{E}\Big\|\sum_{i\leq n} \epsilon_i x_i\Big\| \,.$$

Similar questions in various settings are investigated e.g. in [71], [14]. We will approach this question through a comparison principle between Gaussian and Rademacher averages that is of interest in its own right.

Consider a Banach space X of dimension N, and its dual X^*. Consider elements $x_1\,, \cdots, x_n$ in X and assume without loss of generality that they span X.

Consider the norm $\|\cdot\|_2$ on X such that its unit ball is the set

$$\Big\{\sum_{i\leq n} \alpha_i x_i \;;\; \sum_{i\leq n} \alpha_i^2 \leq 1\Big\} \,.$$

Let us denote by $\|\cdot\|_2$ the dual of this norm on X^*, so that

$$\begin{aligned} \|x^*\|_2 &= \sup\Big\{\Big|x^*\Big(\sum_{i\leq n} \alpha_i x_i\Big)\Big| \;;\; \sum_{i\leq n} \alpha_i^2 \leq 1\Big\} \\ &= \Big(\sum_{i\leq n} x^*(x_i)^2\Big)^{1/2} \,. \end{aligned} \tag{6.36}$$

This norm arises from the dot product given by

$$(x^*, y^*) = \sum_{i\leq n} x^*(x_i) y^*(x_i) \,.$$

Consider an orthonormal basis $(e_j^*)_{j\leq N}$ of X^* for this dot product. Then

$$x^* = \sum_{j\leq N} (x^*, e_j^*)\, e_j^* \;;\; \|x^*\|_2^2 = \sum_{j\leq N} (x^*, e_j^*)^2 \,.$$

Thus

$$\begin{aligned}\|x\|_2 &= \sup\{|x^*(x)| \ ; \ \|x^*\|_2 \le 1\}\\ &= \sup\Big\{\big|\sum_{j\le N} \beta_j e_j^*(x)\big| \ ; \ \sum_{j\le N} \beta_j^2 \le 1\Big\} = \big(\sum_{j\le N} e_j^*(x)^2\big)^{1/2} .\end{aligned}$$

We note that

$$\begin{aligned}\sum_{i\le n} \|x_i\|_2^2 &= \sum_{i\le n}\sum_{j\le N} e_j^*(x_i)^2 \\ &= \sum_{j\le N}\sum_{i\le n} e_j^*(x_i)^2 = N ,\end{aligned} \tag{6.37}$$

using (6.36) with $x^* = e_j^*$ in the last inequality.

It is of interest to consider a subset T of X as a subset of the Hilbert space $(X, \|\cdot\|_2)$. One can then define the usual quantity $g(T)$, that is concretely given by

$$g(T) = \mathsf{E} \sup_{t\in T} \sum_{j\le N} g_j e_j^*(t) , \tag{6.38}$$

where $(g_i)_{j\le N}$ are independent standard normal r.v. (Interestingly, this formula will not be needed in the sequel.)

Lemma 6.2.1. *If $T = \{x_1, \cdots, x_n\}$ then*

$$g(T) \le L\sqrt{\log(N+1)} . \tag{6.39}$$

If the sequence $(\|x_i\|_2)_{i\ge1}$ is non-increasing, if $M = N\log N$ and if we write $T' = \{x_i \ ; \ M \le i \le n\}$ we have

$$g(T') \le L . \tag{6.40}$$

Proof. Both results are based on the fact that if $T = \{t_k \ ; \ k \ge 1\}$ then

$$g(T) \le L \sup_{k\ge1}\big(\|t_k\|_2 \sqrt{\log(k+1)}\big) ,$$

as shown in Proposition 2.1.7. We observe that it is obvious from the definition of $\|\cdot\|_2$ that $\|x_i\|_2 \le 1$. Assuming without loss of generality that the sequence $(\|x_i\|)_{i\ge1}$ is non-increasing, we see from (6.37) that $\|x_i\| \le \sqrt{N/i}$. Thus

$$\begin{aligned} g(T) &\le L \sup_{k\ge1}\Big(\min\Big(1, \sqrt{\frac{N}{k}}\Big)\sqrt{\log(k+1)}\Big) \le L\sqrt{\log(N+1)} \\ g(T') &\le L \sup_{k\ge1}\Big(\sqrt{\frac{N}{M+k}}\sqrt{\log(k+1)}\Big) \le L\sqrt{\frac{N}{M}\log M} \le L .\end{aligned}$$

$\square$

In the next statement, we define $T = \{x_1, \cdots, x_n\}$, and, for a subset I of $\{1, \cdots, n\}$ we write

$$T_I = \{x_i \,,\, i \le n \,,\, i \notin I\} \,.$$

Theorem 6.2.2. *We have*

$$\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\le n} \epsilon_i x_i\Big\|(1 + g(T)) \,. \tag{6.41}$$

More generally, for any subset I of $\{1, \cdots, n\}$ we have

$$\mathsf{E}\Big\|\sum_{i\notin I} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\notin I} \epsilon_i x_i\Big\| \left(1 + \frac{\mathsf{E}\|\sum_{i\le n} g_i x_i\|}{\mathsf{E}\|\sum_{i\notin I} g_i x_i\|} g(T_I)\right) \,. \tag{6.42}$$

Of course (6.41) is the special case of (6.42) where $I = \emptyset$. Using (6.39) we see that (6.41) improves the known inequality

$$\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \le L\sqrt{\log(N+1)}\mathsf{E}\Big\|\sum_{i\le n} \epsilon_i x_i\Big\| \,. \tag{6.43}$$

Corollary 6.2.3. *There exists a subset I of $\{1, \cdots, n\}$ such that* $\operatorname{card} I \le N\log(N+1)$ *and that* **either** *of the following holds true*

$$\mathsf{E}\Big\|\sum_{i\notin I} g_i x_i\Big\| \le \frac{1}{2}\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \tag{6.44}$$

or

$$\mathsf{E}\Big\|\sum_{i\notin I} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\notin I} \epsilon_i x_i\Big\| \,. \tag{6.45}$$

Proof. By (6.40) we can find a set I with the required cardinality such that $g(T_I) \le L$, so that if (6.44) fails, (6.45) follows from (6.42). $\square$

Corollary 6.2.4. *Consider an operator U from X to Y, and vectors $(x_i)_{i\le n}$ of X such that*

$$A\mathsf{E}\Big\|\sum_{i\le n} \epsilon_i x_i\Big\| < \Big(\sum_{i\le n} \|U(x_i)\|^2\Big)^{1/2} \,. \tag{6.46}$$

Then we can find vectors $(y_j)_{j\le M}$ of X such that

$$\frac{A}{L}\mathsf{E}\Big\|\sum_{j\le M} \epsilon_j y_j\Big\| < \Big(\sum_{j\le M} \|U(y_j)\|^2\Big)^{1/2} \tag{6.47}$$

and $M \le N\log N\log\log N$.

For every $A < C_2^r(U)$, there exists vectors $x_1, \ldots, x_n$ such that (6.46) is satisfied, and (6.47) means that within the loss of a constant factor one can take $n = M$. In other words, the “Rademacher cotype 2 constant of U can essentially be computed on M vectors”.

Of course, one should ask whether it would actually suffice to consider LN vectors.

Proof. The first part of the proof consists in showing that we can find a subset J of $\{1, \cdots, n\}$ with $\mathrm{card} J \le M$ and

$$\mathsf{E}\Big\|\sum_{i \notin J} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\le n} \epsilon_i x_i\Big\| \,. \tag{6.48}$$

To this aim, consider the largest integer k_0 with $2^{k_0} \le \sqrt{\log(N+1)}$. Using Corollary 6.2.3, by induction over k, for $k \le k_0$ we construct subsets I_k of $\{1, \cdots, n\}$ with $\mathrm{card} I_k \le LN\log N$ and either

$$\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_k} g_i x_i\Big\| \le \frac{1}{2}\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_{k-1}} g_i x_i\Big\|.$$

or

$$\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_k} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_{k-1}} \epsilon_i x_i\Big\|. \tag{6.49}$$

If at one step (6.49) holds, we stop the construction. Taking $J = I_1 \cup \ldots \cup I_k$ we see that $\mathrm{card}\, J \le kN\log N \le M$ and that (6.48) holds. Otherwise, for $J = I_1 \cup \ldots \cup I_{k_0}$, we get $\mathrm{card}\, J \le k_0 N \log N \le M$ and by (6.43) that

$$\mathsf{E}\Big\|\sum_{i\notin J} g_i x_i\Big\| \le 2^{-k_0}\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \le 2^{-k_0} L\sqrt{\log(N+1)}\mathsf{E}\Big\|\sum_{i\le n}\epsilon_i x_i\Big\| \,,$$

and this proves (6.48) by the choice of k_0.

Now that we have proved (6.48) we consider 2 cases.

Case 1. We have

$$\sum_{i\in J}\|U(x_i)\|^2 \ge \frac{1}{2}\sum_{i\le n}\|U(x_i)\|^2 \,.$$

Then we have

$$\frac{A}{2}\mathsf{E}\Big\|\sum_{i\in J}\epsilon_i x_i\Big\| \le \frac{A}{2}\mathsf{E}\Big\|\sum_{i\le n}\epsilon_i x_i\Big\| < \frac{1}{2}\Big(\sum_{i\le n}\|U(x_i)\|^2\Big)^{1/2} \le \Big(\sum_{i\in J}\|U(x_i)\|^2\Big)^{1/2} \,,$$

and this proves (6.47).

Case 2. We have

$$\sum_{i\notin J}\|U(x_i)\|^2 \ge \frac{1}{2}\sum_{i\le n}\|U(x_i)\|^2 \,.$$

Then (6.46) yields

$$\frac{A}{2}\mathsf{E}\Big\|\sum_{i\le n}\epsilon_i x_i\Big\| < \Big(\sum_{i\notin J}\|U(x_i)\|^2\Big)^{1/2}$$

and combining with (6.48) we see that

$$\frac{A}{L}\mathsf{E}\Big\|\sum_{i\notin J} g_i x_i\Big\| < \Big(\sum_{i\notin J}\|U(x_i)\|^2\Big)^{1/2} . \tag{6.50}$$

To conclude the proof, we use that the Gaussian cotype 2 constant of U can be computed on N vectors [71], so that by (6.50) we can find N vectors $y_1, \dots, y_N$ of X such that

$$\frac{A}{L}\mathsf{E}\Big\|\sum_{j\le N} g_j y_j\Big\| \le \Big(\sum_{j\le N}\|U(y_j)\|^2\Big)^{1/2} ,$$

which by (4.3) implies (6.47). □

The proof of Theorem 6.2.2 will use the following general fact. We recall that $N_0 = 1$ and that $N_n = 2^{2^n}$ for $n \ge 1$.

Lemma 6.2.5. *Consider a set T provided with two distances d and d'. Assume that for a certain number S and every $n \ge 0$, every ball $B_d(t,a)$ of T can be covered by N_n sets of d'-diameter at most $aS2^{-n/2}$. Then we have*

$$\gamma_1(T,d') \le LS\gamma_2(T,d) .$$

Proof. Consider an admissible sequence $(\mathcal{B}_n)$ of T with

$$\forall t \in T , \sum_{n\ge 0} 2^{n/2}\Delta(\mathcal{B}_n(t),d) \le 2\gamma_2(T,d) .$$

We construct by induction an increasing sequence of partitions $(\mathcal{C}_n)$ satisfying

$$\operatorname{card}\mathcal{C}_n \le N_{n+2} \tag{6.51}$$

$$\forall C \in \mathcal{C}_n , \exists B \in \mathcal{B}_n , C \subset B , \Delta(C,d') \le S2^{-n/2}\Delta(B,d) . \tag{6.52}$$

First, we set $\mathcal{C}_0 = \{T\}$. We note that using the hypothesis for $a = \Delta(T,d)$ and $n = 0$ we have

$$\Delta(T,d') \le S\Delta(T,d) . \tag{6.53}$$

Thus (6.52) is true for $n = 0$. Assuming that $\mathcal{C}_n$ has been constructed, we split each element C of $\mathcal{C}_n$ as follows. First we split C in the sets $C\cap B$, $B \in \mathcal{B}_{n+1}$. Then we split each set $C \cap B$ in N_{n+1} pieces C' such that

$$\Delta(C',d') \le S2^{-(n+1)/2}\Delta(C\cap B,d) .$$

This is possible by hypothesis, and this completes the construction of $\mathcal{C}_{n+1}$. Clearly, $\mathcal{C}_{n+1}$ consists of at most $N_{n+2}\cdot N_{n+1}^2 = N_{n+3}$ sets and it is obvious that (6.51) and (6.52) hold for $n+1$. A consequence of (6.52) is that

$$\forall t , \Delta(C_n(t),d') \le S2^{-n/2}\Delta(B_n(t),d)$$

and thus

$$\sum_{n\geq 0} 2^n \Delta(C_n(t), d') \leq S \sum_{n\geq 0} 2^{n/2} \Delta(B_n(t), d)$$
$$\leq 2S\gamma_2(T, d) .$$

Using (6.53) and Lemma 1.3.3 then yields the result. □

Proof of Theorem 6.2.2. We prove (6.42). On X^* consider the norm given by

$$\|x^*\|_I = \sup_{i\notin I} |x^*(x_i)| .$$

It follows from the Pajor-Tomczak reverse Sudakov minoration inequality ([18], equation (3.15)) that for $n \geq 0$ the unit ball of $(X^*, \|\cdot\|_2)$ can be covered by N_n balls for $\|\cdot\|_I$ of radius $Lg(T_I)2^{-n/2}$. Thus, by Lemma 6.2.5 we have

$$\gamma_1(X_1^*, d_I) \leq Lg(T_I)\gamma_2(X_1^*, \|\cdot\|_2) ,$$

where of course d_I is the (quasi-) distance associated to the norm $\|\cdot\|_I$ and where X_1^* is the unit ball of X^*. Using Theorem 2.1.1 for the process given for x^* in X_1^* by $X_{x^*} = \sum_{i\leq n} g_i x^*(x_i)$ we get

$$\gamma_1(X_1^*, d_I) \leq Lg(T_I)\mathsf{E}\Big\|\sum_{i\leq n} g_i x_i\Big\| . \tag{6.54}$$

Consider now the set

$$T^\sim = \{(x^*(x_i))_{i\notin I} \ ; \ x^* \in X_1^*\} .$$

It should be obvious that

$$g(T^\sim) = \mathsf{E}\Big\|\sum_{i\notin I} g_i x_i\Big\| \ ; \ b(T^\sim) = \mathsf{E}\Big\|\sum_{i\notin I} \epsilon_i x_i\Big\| \ ;$$
$$\gamma_1(T^\sim, d_\infty) = \gamma_1(X_1^*, d_I) . \tag{6.55}$$

We appeal to Theorem 4.2.1 to see that

$$g(T^\sim) \leq L\big(b(T^\sim) + \sqrt{b(T^\sim)\gamma_1(T^\sim, d_\infty)}\big)$$
$$\leq L\big(b(T^\sim) + \sqrt{b(T^\sim)g(T^\sim)A}\big)$$

where

$$A = g(T_I)\frac{\mathsf{E}\|\sum_{i\leq n} g_i x_i\|}{\mathsf{E}\|\sum_{i\notin I} g_i x_i\|} ,$$

using (6.54) and (6.55). Using the inequality $\sqrt{xy} \leq cx + y/c$, we conclude that

$$g(T^\sim) \leq Lb(T^\sim) + Lb(T^\sim)A + \frac{1}{2}g(T^\sim) ,$$

so that $g(T^\sim) \leq L(1 + A)b(T^\sim)$. □

6.3 Restriction of Operators

We consider $q > 1$, the space ℓ_N^q, and its canonical basis $(e_i)_{i\le N}$. Consider a Banach space X and an operator $U : \ell_N^q \to X$. We will give conditions under which there are large subsets J of $\{1, \cdots, N\}$ such that the norm of the restriction U_J to the span of the vectors $(e_i)_{i\in J}$ is much smaller than the norm of U. We denote by X_1^* the unit ball of the dual of X, by p the conjugate exponent of q. Setting $x_i = U(e_i)$, we have

$$\begin{aligned}\|U_J\| &= \sup\Big\{\sum_{i\in J} \alpha_i x^*(x_i)\ ;\ \sum_{i\in J} |\alpha_i|^q \le 1\,,\, x^* \in X_1^*\Big\} \\ &= \sup\Big\{\big(\sum_{i\in J} |x^*(x_i)|^p\big)^{1/p}\ ;\ x^* \in X_1^*\Big\}\,.\end{aligned} \tag{6.56}$$

The set J will be constructed by a random choice. That is, given a number $\delta > 0$, we consider i.i.d. r.v. $(\delta_i)_{i\le N}$ with

$$\mathsf{P}(\delta_i = 1) = \delta\ ;\ \mathsf{P}(\delta_i = 0) = 1 - \delta\,, \tag{6.57}$$

and we set $J = \{i \le N\ ;\ \delta_i = 1\}$. (The r.v. δ_i are often called selectors.) Thus, using (6.56), we have

$$\|U_J\|^p = \sup_{t\in T} \sum_{i\le N} \delta_i |t_i|^p\,, \tag{6.58}$$

where

$$T = \{(x^*(x_i))_{i\le N}\ ;\ x^* \in X_1^*\}\,. \tag{6.59}$$

Let us observe that, by interversion of the supremum and the expectation, we have

$$\begin{aligned}\mathsf{E}\|U_J\|^p &\ge \sup_{t\in T} \mathsf{E}\Big(\sum_{i\le N} \delta_i |t_i|^p\Big) \\ &= \delta \sup_{t\in T} \sum_{i\le N} |t_i|^p\,.\end{aligned} \tag{6.60}$$

This demonstrates the relevance of the quantity $\sup_{t\in T} \sum_{i\le N} |t_i|^p$ and why we need to control it from above if we want to control $\mathsf{E}\|U_J\|^p$ from above.

For a subset T of $\mathbb{R}^N$, we set

$$|T|^p = \{(|t_i|^p)_{i\le N}\ ;\ t \in T\}\,.$$

Thus, if T is the set (6.59) we have by (6.58) that

$$\|U_J\|^p = \sup_{t\in |T|^p} \sum_{i\le N} \delta_i t_i\,. \tag{6.61}$$

This shows that to control $\mathsf{E}\|U_J\|$ we need information on the set $|T|^p$. On the other hand, information we might gather from the properties of X as a Banach space is likely to bear on T. The link between the properties of T and $|T|^p$ is provided in Theorem 6.3.1 below, that transfers a certain "smallness" property of T into an appropriate smallness property of $|T|^p$.

We recall from Section 2.5 that for a subset I of $\{1,\cdots,N\}$ and for $a>0$ we write

$$W(I,a)=\{(t_i)_{i\le N}\ ;\ i\notin I\Rightarrow t_i=0\,,\forall i\in I\,,\,|t_i|\le a\}\,.$$

Theorem 6.3.1. *Consider a subset T of $\mathbb{R}^N$ with $0\in T$. Assume that there exists an admissible sequence $(\mathcal{B}_n)$ of T such that*

$$\forall t\in T\,,\ \sum_{n\ge 0}2^n\Delta^p(B_n(t),d_\infty)\le A \tag{6.62}$$

and let

$$B=\max\Bigl(A,\sup_{t\in T}\sum_{i\le N}|t_i|^p\Bigr)\ . \tag{6.63}$$

Then we can find a family $\mathcal{F}$ of couples (I,a) with

$$\forall (I,a)\in\mathcal{F}\,,\ a\,\mathrm{card}\,I\le B/A \tag{6.64}$$

$$\forall n\ge 0\,,\ \mathrm{card}\ \{(I,a)\in\mathcal{F}\ ;\ a\ge\ 2^{-n}\}\le N_{n+2} \tag{6.65}$$

$$|T|^p\subset K(p)A\ \mathrm{conv}\bigcup_{(I,a)\in\mathcal{F}}W(I,a)\ . \tag{6.66}$$

Proof. The proof resembles that of Theorem 2.6.12. Consider the largest integer τ for which $2^\tau\le B/A$. Since $B\ge A$, we have $\tau\ge 0$, and $2^{-\tau}\le 2A/B$.

The set $|T|^p$ is a subset of the ball W of $L^1(\mu)$ of center 0 and radius B, where μ is the counting measure on $\{1,\dots,N\}$. So we can use Theorem 2.6.4 and homogeneity to find an admissible sequence of partitions $(\mathcal{C}_n)$ of T and for each $C\in\mathcal{C}_n$ an integer $\ell(C)\in\mathbb{Z}$, such that if for $t\in W$ we set

$$\ell(t,n)=\ell(C_n(t)) \tag{6.67}$$

we have

$$\forall t\in T\,,\ \mathrm{card}\,\{i\le N\,;\,|t_i|^p\ge 2^{-\ell(t,n)}\}\le 2^{n+\tau}\le\frac{2^nB}{A} \tag{6.68}$$

$$\forall t\in T\,,\ \sum_{n\ge 0}2^{n-\ell(t,n)}\le 12\cdot 2^{-\tau}B\le LA\ . \tag{6.69}$$

Using (6.62) we see that the sequence of partitions $\mathcal{A}_n$ generated by $\mathcal{B}_n$ and $\mathcal{C}_n$ satisfies

$$\forall t\in T\,,\ \sum_{n\ge 0}2^n\Delta^p(A_n(t),d_\infty)\le A\ . \tag{6.70}$$

Moreover, this sequence is increasing and card $\mathcal{A}_n \le N_{n+1}$. Also, the integers $\ell(t,n)$ depend only on $A_n(t)$.

For $A \in \mathcal{A}_n$, $n \ge 0$, let us choose in an arbitrary manner $u(A) \in A$, and set $\pi_n(t) = u(A_n(t))$. We write $\pi_n(t) = (\pi_{n,i}(t))_{i \le N}$ and we define

$$I_0(t) = \{i \le N \; ; \; |\pi_{0,i}(t)|^p \ge 2^{-\ell(t,0)}\} . \tag{6.71}$$

For $n \ge 1$ we define

$$I_n(t) = \{i \le N \; ; \; |\pi_{n,i}(t)|^p \ge 2^{-\ell(t,n)} \, , \, |\pi_{n-1,i}(t)|^p < 2^{-\ell(t,n-1)}\} .$$

Thus, for $n \ge 1$ and $i \in I_n(t)$, we have

$$\begin{aligned} |t_i| &\le |t_i - \pi_{n-1,i}(t)| + |\pi_{n-1,i}(t)| \\ &\le \Delta(A_{n-1}(t), d_\infty) + 2^{-\ell(t,n-1)/p} \end{aligned}$$

and hence

$$|t_i|^p \le K(p)(\Delta(A_{n-1}(t), d_\infty)^p + 2^{-\ell(t,n-1)}) := c(t,n) . \tag{6.72}$$

Since $0 \in T$, this remains true for $n = 0$ if we define $c(t,0) = \Delta(T, d_\infty)^p$. From (6.70) and (6.69) we get

$$\forall t \in T \, , \sum_{n \ge 0} 2^n c(t,n) \le K(p) A . \tag{6.73}$$

We consider the family $\mathcal{F}$ of all pairs $(I_n(t), 2^{-n})$. Thus by (6.68) if $(I,a) \in \mathcal{F}$ we have $a \,\mathrm{card}\, I \le B/A$. Since the set $I_n(t)$ depends only on $A_n(t)$, there are at most N_{n+1} sets of this type, and this proves (6.65), using that $\sum_{k \le n} N_{k+1} \le N_{n+2}$.

Finally,

$$i \in I_n(t) \Rightarrow |t_i|^p \le c(t,n) \tag{6.74}$$

and this implies (6.66) as in the proof of Theorem 2.6.12. □

The smallness criterion provided by (6.66) is perfectly adapted to the control of $\mathsf{E}\|U_J\|^p$.

Theorem 6.3.2. *Consider the set T of (6.59), and assume that (6.62) and (6.63) hold. Consider $\epsilon > 0$ and*

$$\delta = \frac{A}{B \epsilon N^\epsilon \log N} . \tag{6.75}$$

Assume that $\delta \le 1$. Then if the r.v. $(\delta_i)_{i \le N}$ are as in (6.57) and $J = \{i \le N \; ; \; \delta_i = 1\}$, for $v > 0$ we have

$$\mathsf{P}\Big(\|U_J\|^p \ge v K(p) \frac{A}{\epsilon \log N}\Big) \le L \exp\Big(-\frac{v}{L}\Big)$$

and in particular

$$\mathsf{E}\|U_J\|^p \le K(p) \frac{A}{\epsilon \log N} . \tag{6.76}$$

Lemma 6.3.3. *Consider a fixed set I. If $u \ge 6\delta \operatorname{card} I$ we have*

$$P\Big(\sum_{i\in I} \delta_i \ge u\Big) \le \exp\Big(-\frac{u}{2}\log\frac{u}{2\delta \operatorname{card} I}\Big) . \tag{6.77}$$

Proof. We are dealing here with the tails of the binomial law and (6.77) follows from the Chernov bounds. For a direct proof, considering $\lambda > 0$ we write

$$\mathsf{E}\exp\lambda\delta_i \le 1+\delta e^{\lambda} \le \exp(\delta e^{\lambda})$$

so that we have

$$\mathsf{E}\exp\lambda\sum_{i\in I}\delta_i \le \exp(\delta e^{\lambda}\operatorname{card} I)$$

and

$$\mathsf{P}\Big(\sum_{i\in I}\delta_i \ge u\Big) \le \exp(\delta e^{\lambda}\operatorname{card} I - \lambda u) .$$

We take $\lambda = \log(u/(2\delta\operatorname{card} I))$, so that $\lambda \ge 1$ and $\delta e^{\lambda}\operatorname{card} I = u/2 \le \lambda u/2$. □

Proof of Theorem 6.3.2. Consider the family $\mathcal{F}$ provided by Theorem 6.3.1. For $(I,a)\in\mathcal{F}$, we have $a\operatorname{card} I \le B/A$ so that

$$\delta\operatorname{card} I \le \frac{1}{a\epsilon N^{\epsilon}\log N} .$$

Considering $v \ge 6$, we use (6.77) for $u = v/(a\epsilon\log N) \ge 6\delta N^{\epsilon}\operatorname{card} I$ to obtain

$$\begin{aligned}\mathsf{P}\Big(a\sum_{i\in I}\delta_i \ge \frac{v}{\epsilon\log N}\Big) &\le \exp\Big(-\frac{v}{2a\epsilon\log N}\log(N^{\epsilon})\Big) \\ &= \exp\Big(-\frac{v}{2a}\Big) .\end{aligned} \tag{6.78}$$

By (6.65) and a simple computation we have for v and L large enough that

$$\sum_{(I,a)\in\mathcal{F}}\exp\Big(-\frac{v}{2a}\Big) \le L\exp\Big(-\frac{v}{L}\Big) . \tag{6.79}$$

Thus, if we define the event

$$\Omega(v)\ :\ \forall (I,a)\in\mathcal{F}\,,\ a\sum_{i\in I}\delta_i \le \frac{v}{\epsilon\log N}\,,$$

we see from (6.78) and (6.79) that $\mathsf{P}(\Omega(v)^c) \le L\exp(-v/L)$. When $\Omega(v)$ occurs, for t in $W(I,a)$ we have $\sum_{i\le N}\delta_i t_i \le a\sum_{i\in I}\delta_i \le v/(\epsilon\log N)$. By (6.66) for $t\in |T|^p$ we have $\sum_{i\le N}\delta_i t_i \le K(p)vA/(\epsilon\log N)$, and by (6.61) we have $\|U_J\|^p \le K(p)vA/(\epsilon\log N)$. □

Theorem 6.3.4. *Consider $1 < q \le 2$ and its conjugate $p \ge 2$. Consider a Banach space X such that X^* is p-convex (see Definition 3.1.2). Consider vectors $x_1, \cdots, x_N$ of X, and $S = \max_{i \le N} \|x_i\|$. Denote by U the operator $\ell_N^q \to X$ such that $U(e_i) = x_i$. For a number $C > 0$, denote by $\|\cdot\|_C$ the norm on X such that the unit ball of the dual norm is*

$$\Big\{x^* \in X^*, \|x^*\| \le 1 \,;\, \sum_{i \le N} |x^*(x_i)|^p \le C\Big\} . \tag{6.80}$$

Then, for a number $K(\eta, p)$ depending only on p and on the constant η in Definition 3.1.2, if $B = \max(K(\eta,p) S^p \log N, C)$, and if

$$\delta = \frac{S^p}{B \epsilon N^\epsilon} \le 1 , \tag{6.81}$$

we have

$$\mathsf{E}\|U_J\|_C^p \le K(\eta, p) \frac{S^p}{\epsilon} . \tag{6.82}$$

It is remarkable that the right-hand side of (6.82) does not depend on $\|U\|_C$ but only on $S = \max_{i \le n} \|U(e_i)\|$. On the other hand, since $\|U\|_C \le C$ we should think of δ as depending on $\|U\|_C$.

Lemma 6.3.5. *Consider the (quasi) distance d_∞ on X_1^* defined by*

$$d_\infty(x^*, y^*) = \max_{i \le N} |x^*(x_i) - y^*(x_i)| .$$

Then

$$e_k(X_1^*, d_\infty) \le K(p, \eta) S 2^{-k/p} (\log N)^{1/p} \tag{6.83}$$

or, equivalently,

$$\log N(X_1^*, d_\infty, \epsilon) \le K(p, \eta) \Big(\frac{S}{\epsilon}\Big)^p \log N . \tag{6.84}$$

Here X_1^* is the unit ball of X^*, $N(X_1^*, d_\infty, \epsilon)$ is the smallest number of balls for d_∞ of radius ϵ needed to cover X_1^* and e_k is defined in (1.13).

It would be nice to have a simple proof of this statement. The only proof we know is somewhat indirect. It involves geometric ideas. First, one proves a "duality" result, namely that if W denotes the convex hull of the points $(\pm x_i)_{i \le N}$, it suffices to show that

$$\log N(W, \|\cdot\|, \epsilon) \le K(p, \eta) \Big(\frac{S}{\epsilon}\Big)^p \log N . \tag{6.85}$$

This duality result is proved in [5], Proposition 2, *(ii)*. We do not reproduce the simple and very nice argument, that is not related to the ideas of this work. One then observes that X is of type p because X^* is p-convex, with a type p constant depending only on p and η, and a beautiful probabilistic argument of Maurey, that is reproduced e.g. in [60], Lemma 3.2 then yields (6.85).

Proof of Theorem 6.3.4. We combine (6.84) with Theorem 3.1.3 (used for $\alpha = p$) to see that

$$\gamma_{p,p}(X_1^*, d_\infty) \le K(p,\eta) S (\log N)^{1/p} ,$$

i.e. there exists an admissible sequence $(\mathcal{B}_n)$ on X_1^*, such that

$$\forall t \in X_1^* , \sum_{n \ge 0} 2^n \Delta^p(B_n(t), d_\infty) \le K(p,\eta) S^p \log N := A .$$

The set T corresponding to the norm (6.80) is

$$T = \Big\{ (x^*(x_i))_{i \le N} \ ; \ \|x^*\| \le 1 \ ; \ \sum_{i \le N} |x^*(x_i)|^p \le C \Big\} .$$

Thus we can use Theorem 6.3.2 with $B = \max(A, C)$. □

To conclude this section, we describe an example showing that Theorem 6.3.4 is very close to being optimal. Consider two integers r, m and $N = rm$. We divide $\{1, \cdots, N\}$ into m disjoint subsets $I_1, \cdots, I_m$ of cardinality r. We consider $1 < q \le 2$ and the operator $U : \ell_N^q \to \ell_m^q = X$ such that $U(e_i) = e_j$ for $i \in I_j$, where $(e_i)_{i \le N}$, $(e_j)_{j \le m}$ are the canonical bases of ℓ_N^q and ℓ_m^q respectively. It is classical [20] that $X^* = \ell_m^p$ is p-convex. Consider δ with $\delta^r = 1/m$. Then

$$\mathsf{P}(\exists j \le m \ ; \ \forall i \in I_j , \ \delta_j = 1) = 1 - \Big(1 - \frac{1}{m}\Big)^m \ge \frac{1}{L} ,$$

and when this event occurs we have $\|U_J\| \ge r^{1/p}$, since $\| \sum_{i \in I_j} e_i \| = r^{1/q}$ and $\|U_J(\sum_{i \in I_j} e_i)\| = r$. Thus

$$\mathsf{E}\|U_J\|^p \ge \frac{r}{L} . \tag{6.86}$$

On the other hand, let us apply Theorem 6.3.4 to this situation. To ensure (6.80) one has to take $C = r$, so that $B = r$ whenever $K(q) \log N$. If ϵ is such that

$$\delta = \frac{1}{m^{1/r}} = \frac{1}{r \epsilon N^\epsilon} = \frac{1}{r \epsilon m^\epsilon r^\epsilon} ,$$

then ϵ is about $1/r$, and (6.86) shows that (6.82) gives the exact order of $\|U_J\|$ in this case.

6.4 The $\Lambda(p)$ Problem

We denote by λ the uniform measure on $[0,1]$. Consider functions $(x_i)_{i \le N}$ on $[0,1]$ such that

$$\forall i \leq N \,,\ \|x_i\|_\infty \leq 1 \tag{6.87}$$

$$\text{the sequence } (x_i)_{i \leq N} \text{ is orthogonal in } L^2 = L^2(\lambda) \,. \tag{6.88}$$

Consider a number $p > 2$. J. Bourgain [4] proved the remarkable fact that there exists a subset J of $\{1, \cdots, N\}$ with card $J = N^{2/p}$, for which we have an estimate

$$\forall (\alpha_i)_{i \in J} \,,\ \Big\|\sum_{i \in J} \alpha_i x_i\Big\|_p \leq K(p)\Big(\sum_{i \in J} \alpha_i^2\Big)^{1/2} \,, \tag{6.89}$$

where $\|\cdot\|_p$ denotes the norm in $L^p(\lambda)$. The most interesting case of application of this theorem is the case of the trigonometric system. Even in that case, no simpler proof is known. Bourgain's argument is probabilistic, showing in fact that a random choice of J works with positive probability.

We will give a sharpened version of (6.89). Consider r.v. δ_i as in (6.57) with $\delta = N^{2/p-1}$, and $J = \{i \leq N \,;\ \delta_i = 1\}$.

Theorem 6.4.1. *Consider $p < p_1 < \infty$ and $p < p' < 2p$. Then there is a r.v. $W \geq 0$ with $\mathsf{E}W \leq K$ such that for any numbers $(\alpha_i)_{i \in J}$ with $\sum_{i \in J} \alpha_i^2 \leq 1$ we can write*

$$f := \sum_{i \in J} \alpha_i x_i = f_1 + f_2 + f_3 \tag{6.90}$$

where

$$\|f_1\|_{p_1} \leq W \tag{6.91}$$

$$\|f_2\|_2 \leq W\sqrt{\log N} N^{1/p-1/2} \,;\ \|f_2\|_\infty \leq WN^{1/p'} \tag{6.92}$$

$$\|f_3\|_2 \leq WN^{1/p-1/2} \,;\ \|f_3\|_\infty \leq WN^{1/p} \,. \tag{6.93}$$

Here, as well as in the rest of this section, K denotes a number depending only on p, p' and p_1, that need not be the same at each occurrence. To understand Theorem 6.4.1, it helps to keep in mind that by (6.19), for any function h we have

$$\|h\|_{p,1} \leq K(p)\|h\|_2^{2/p}\|h\|_\infty^{1-2/p} \,. \tag{6.94}$$

Thus (6.93) implies that $\|f_3\|_{p,1} \leq KW$ and (6.92) implies that $\|f_2\|_{p,1} \leq KWN^{-1/K}$. Since $\|f_1\|_{p,1} \leq K\|f_1\|_{p_1}$ by (6.20), (6.90) implies that $\|f\|_{p,1} \leq KW$, so that we have the estimate

$$\forall (\alpha_i)_{i \in J} \,,\ \Big\|\sum_{i \in J} \alpha_i x_i\Big\|_{p,1} \leq KW\Big(\sum_{i \in J} \alpha_i^2\Big)^{1/2} \,. \tag{6.95}$$

Moreover, since $\mathsf{P}(\operatorname{card} J \geq N^{2/p}) \geq 1/L$, with positive probability we have both $\operatorname{card} J \geq N^{2/p}$ and $W \leq K$ and in this case we see using (6.21) that (6.95) improves upon (6.89) . Thus Theorem 6.4.1 sharpens Bourgain's result.

Moreover Theorem 6.4.1 shows the exact reason why we cannot increase p in (6.89): the function f of (6.90) might take a value about $N^{1/p}$ on a set of measure about $1/N$. We believe (but cannot prove) that the lower order term f_2 is not needed in (6.90).

We consider the operator $U : \ell^2_N \to L^p$ given by $U(e_i) = x_i$, and we denote by U_J its restriction to ℓ^2_J.

We choose once and for all $p_2 > p_1$. (This might be the time to mention that there is some room in the proof, and that some of the choices we make are simply convenient and in no way canonical.) We consider on L^{p_2} the norms $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ such that the unit ball of the dual norm is given respectively by

$$\Big\{x^* \in L^{q_2} \; ; \; \|x^*\|_{q_2} \le 1 \, , \sum_{i \le N} x^*(x_i)^2 \le N^{1/2-1/p}\Big\} \,. \tag{6.96}$$

$$\Big\{x^* \in L^{q_2} \; ; \; \|x^*\|_{q_2} \le 1 \, , \sum_{i \le N} x^*(x_i)^2 \le N^{1-2/p}\Big\} \,, \tag{6.97}$$

where q_2 is the conjugate exponent of p_2.

Lemma 6.4.2. *We have*

$$\mathsf{E}\|U_J\|_{(1)} \le K \; ; \; \mathsf{E}\|U_J\|_{(2)} \le K\sqrt{\log N} \,.$$

Proof. We appeal to Theorem 6.3.4 with $p = 2$. We recall the classical fact that L^{q_2} is 2-convex [20], and we observe that it is enough to prove the result for N large enough. We then use (6.82) with $S = 1$, noting that for N large enough we have $B = \max(C, K(\eta, p)\log N) = C$ and

$$\delta = N^{2/p-1} \le \frac{1}{B\epsilon N^\epsilon}$$

when $C = N^{1/2-1/p}$ and $\epsilon = 1/2 - 1/p$ (in which case $S^2/\epsilon \le K$) and also when $C = N^{1-2/p}$ and $\epsilon = 1/\log N$ (in which case $S^2/\epsilon \le L\log N$). □

We recall the norm

$$\|f\|_{\psi_2} = \inf\Big\{c > 0 \; ; \; \int \exp\Big(\frac{f^2}{c^2}\Big)\mathrm{d}\lambda \le 2\Big\} \,. \tag{6.98}$$

We denote by $\|\cdot\|^*_{\psi_2}$ the dual norm.

We consider $a = N^{-1/p'}$, $b = N^{1/p'-1/p}$, and the norm $\|\cdot\|_{(3)}$ on L^p such that the unit ball of the dual norm is the set

$$Z = \Big\{x^* \in L^q(\lambda) \; ; \; \|x^*\|_1 \le a \, , \; \|x^*\|^*_{\psi_2} \le b \; ; \; \sum_{i \le N} x^*(x_i)^2 \le N^{1-2/p}\Big\} \,, \tag{6.99}$$

where q is the conjugate of p.

Lemma 6.4.3. *We have* $\mathsf{E}\|U_J\|_{(3)} \le L$.

This uses arguments really different from those of Lemma 6.4.2, and the proof will be given at the end of this section.

Lemma 6.4.4. *Assume that* $\|f\|_{(1)} \le 1$, $\|f\|_{(2)} \le \sqrt{\log N}$, $\|f\|_{(3)} \le 1$. *Then we can write* $f = f_1 + f_2 + f_3$*, where*

$$\|f_1\|_{p_1} \le K \tag{6.100}$$

$$\|f_2\|_2 \le K\sqrt{\log N} N^{1/p-1/2} \; ; \; \|f_2\|_\infty \le K N^{1/p'} \tag{6.101}$$

$$\|f_3\|_2 \le K N^{1/p-1/2} \; ; \; \|f_3\|_\infty \le K N^{1/p} \; . \tag{6.102}$$

Proof of Theorem 6.4.1. This is an obvious consequence of the previous three lemmas, with

$$W = \|U_J\|_{(1)} + \frac{1}{\sqrt{\log N}} \|U_J\|_{(2)} + \|U_J\|_{(3)} \; .$$

□

Proof of Lemma 6.4.4. Since $\|f\|_{(1)} \le 1$, by duality we can write $f = u_1 + u_2$ where $\|u_1\|_{p_2} \le 1$ and $u_2 = \sum_{i \le N} \beta_i x_i$ with $\sum_{i \le N} \beta_i^2 \le N^{1/p-1/2}$. By (6.87) and (6.88) we have $\|u_2\|_2^2 \le N^{1/p-1/2}$. Using that

$$\lambda(\{|f| \ge t\}) \le \lambda\Big(\Big\{|u_1| \ge \frac{t}{2}\Big\}\Big) + \lambda\Big(\Big\{|u_2| \ge \frac{t}{2}\Big\}\Big) \tag{6.103}$$

we see that

$$\lambda(\{|f| \ge t\}) \le K(t^{-p_2} + t^{-2} N^{1/p-1/2}) \le K t^{-p_2}$$

for $t \le c_1 = N^\alpha$, where $\alpha(p_2 - 2) = 1/2 - 1/p$. In particular we have

$$\|f \mathbf{1}_{\{|f| \le c_1\}}\|_{p_1} \le K \; . \tag{6.104}$$

Since $\|f\|_{(2)} \le \sqrt{\log N}$, by duality we can write $f = v_1 + v_2$, where $\|v_1\|_{p_2} \le \sqrt{\log N}$ and $v_2 = \sum_{i \le N} \beta_i x_i$, with $\sum_{i \le N} \beta_i^2 \le (\log N) N^{2/p-1}$, so that $\|v_2\|_2 \le \sqrt{\log N} N^{1/p-1/2}$. Let $c_2 = 3N^{1/p'}$. We next show that

$$\begin{aligned} |f| \mathbf{1}_{\{c_1 \le |f| \le c_2\}} &\le 2|v_1| \mathbf{1}_{\{|v_1| \ge c_1/2\}} + 2|v_2| \mathbf{1}_{\{|v_2| \le 2c_2\}} \\ &:= h_1 + h_2 \; . \end{aligned} \tag{6.105}$$

To see this, we can assume that $c_1 \le c_2$. Assume first that $|v_2| > 2c_2$. Then if $|f| = |v_1 + v_2| \le c_2$, we have $|v_1| > c_2$ so that since $|f| \le c_2$, then $|f| \le c_2 \le |v_1| \mathbf{1}_{\{|v_1| \ge c_2\}}$ and hence (6.105) holds true since $c_2 \ge c_1$. Hence to prove (6.105) we can assume that $|v_2| \le 2c_2$. Since $|f| \le |v_1| + |v_2|$, we are done if $|v_1| \le |v_2|$, since then $|f| \le 2|v_2|$ and $|f| \le 2|v_2| \mathbf{1}_{\{|v_2| \le 2c_2\}}$. If $|v_1| \ge |v_2|$, then $|f| \le 2|v_1|$, so $|f| \mathbf{1}_{\{c_1 \le |f|\}} \le 2|v_1| \mathbf{1}_{\{|v_1| \ge c_1/2\}}$, finishing the proof of (6.105).

Since $\|v_1\|_{p_2} \le \sqrt{\log N}$, we have

$$\lambda(\{|v_1| \ge t\}) \le (\log N)^{p_2/2} t^{-p_2}$$

and a straightforward computation based on the formula

$$\|h\|_{p_1}^{p_1} = \int p_1 t^{p_1-1} \lambda(\{|h| \ge t\}) \mathrm{d}t \tag{6.106}$$

yields $\|h_1\|_{p_1} \le K$. Since $\|h_2\|_\infty \le 2c_2 \le 6N^{1/p'}$ and $\|h_2\|_2 \le 2\|v_2\|_2 \le 2\sqrt{\log N} N^{1/p-1/2}$, we see from (6.105) that

$$f\mathbf{1}_{\{c_1 \le |f| \le c_2\}} = g_1 + g_2 \tag{6.107}$$

where $\|g_1\|_{p_1} \le K$, $\|g_2\|_\infty \le 6N^{1/p'}$ and $\|g_2\|_2 \le 2\sqrt{\log N} N^{1/p-1/2}$.

Since $\|f\|_{(3)} \le 1$, by duality we have $f = w_1 + w_2 + w_3$ with $\|w_1\|_\infty \le a^{-1} = N^{1/p'}$, $\|w_2\|_{\psi_2} \le b^{-1} = N^{1/p-1/p'}$ and $w_3 = \sum_{i \le N} \beta_i x_i$ with $\sum_{i \le N} \beta_i^2 \le N^{2/p-1}$.

Thus

$$\|w_3\|_2 \le N^{1/p-1/2} \tag{6.108}$$

and, using (6.87)

$$\|w_3\|_\infty \le \sum_{i \le N} |\beta_i| \le N^{1/2} (\sum_{i \le N} \beta_i^2)^{1/2} \le N^{1/p} \, .$$

We note that

$$|f|\mathbf{1}_{\{|f| \ge c_2\}} \le 3|w_3| + 2|w_2|\mathbf{1}_{\{|w_2| \ge c_2/3\}} \, . \tag{6.109}$$

To see this, we first observe that this is obvious if $|w_2| > c_2/3$, because then $|w_1| \le N^{1/p'} = c_2/3 \le |w_2|$, so $|f| \le |w_3| + 2|w_2|$. If now $|w_2| \le c_2/3$, since $|w_1| \le c_2/3$, when $|f| \ge c_2$, we must have $|w_3| \ge c_2/3$ and hence $|f| \le |w_1| + |w_2| + |w_3| \le 2c_2/3 + |w_3| \le 3|w_3|$, finishing the proof of (6.109).

By definition of $\|\cdot\|_{\psi_2}$, and since $\|w_2\|_{\psi_2} \le b^{-1}$, we have

$$\int \exp(w_2^2 b^2) \mathrm{d}\lambda \le 2$$

so that

$$\lambda(\{|w_2| \ge t\}) \le 2\exp(-t^2 b^2) \, . \tag{6.110}$$

Since $p' < 2p$ we have $1/p - 1/p' < 1/p'$ and recalling the values of b and c_2 one checks from (6.110) and (6.106), with huge room to spare, that $h_3 = 2|w_2|\mathbf{1}_{\{|w_2| \ge c_2/3\}}$ satisfies $\|h_3\|_{p_1} \le K$. Thus from (6.109) we see that

$$f\mathbf{1}_{\{|f| \ge c_2\}} = g_3 + g_4$$

where $\|g_3\|_2 \le KN^{1/p-1/2}$, $\|g_3\|_\infty \le KN^{1/p}$, and $\|g_4\|_{p_1} \le K$. Combining with (6.104) and (6.107) finishes the proof. □

We turn to the proof of Lemma 6.4.3.

Lemma 6.4.5. *Consider independent standard Gaussian r.v.* $(g_i)_{i\le N}$. *Then, given a set* I *we have*

$$\mathsf{E}\Big\|\sum_{i\in I} g_i x_i\Big\|_{\psi_2} \le L\sqrt{\operatorname{card} I}\,.$$

Proof. We have

$$\mathsf{E}\int \exp\frac{(\sum_{i\in I} g_i x_i)^2}{3\operatorname{card} I}\mathrm{d}\lambda = \int \mathsf{E}\exp\frac{(\sum_{i\in I} g_i x_i)^2}{3\operatorname{card} I}\mathrm{d}\lambda \le L \tag{6.111}$$

because for each $t\in[0,1]$, $g=\sum_{i\in I} g_i x_i(t)$ is a Gaussian r.v. with $\mathsf{E}g^2\le \operatorname{card} I$.

For $u\ge 1$, we have $e^{f^2/u}\le 1+e^{f^2}/u$. We use this for $u=\int\exp f^2 d\lambda$ and integrate to see that

$$\|f\|_{\psi_2}\le\int\exp f^2\mathrm{d}\lambda\,.$$

Taking $f=(3\operatorname{card} I)^{-1/2}\sum_{i\in I} g_i x_i$ and combining with (6.111) yields the result. □

Proof of Lemma 6.4.3. The beginning of the proof uses arguments similar to the Giné-Zinn Theorem (Theorem 2.7.8). We recall the set Z of (6.99), and we set

$$T=\{(x^*(x_i)^2)_{i\le N}\ ;\ x^*\in Z\}$$

so that $\sum_{i\le N} t_i\le N^{1-2/p}$ for $t\in T$, and hence $\delta\sum_{i\le N} t_i\le 1$. Thus

$$\begin{aligned}\mathsf{E}\|U_J\|^2_{(3)} &= \mathsf{E}\sup_{t\in T}\sum_{i\le N}\delta_i t_i\\ &\le 1+\mathsf{E}\sup_{t\in T}\sum_{i\le N}(\delta_i-\delta)t_i\ .\end{aligned}$$

Consider an independent sequence $(\delta'_i)_{i\le N}$ distributed like $(\delta_i)_{i\le N}$. Then, by Jensen's inequality we have

$$\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}(\delta_i-\delta)t_i\Big|\le\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}(\delta_i-\delta'_i)t_i\Big|\ .$$

Consider independent Bernoulli r.v. $(\epsilon_i)_{i\le N}$, independent of the r.v. δ_i and δ'_i. Since the sequences $(\delta_i-\delta'_i)_{i\le N}$ and $(\epsilon_i(\delta_i-\delta'_i))_{i\le N}$ have the same distribution, we have

$$\begin{aligned}\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}(\delta_i-\delta'_i)t_i\Big| &= \mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}\epsilon_i(\delta_i-\delta'_i)t_i\Big|\\ &\le 2\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}\epsilon_i\delta_i t_i\Big|\end{aligned}$$

$$\begin{aligned}
&= 2\mathsf{E}\sup_{t\in T}\Big|\sum_{i\in J}\epsilon_i t_i\Big| \\
&\le \sqrt{2\pi}\mathsf{E}\sup_{t\in T}\Big|\sum_{i\in J} g_i t_i\Big| \\
&\le L\mathsf{E}\sup_{t\in T}\sum_{i\in J} g_i t_i \ ,
\end{aligned}$$

using Proposition 4.1.2 and Lemma 1.2.8, and since $0 \in T$. Since $\mathsf{E}\operatorname{card} J = N^{2/p}$, it suffices to show that given a set I we have

$$\mathsf{E}\sup_{t\in T}\sum_{i\in I} g_i t_i = \mathsf{E}\sup_{x^*\in Z}\sum_{i\in I} g_i x^*(x_i)^2 \le Lab\sqrt{\operatorname{card} I}\ . \tag{6.112}$$

We have $|x^*(x_i)| \le a$ since $\|x^*\|_1 \le a$ and $\|x_i\|_\infty \le 1$, so that for a fixed set I we have

$$\begin{aligned}
d_1(x^*, y^*) &:= \Big(\sum_{i\in I}\big(x^*(x_i)^2 - y^*(x_i)^2\big)^2\Big)^{1/2} \\
&\le 2a\Big(\sum_{i\in I}\big(x^*(x_i) - y^*(x_i)\big)^2\Big)^{1/2} := 2ad_2(x^*, y^*)\ .
\end{aligned}$$

Thus

$$\gamma_2(Z, d_1) \le La\gamma_2(Z, d_2)$$

and, by Theorem 1.2.4 and Theorem 2.1.1, we have

$$\begin{aligned}
\mathsf{E}\sup_{x^*\in Z}\sum_{i\in I} g_i x^*(x_i)^2 &\le La\mathsf{E}\sup_{x^*\in Z} x^*\Big(\sum_{i\in I} g_i x_i\Big) \\
&\le Lab\mathsf{E}\Big\|\sum_{i\in I} g_i x_i\Big\|_{\psi_2}\ ,
\end{aligned} \tag{6.113}$$

where the second inequality holds because $\|x^*\|^*_{\psi_2} \le b$ for $x^* \in Z$. Combining with Lemma 6.4.5 this proves (6.112) and hence Lemma 6.4.3. □

Remark. One can also deduce (6.113) from the classical comparison theorems for Gaussian r.v., see [18].

6.5 Schechtman's Embedding Theorem

One should think that there are many other potential applications of the material presented in this book to Banach Spaces, but, as far as the author is aware, (with the notable exception of [37]) only Theorem 2.1.1 and its corollary Theorem 2.1.5 have been applied in the literature. The following is a particularly elegant application of Theorem 2.1.1. For any integer n, we denote by $\|\cdot\|_2$ the Euclidean norm on $\mathbb{R}^n$.

Theorem 6.5.1. *(Schechtman's embedding theorem [38]) Consider two integers n and m. Denote by S^{m-1} the unit sphere of $\mathbb{R}^m$. Consider a norm $\|\cdot\|$ on $\mathbb{R}^n$, and assume that $\|\cdot\| \le \|\cdot\|_2$. Denote by X_t the canonical Gaussian process on $\mathbb{R}^m$, and by $(e_j)_{j\le n}$ the canonical basis of $\mathbb{R}^n$. Then for every subset T of S^{m-1} there is a linear operator $U : \mathbb{R}^m \to \mathbb{R}^n$ such that*

$$\forall t \in T\ ,\ 1 - L\epsilon \le \|U(t)\| \le 1 + L\epsilon$$

where

$$\epsilon = \frac{\mathsf{E}\sup_{t\in T} X_t}{\mathsf{E}\|\sum_{j\le n} g_j e_j\|}\ .$$

Discussing all the remarkable consequences of this statement in Banach Space Theory goes beyond the purpose of this work, and we refer to [38] and references therein for this.

Proof. Consider independent standard Gaussian random variables $(g_i)_{i\ge1}$, $(g_{ij})_{i,j\ge1}$, and for $t = (t_1,\ldots,t_m) \in \mathbb{R}^m$ define $C_t \in \mathbb{R}^n$ by

$$C_t = \sum_{i\le m, j\le n} t_i g_{ij} e_j = \sum_{j\le n} e_j \Bigl(\sum_{i\le m} g_{ij} t_i\Bigr)\ ,$$

so that the law of C_t in $\mathbb{R}^n$ is the same for all $t \in S^{m-1}$, because in that case the sequence $(\sum_{i\le m} g_{ij} t_i)_{j\le m}$ is an independent sequence of standard normal r.v. Moreover, for the same reason, when $t \in S^{m-1}$ we have

$$\mathsf{E}\|C_t\| = \mathsf{E}\Bigl\|\sum_{j\le n} g_j e_j\Bigr\|\ . \tag{6.114}$$

We fix $t_0 \in T$, and for $t \in S^{m-1}$ we define

$$Y_t = \|C_t\| - \|C_{t_0}\|\ ,$$

so that $\mathsf{E}Y_t = 0$. The key of the proof is to establish the inequality

$$\forall u > 0\ ,\ \forall s\ ,\ t \in S^{m-1}\ ,\ \mathsf{P}(|Y_s - Y_t| \ge u) \le 2\exp\Bigl(-\frac{u^2}{L\|s-t\|_2^2}\Bigr)\ . \tag{6.115}$$

Once this is proved, we proceed as follows. Since $Y_{t_0} = 0$, it follows from Theorem 2.1.5 that

$$\mathsf{E}\sup_{t\in T}|Y_t| \le L\mathsf{E}\sup_{t\in T} X_t\ . \tag{6.116}$$

It follows from Lemma 2.3.6 that

$$\mathsf{P}\Bigl(\|C_{t_0}\| \ge \frac{1}{2}\mathsf{E}\|C_{t_0}\|\Bigr) \ge \frac{1}{L}\ ,$$

and combining with (6.114) and (6.116), we see that we find a realization of the r.v. such that

$$\sup_{t\in T} |\|C_t\| - \|C_{t_0}\|| \le L\mathsf{E}\sup_{t\in T} X_t$$

$$\|C_{t_0}\| \ge \frac{1}{2}\mathsf{E}\|C_{t_0}\| = \frac{1}{2}\mathsf{E}\Big\|\sum_{j\le n} g_j e_j\Big\| .$$

The operator U given by $U(t) = C_t/\|C_{t_0}\|$ then satisfies our requirements.

The proof of (6.115) is very beautiful. First, we note that for any $x \in \mathbb{R}^n$ and any $b \in \mathbb{R}^m$ the r.v $\|x + C_b\|$ and $\|x - C_b\|$ have the same law because the distribution of C_b is symmetric, and thus

$$\mathsf{E}\|x + C_b\| = \mathsf{E}\|x - C_b\| ,$$

and also

$$\begin{aligned} &\mathsf{P}\big(|\|x + C_b\| - \|x - C_b\|| \ge u\big) \qquad (6.117)\\ &\le \mathsf{P}\Big(|\|x + C_b\| - \mathsf{E}\|x + C_b\|| \ge \frac{u}{2}\Big) + \mathsf{P}\Big(|\|x - C_b\| - \mathsf{E}\|x - C_b\|| \ge \frac{u}{2}\Big)\\ &= 2\mathsf{P}\Big(|\|x + C_b\| - \mathsf{E}\|x + C_b\|| \ge \frac{u}{2}\Big) . \end{aligned}$$

Now,

$$\|x + C_b\| = \sup\big\{x^*(x + C_b)\ ;\ x^* \in W\big\} = \sup\big\{Z_{x^*}\ ;\ x^* \in W\big\} ,$$

where W is the unit ball of the dual of the Banach space $(\mathbb{R}^N, \|\cdot\|)$, and where $Z_{x^*} = x^*(x + C_b)$. The crucial fact now is that Lemma 2.1.3 remains true when the Gaussian process Z_t is not necessarily centered, provided one replaces the condition $\mathsf{E}Z_t^2 \le \sigma^2$ by the condition $\mathsf{E}(Z_t - \mathsf{E}Z_t)^2 \le \sigma^2$. (This property, as Lemma 2.1.3, takes its roots in the remarkable behavior of the canonical Gaussian measure on $\mathbb{R}^k$ with respect to Lipschitz functions [17].) We have

$$\mathsf{E}(Z_{x^*} - \mathsf{E}Z_{x^*})^2 = \mathsf{E}(x^*(C_b))^2 = \sum_{i\le m, j\le n} x^*(e_j)^2 b_i^2 .$$

Since we assume that $\|\cdot\| \le \|\cdot\|_2$, for $x^* \in W$, we have $|x^*(e_j)| \le 1$, and thus $\mathsf{E}(Z_{x^*} - \mathsf{E}Z_{x^*})^2 \le \|b\|_2^2$. We can then deduce from the extension of Lemma 2.1.3 mentioned above that

$$\mathsf{P}\Big(|\|x + C_b\| - \mathsf{E}\|x + C_b\|| \ge \frac{u}{2}\Big) \le 2\exp\Big(-\frac{u^2}{8\|b\|_2^2}\Big) ,$$

and combining with (6.117) we have

$$\mathsf{P}\big(|\|x + C_b\| - \|x - C_b\|| \ge u\big) \le 4\exp\Big(-\frac{u^2}{8\|b\|_2^2}\Big) . \qquad (6.118)$$

Consider finally s and t in S^{m-1}. Writing $a = (s + t)/2$ and $b = (s - t)/2$ we notice that

$$C_s = C_a + C_b \; ; \; C_t = C_a - C_b \; .$$

Most importantly, since $\|s\| = \|t\|$, the vectors a and b are orthogonal, so that by the rotational invariance property of Gaussian measures the random vectors C_a and C_b are independent, and (6.115) follows using (6.118) for $x = C_a$ conditionally on C_a. □

6.6 Further Reading

The paper [65] presents a significant extension of Theorem 6.3.4, but, unfortunately, we did not see how to simplify the original proof using the ideas of the present work (although of course the proof can be translated in the language of the generic chaining).

Rudelson's paper [37] contains a very beautiful construction in the spirit of the present work. Another application of the present methods to Banach space theory is found in [59], but the construction there is unfortunately cluttered by technical complications, which it would be interesting to remove. See also [79].

References

1. Ajtai, M.; Komlós, J.; Tusnády, G. (1984) On optimal matchings. *Combinatorica* **4**, no. 4, pp. 259–264.
2. Arcones, M. A. (1999) The class of Gaussian chaos of order two is closed by taking limits in distribution. Advances in stochastic inequalities (Atlanta, GA, 1997), *Contemp. Math.* **234**, Amer. Math. Soc., Providence, RI, pp. 13-19.
3. Bollobas, B. Combinatorics, Set Systems, Hypergraphs
4. Bourgain, J. (1989) Bounded orthogonal sets and the Λ_p-problem, *Acta. Math.* **162**, pp 227-246.
5. Bourgain, J.; Pajor, A.; Szarek, S.J.; Tomczak-Jaegerman, N. (1989) On the duality problem for entropy numbers, in Geometric aspects of Functional Analysis, *Lecture Notes in Math* **1376**, pp. 53-63.
6. Coffman, E. G. Jr.; Shor, P. W. (1991) A simple proof of the $O(\sqrt{n}\log^{3/4} n)$ upright matching bound. *SIAM J. Discrete Math.* **4** no. 1, pp. 48-57.
7. Dudley, R. M. (1967) The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis* **1**, pp. 290-330.
8. Dudley, R. M. (1999) Uniform central limit theorems. Cambridge Studies in advanced mathematics 63, Cambridge University Press.
9. Diestel, J; Jarchow, H; Tonge, A (1955) Absolutely summing operators. Cambridge Studies in Advanced Mathematics 43, Cambridge University Press, Cambridge.
10. Fernique, X. (1975) Regularité des trajectoires des fonctions aléatoires gaussiennes. Ecole d'été de Probabilités de Saint-Flour, IV-1974, *Lecture Notes in Math.* **480**, pp. 1-96., Springer-Verlag, Berlin, 1975.
11. Fernique, X. (1995) Eléments d'analyse des chaos gaussiens. *Probab. Math. Statist.* **15**, pp. 291-300.
12. Giné, E.; Zinn, J. (1984) Some limit theorems for empirical processes. *Ann. Probab.* **12**, pp. 929-989.
13. Heinrich, P. (2001) The size of diagonal two-order Gaussian chaoses. *Statist. Probab. Lett.* **51** no. 1, pp. 35-39.
14. König, H.; Tzafriri, L. (1981) Some estimates for type and cotype constants. *Math. Ann.* **256** , no. 1, pp. 85-94.
15. Kôno, N. (1980) Sample path properties of stochastic processes. *J. Math. Kyoto Univ.* **20**, pp. 295-313.
16. Latala, R. (1997) Sudakov minoration principle and supremum of some processes. *Geom. funct. anal.* **7**, pp. 936-953.
17. Ledoux, M. (2001) The concentration of measure phenomenon. *Mathematical Surveys and Monographs* **89**. American Mathematical Society, Providence, RI.
18. Ledoux, M.; Talagrand, M. (1991) Probability in a Banach Space. *Springer-Verlag.*

19. Leighton, T.; Shor, P. (1989) Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica* **9**, no. 2, pp. 161-187.
20. Lindenstrauss, J., Tzafriri, L. (1979) Classical Banach Spaces, II. *Springer-Verlag.*
21. Marcus, M. B. (1987) ξ-radial processes and random Fourier series. *Mem. Amer. Math. Soc.* **68** no. 368, viii+181 pp.
22. Marcus, M. B., Pisier, G. (1981) Random Fourier series with applications to harmonic analysis. *Annals of Mathematics Studies* **101**. Princeton University Press.
23. Marcus, M. B.; Pisier, G. (1984) Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. *Acta Math.* **152**, no. 3-4, pp. 245-301.
24. Marcus, M.; Rosinski, J. (2003) Sufficient conditions for boundedness of moving averages, Manuscript.
25. Marcus, M.; Talagrand, M. (1998) Continuity conditions for a class of Gaussian chaoses related to Wick squares. *J. Th. Probab.* **11**, pp. 157-179.
26. Maurey, B. (1974) Type et cotype dans les espaces munis de structures locales inconditionelles. Séminaire Maurey-Schwartz 1973-74, Exposé 12, Paris, Ecole Polytechnique 1974.
27. Mendelson, S. Vershynin, R. (2003) Entropy and the combinatorial dimension. *Inv. Math.* **152**. no 3, pp. 37-55.
28. Montgomery-Smith, S.J. (1988) The cotype of operators form $C(K)$, Ph. D. thesis, Cambridge.
29. Montgomery-Smith, S. J. (1989) The Gaussian cotype of operators from $C(K)$. *Israel J. Math.***68**, no. 1, pp. 123-128.
30. Montgomery-Smith, S.J.; Talagrand, M. (1990) The Rademacher cotype of Operators from ℓ_∞^N. *Israel J. Math.* **68**, pp. 123-128.
31. Ossiander, M (1987) A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.***15**, no. 3, pp. 897-919.
32. Pisier, G.(1986) Factorization of operators through $L_{p,\infty}$ or $L_{p,1}$ and non-comutative generalizations. *Math. Ann.* **276**, pp. 105-136.
33. Pisier, G. (1980) Conditons d'entropie assurant la continuité de certains processus et applications à l'analyse harmonique. Séminaire d'analyse fonctionelle 1979-1980, Ecole Polytechnique, Paris.
34. Rhee, W.; Talagrand, M. (1988) Exact bounds for the stochastic upward matching problem. *Trans. Amer. Math. Soc.* **109**, pp. 109-126.
35. Rosinski, J. On series representations of infinitely divisible random vectors, *Ann. Probab.* **18**, pp. 405-430.
36. Rosinski, J. (2001) Series representation of Lévy processes from the perspective of point processes. In Lévy processes, theory and applications, O.E. Barndorff-Nielsen, T. Mikosch and S.I. Resnick editors, Birkhäuser, 2001, pp. 401-415.
37. Rudelson, M. (1999) Almost orthogonal submatrices of an orthogonal matrix. *Israel J. Math.* **111**, pp. 143-155.
38. Schechman, G. (2004) Two remarks concerning embedding subsets of Euclidean space into Banach spaces, manuscript.
39. Shor, P. W.; Yukich, J. E. (1991) Minimax grid matching and empirical measures. *Ann. Probab.* **19**, no. 3, pp. 1338-1348.
40. Shor, P.W. (1991) How to pack better than Best Fit: Tight bounds for average-case on-line bin packing, *Proc. 32nd Annual Symposium on Foundations of Computer Science*, pp. 752-759.
41. Talagrand, M. (1987) Regularity of Gaussian processes. *Acta Math.* **159**, pp. 99-149.

42. Talagrand, M. (1988) The structure of sign-invariant *GB*-sets, and of certain Gaussian measures. *Ann. Probab.* **16**, pp. 172-179.
43. Talagrand, M. (1988) Necessary conditions for sample boundedness of p-stable processes. *Ann. Probab.* **16**, pp. 1584-1595.
44. Talagrand, M. (1988) Donsker classes of sets. *Probab. Theo. Relat. Fields* **78**, pp. 179-191.
45. Talagrand, M. (1989) On subsets of L^p and p-stable processes. *Ann. Inst. H. Poincaré* **25**, pp. 153-166.
46. Talagrand, M. (1990) Sample boundedness of stochastic processes under increment conditions. *Ann. Probab.* **18**, pp. 1-49.
47. Talagrand, M. (1991) A new isoperimetric inequality for product measure and the concentration of measure phenomenon. *Israel Seminar (GAFA)*, Springer Verlag Lecture Notes in Math. 1469, pp. 94-124.
48. Talagrand, M. (1992) Cotype of Operators from $C(K)$. *Invent. Math.* **107**, pp. 1-40.
49. Talagrand, M. (1992) Necessary and sufficient conditions for sample continuity of Random Fourier series and of harmonic infinitely divisible processes. *Ann. Probab.* **20**, pp. 1-28.
50. Talagrand, M. (1992) Sudakov-type minoration for Gaussian chaos. *Israël J. Math* **79**, pp. 207-224.
51. Talagrand, M. (1992) The Ajtai-Komlos-Tusnady matching theorem for general measures. *Probability in a Banach space* **8**, Progress in Probability, Birkhäusser, pp. 39-54.
52. Talagrand, M. (1992) A simple proof of the majorizing measure theorem. *Geometric and Functional analysis* **2**, pp. 119-125.
53. Talagrand, M. (1993) Regularity of infinitely divisible processes. *Ann. Probab.* **21**, pp. 362-432.
54. Talagrand, M. (1994) The supremum of certain canonical processes. *Amer. J. Math.* **116**, pp. 283-325.
55. Talagrand, M. (1994) Matching theorems and discrepency computations using majorizing measures. *J. Amer. Math. Soc.* **7**, pp. 455-537.
56. Talagrand, M. (1994) Constructions of Majorizing measures Bernoulli processes and cotype. *Geometric and Functional Analysis* **4**, pp. 660-717.
57. Talagrand, M. (1944) The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.* **22**, pp. 919-959.
58. Talagrand, M. (1994) A remark on Sudakov minoration for chaos. *Probability in Banach Spaces* **9**, Birkhäuser, pp. 341-346.
59. Talagrand, M. (1995) Embedding subspaces of L_p into ℓ_p^n. *Operator theory: Advances and applications* **77**, Birkhäuser, pp. 311-326.
60. Talagrand, M. (1995) Sections of smooth convex bodies via majorizing measures. *Acta. Math* **175**, pp. 273-306.
61. Talagrand, M. (1995) Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques I.H.E.S. **81**, pp. 73-205.
62. Talagrand, M. (1996) Majorizing measures: The generic chaining. *Ann. Probab.* **24**, pp. 1049-1103.
63. Talagrand, M. (1996) The Glivenko-Cantelli problem, ten years later *J. Th. Probab.* **9**, pp. 371-384.
64. Talagrand, M. (1998) A general lower bound for Gaussian chaos. *J. Th. Probab.* **11**, pp. 157-179.
65. Talagrand, M. (1999) Selecting a proportion of characters. *Israel J. Math* **108**, pp. 275-284.
66. Talagrand, M. (2001) Majorizing measures without measures. *Ann. Probab.* **29**, pp. 411-417.

67. Talagrand, M. (2002) Gaussian averages, Bernoulli averages, and Gibbs' measures. Random structures and algorithms (Poznan, 2001). *Random Structures and Algorithms* 21, no. 3-4, pp. 197-204.
68. Talagrand, M. (2003) Vapnik-Chervonenkis type conditions and uniform Donsker classes of functions. *Ann. Probab.* **31**, no. 2, pp. 1565-1582.
69. Talagrand, M. (2004) Selector processes on classes of sets. *Probab. Theo. Relat. Fields*, to appear.
70. Talagrand, M; Yukich, J. (1993) The integrability of the square exponential transportation cost. *Ann. Applied Probab.* **3**, pp. 1100-1111.
71. Tomczak-Jaegermann, N. (1989) Banach-Mazur distances and finite-dimensional operator ideals. *Pitman monographs and surveys in pure and applied mathematics* **38**, J. Wiley, New York
72. Weber, M. (1994) GB and GC sets in ergodic theory, Probability in Banach spaces, 9 (Sandjberg, 1993) *Progr. Probab.* **35**, Birkhäuser Boston, Boston, MA. pp. 129–151.
73. Weber, M. (1996) Coupling of the GB set property for ergodic averages, *J. Theoret. Probab.***9**, no. 1, pp. 105–112.
74. Weber, M. (1998) Entropie métrique et convergence presque partout, *Travaux en Cours* **58**. Hermann, Paris, vi+151 pp.
75. Weber, M. (2000) Some theorems related to almost sure convergence of orthogonal series. *Indag. Math. (N.S.)* **11**, no. 2, pp. 293-311.
76. Weber, M. (2000) Estimating random polynomials by means of metric entropy methods. *Math. Inequal. Appl.* **3**, no. 3, pp. 443-457.
77. Weber, M, (2001) Sur le caractère gaussien de la convergence presque partout, *Fund. Math.* **167**, no. 1, pp. 23-54.
78. Weber, M, (2003) Uniform bounds under increment conditions, *Trans. Amer. Math. Soc.*, to appear.
79. Zvavitch, A. (2000) More on embedding subspaces of L_p into l_p^N, $0 < p < 1$. Geometric aspects of functional analysis, *Lecture Notes in Math.* **1745**, Springer-Verlag, Berlin, pp. 269-280.

Index

Springer Monographs in Mathematics

This series publishes advanced monographs giving well-written presentations of the "state-of-the-art" in fields of mathematical research that have acquired the maturity needed for such a treatment. They are sufficiently self-contained to be accessible to more than just the intimate specialists of the subject, and sufficiently comprehensive to remain valuable references for many years. Besides the current state of knowledge in its field, an SMM volume should also describe its relevance to and interaction with neighbouring fields of mathematics, and give pointers to future directions of research.

Abhyankar, S.S. **Resolution of Singularities of Embedded Algebraic Surfaces** 2nd enlarged ed. 1998
Alexandrov, A.D. **Convex Polyhedra** 2005
Andrievskii, V.V.; Blatt, H.-P. **Discrepancy of Signed Measures and Polynomial Approximation** 2002
Angell, T.S.; Kirsch, A. **Optimization Methods in Electromagnetic Radiation** 2004
Ara, P.; Mathieu, M. **Local Multipliers of C*-Algebras** 2003
Armitage, D.H.; Gardiner, S.J. **Classical Potential Theory** 2001
Arnold, L. **Random Dynamical Systems** corr. 2nd printing 2003 (1st ed. 1998)
Arveson, W. **Noncommutative Dynamics and E-Semigroups** 2003
Aubin, T. **Some Nonlinear Problems in Riemannian Geometry** 1998
Auslender, A.; Teboulle M. **Asymptotic Cones and Functions in Optimization and Variational Inequalities** 2003
Bang-Jensen, J.; Gutin, G. **Digraphs** 2001
Baues, H.-J. **Combinatorial Foundation of Homology and Homotopy** 1999
Brown, K.S. **Buildings** 3rd printing 2000 (1st ed. 1998)
Cherry, W.; Ye, Z. **Nevanlinna's Theory of Value Distribution** 2001
Ching, W.K. **Iterative Methods for Queuing and Manufacturing Systems** 2001
Crabb, M.C.; James, I.M. **Fibrewise Homotopy Theory** 1998
Chudinovich, I. **Variational and Potential Methods for a Class of Linear Hyperbolic Evolutionary Processes** 2005
Dineen, S. **Complex Analysis on Infinite Dimensional Spaces** 1999
Dugundji, J.; Granas, A. **Fixed Point Theory** 2003
Elstrodt, J.; Grunewald, F. Mennicke, J. **Groups Acting on Hyperbolic Space** 1998
Edmunds, D.E.; Evans, W.D. **Hardy Operators, Function Spaces and Embeddings** 2004
Fadell, E.R.; Husseini, S.Y. **Geometry and Topology of Configuration Spaces** 2001
Fedorov, Y.N.; Kozlov, V.V. **A Memoir on Integrable Systems** 2001
Flenner, H.; O'Carroll, L. Vogel, W. **Joins and Intersections** 1999
Gelfand, S.I.; Manin, Y.I. **Methods of Homological Algebra** 2nd ed. 2003
Griess, R.L.Jr. **Twelve Sporadic Groups** 1998
Gras, G. **Class Field Theory** corr. 2nd printing 2005
Hida, H. ***p*-Adic Automorphic Forms on Shimura Varieties** 2004
Ischebeck, F.; Rao, R.A. **Ideals and Reality** 2005
Ivrii, V. **Microlocal Analysis and Precise Spectral Asymptotics** 1998
Jech, T. **Set Theory** (3rd revised edition 2002)
Jorgenson, J.; Lang, S. **Spherical Inversion on SLn (R)** 2001
Kanamori, A. **The Higher Infinite** corr. 2nd printing 2005 (2nd ed. 2003)
Kanovei, V. **Nonstandard Analysis, Axiomatically** 2005
Khoshnevisan, D. **Multiparameter Processes** 2002
Koch, H. **Galois Theory of *p*-Extensions** 2002
Kozlov, V.; Maz'ya, V. **Differential Equations with Operator Coefficients** 1999
Landsman, N.P. **Mathematical Topics between Classical & Quantum Mechanics** 1998
Leach, J.A.; Needham, D.J. **Matched Asymptotic Expansions in Reaction-Diffusion Theory** 2004
Lebedev, L.P.; Vorovich, I.I. **Functional Analysis in Mechanics** 2002
Lemmermeyer, F. **Reciprocity Laws: From Euler to Eisenstein** 2000
Malle, G.; Matzat, B.H. **Inverse Galois Theory** 1999
Mardesic, S. **Strong Shape and Homology** 2000

Margulis, G.A. **On Some Aspects of the Theory of Anosov Systems** 2004
Murdock, J. **Normal Forms and Unfoldings for Local Dynamical Systems** 2002
Narkiewicz, W. **Elementary and Analytic Theory of Algebraic Numbers** 3rd ed. 2004
Narkiewicz, W. **The Development of Prime Number Theory** 2000
Parker, C.; Rowley, P. **Symplectic Amalgams** 2002
Peller, V. (Ed.) **Hankel Operators and Their Applications** 2003
Prestel, A.; Delzell, C.N. **Positive Polynomials** 2001
Puig, L. **Blocks of Finite Groups** 2002
Ranicki, A. **High-dimensional Knot Theory** 1998
Ribenboim, P. **The Theory of Classical Valuations** 1999
Rowe, E.G.P. **Geometrical Physics in Minkowski Spacetime** 2001
Rudyak, Y.B. **On Thom Spectra, Orientability and Cobordism** 1998
Ryan, R.A. **Introduction to Tensor Products of Banach Spaces** 2002
Saranen, J.; Vainikko, G. **Periodic Integral and Pseudodifferential Equations with Numerical Approximation** 2002
Schneider, P. **Nonarchimedean Functional Analysis** 2002
Serre, J-P. **Complex Semisimple Lie Algebras** 2001 (reprint of first ed. 1987)
Serre, J-P. **Galois Cohomology** corr. 2nd printing 2002 (1st ed. 1997)
Serre, J-P. **Local Algebr**a 2000
Serre, J-P. **Trees** corr. 2nd printing 2003 (1st ed. 1980)
Smirnov, E. **Hausdorff Spectra in Functional Analysis** 2002
Springer, T.A. Veldkamp, F.D. **Octonions, Jordan Algebras, and Exceptional Groups** 2000
Sznitman, A.-S. **Brownian Motion, Obstacles and Random Media** 1998
Taira, K. **Semigroups, Boundary Value Problems and Markov Processes** 2003
Talagrand, M. **The Generic Chaining** 2005
Tits, J.; Weiss, R.M. **Moufang Polygons** 2002
Uchiyama, A. **Hardy Spaces on the Euclidean Space** 2001
Üstünel, A.-S.; Zakai, M. **Transformation of Measure on Wiener Space** 2000
Yang, Y. **Solitons in Field Theory and Nonlinear Analysis** 2001