Michel TALAGRAND

Upper and Lower Bounds for Stochastic Processes

Decomposition Theorems

June 8, 2021

Springer-Verlag Berlin Heidelberg NewYork London Paris Tokyo HongKong Barcelona Budapest Dedicated to the memory of Xavier Fernique.



Fig. -1.1. Xavier Fernique (1934-2020), by Pierre Fernique.

Contents

0.	Intr	oducti	ion	1
1.	$\mathbf{W}\mathbf{h}$	at is tl	his Book About?	5
	1.1	Philos	ophy	5
	1.2	What	is Chaining?	5
	1.3	The K	olmogorov Conditions	6
	1.4	Chaini	ing in a Metric Space: Dudley's Bound	9
	1.5	Overal	ll Plan of the Book	12
	1.6	Does t	his Book Contain any Ideas?	12
	1.7	Overv	iew by Chapters	13
		1.7.1	Gaussian Processes and the Generic Chaining	13
		1.7.2	Trees and Other Measures of Size	15
		1.7.3	Matching Theorems	15
		1.7.4	Warming up With <i>p</i> -stable Processes	16
		1.7.5	Bernoulli Processes	16
		1.7.6	Random Fourier Series and Trigonometric Sums	17
		1.7.7	Partition Scheme for Families of Distances	17
		1.7.8	Peaky parts of Functions	18
		1.7.9	Proof of the Bernoulli Conjecture	18
		1.7.10	Random Series of Functions	18
		1.7.11	Infinitely Divisible Processes	18
		1.7.12	Unfulfilled Dreams	19
		1.7.13	Empirical Processes	19
		1.7.14	Gaussian Chaos	19
		1.7.15	Convergence of Orthogonal Series; Majorizing Measures	19
		1.7.16	Shor's Matching Theorem	20
		1.7.17	The Ultimate Matching Theorem in Dimension Three .	20
		1.7.18	Applications to Banach Space Theory	20

Part I. The Generic Chaining

IV Contents

2.	Gau	ssian Processes and the Generic Chaining	23
	2.1	Overview	23
	2.2	Measuring the Size of the Supremum	23
	2.3	The Union Bound and Other Basic Facts	24
	2.4	The Generic Chaining	29
	2.5	Entropy Numbers	33
	2.6	Rolling Up our Sleeves: Chaining in the Simplex	38
	2.7	Admissible Sequences of Partitions	41
	2.8	Functionals	45
	2.9	Partitioning Schemes	48
	2.10	Gaussian Processes: The Majorizing Measure Theorem	57
	2.11	Gaussian Processes as Subsets of a Hilbert Space	61
	2.12	Dreams	68
	2.13	A First Look at Ellipsoids	70
	2.14	Rolling Up our Sleeves: Chaining on Ellipsoids	75
	2.15	Continuity of Gaussian Processes	77
	2.16	Notes and Comments	81
3.	Tr€	ees and Other Measures of Size	83
	3.1	Trees	83
		3.1.1 Separated Trees	84
		3.1.2 Organized trees	85
		3.1.3 Majorizing Measures	88
	3.2	Rolling Up our Sleeves: Trees in Ellipsoids	91
	3.3	Fernique's Functional	94
		3.3.1 Fernique's Functional	94
		3.3.2 Fernique's Convexity Argument	95
		3.3.3 From Majorizing Measures to Sequences of Partitions .	97
	3.4	Witnessing Measures	99
	3.5	An inequality of Fernique	100
4	Mat	taking Theorems	105
4.	1 1 1	The Ellipsoid Theorem	105
	1.1 1.2	Partitioning Scheme II	110
	1.2 1.3	Matchings	113
	4.0 4.4	Discrepancy Bounds	116
	45	The Aitai-Komlós-Tusnády Matching Theorem	117
	1.0	4.5.1 The Long and Instructive Way	110
		4.5.2 The Short and Magic Way	128
	4.6	Lower Bound for the Aitai-Komlós-Tusnády Theorem	131
	$\frac{1.0}{4.7}$	The Leighton-Shor Grid Matching Theorem	138
	4.8	Lower Bound for the Leighton-Shor Theorem	145
	4.0 1 0	For the Expert Only	148
	4.10	Notes and Comments	151
	7,10		τUΤ

Par	t II.	Some Dreams Come True	
5.	War 5.1 5.2 5.3 5.4 5.5 5.6	rming Up With p-Stable Processesp-Stable Processes as Conditionally Gaussian ProcessesA Lower Bound for p-Stable ProcessesPhilosophySimplification Through Abstraction1-Stable ProcessesWhere Do We Stand?	$155 \\ 155 \\ 156 \\ 157 \\ 158 \\ 161 \\ 162$
6.	Ber: 6.1 6.2 6.3 6.4 6.5 6.6 6.7 6.8 6.9	noulli ProcessesBernoulli r.v.sBoundedness of Bernoulli ProcessesConcentration of MeasureSudakov MinorationComparison PrincipleControl in ℓ^{∞} NormPeaky Parts of FunctionsDiscrepancy Bounds for Empirical ProcessesNotes and Comments	$163 \\ 163 \\ 164 \\ 167 \\ 169 \\ 174 \\ 175 \\ 178 \\ 183 \\ 186$
7.	Ram 7.1 7.2 7.3 7.4 7.5	dom Fourier Series and Trigonometric SumsTranslation Invariant DistancesBasics7.2.1 Simplification Through Abstraction7.2.2 Setting7.2.3 Upper Bounds in the Bernoulli case7.2.4 Lower Bounds in the Gaussian CaseRandom Distances7.3.1 Basic Principles7.3.2 A General Upper Bound7.3.3 A Side StoryThe Marcus-Pisier Theorem7.4.1 The Marcus-Pisier Theorem7.4.2 Applications of the Marcus-Pisier Theorem7.5.1 General Setting7.5.2 Families of Distances7.5.3 Lower Bounds7.5.4 Upper Bounds7.5.5 Highlighting the Magic of Theorem 7.5.57.5.6 Combining Upper and Lower Bounds7.5.7 An Example: Tails in u^{-p}	187 189 192 192 193 194 195 195 198 200 202 202 204 206 208 210 213 213 215 215

VI Contents

		7.5.9 Convergence	218
	7.6	A Primer on Random Sets	219
	7.7	Proofs, Lower Bounds	221
	7.8	Proofs, Upper Bounds	227
		7.8.1 Road Map	227
		7.8.2 A Key Step	227
		7.8.3 Road Map: An overview of Decomposition Theorems	229
		7.8.4 Decomposition Theorem in the Bernoulli Case	230
		7.8.5 Upper Bounds in the Bernoulli case	233
		7.8.6 The Main Upper Bound	234
		7.8.7 Sums With Few Non-Zero Terms	237
	7.9	Proof of the Decomposition Theorem	241
		7.9.1 Constructing Decompositions	241
		7.9.2 Proof of Proposition 7.9.1	243
	7.10	Proofs, Convergence	246
	7.11	Further Proofs	250
		7.11.1 Alternate Proof of Proposition 7.5.13	250
		7.11.2 Proof of Theorem 7.5.18	254
	7.12	Explicit Computations	254
	7.13	Vector-valued Series: A Theorem of Fernique	259
	7.14	Notes and Comments	263
_	_		
8.	Part	titioning Scheme and Families of Distances	265
	8.1	The Partitioning Scheme	265
	8.2	Tail Inequalities	270
	8.3	The Structure of Certain Canonical Processes	277
0	Doal	ky Part of Functions	285
3.	0 1	Road Man	285
	9.1	Peaky Part of Functions II	285
	9.2 0.3	Philosophy	200
	9.5	Chaining for Bornoulli Processos	295
	9.4 0.5	Notes and Comments	200
	5.0		250
10.	Pro	of of the Bernoulli Conjecture	299
	10.1	Latała's Principle	300
	10.2	Philosophy, I	304
	10.3	Chopping Maps and Functionals	304
		10.3.1 Chopping Maps	304
		10.3.2 Basic Facts	309
		10.3.3 Functionals	312
	10.4	Philosophy, II	315
	10.5	Latała's Step	316
	10.6	Philosophy, III	319
	10.7	A Decomposition Lemma	320
		-	

	10.8 Building the Partitions10.9 Philosophy, IV	$323 \\ 327$
	10.10The Key Inequality	328
	10.11Philosophy, V	333
	10.12Proof of the Latała-Bednorz Theorem	334
	10.13Philosophy, VI	337
	10.14A Geometric Characterisation of $b(T)$	337
	10.15Lower Bounds from Measures	340
	10.16Notes and Comments	343
11.	Bandom Series of Functions	345
тт.	11.1 Road Man	345
	11.2 Random Sories of Functions: Conoral Sotting	340
	11.2 Arandom Series of Functions. General Setting	240
	11.5 Organization of the Chapter	040 940
	11.4 The Main Lemma	349 951
	11.5 Construction of the Majorizing Measure Using Convexity	301
	11.6 From Majorizing Measures to Partitions	353
	11.7 The General Lower Bound	355
	11.8 The Gine-Zinn Inequalities	350
	11.9 Proof of The Decomposition Theorem for Empirical Processes	358
	11.101 he Decomposition Theorem for Random Series	360
	11.11Selector Processes, and Why They Matter	362
	11.12Proving the Generalized Bernoulli Conjecture	363
	11.13Notes and Comments	367
12.	Infinitely Divisible Processes	369
	12.1 Poisson r.v.s and Poisson Point Processes	369
	12.2 A Shortcut to Infinitely Divisible Processes	371
	12.3 Overview of Results	373
	12.3.1 The Main Lower Bound	373
	12.3.2 The Decomposition Theorem for Infinitely Divisible	
	Processes	374
	12.3.3 Upper Bounds Through Bracketing	376
	12.3.4 Harmonizable Infinitely Divisible Processes	376
	12.3.5 Example: Harmonizable <i>p</i> -Stable Processes	378
	12.4 Proofs: The Bracketing Theorem	379
	12.5 Proofs: The Decomposition Theorem for Infinitely Divisible	
	Processes	380
	12.6 Notes and Comments	383
13.	Unfulfilled Dreams	385
	13.1 Positive Selector Processes	385
	13.2 Explicitly Small Events	386
	13.3 My Lifetime Favorite Problem	388
	13.4 Classes of Sets	300
		000

 \mathbf{VII}

Contents

Part III. Practicing

14.	Empirical Processes, II	399
	14.1 Bracketing	399
	14.2 The Class of Squares of a Given Class	401
	14.3 When Not to Use Chaining	412
15	Gaussian Chaos	491
10.	15.1 Order Two Gaussian Chaos	421
	15.1 1 Basic Facts	421
	15.1.2 When T is Small for the Distance d_{12}	425
	15.1.2 Covering Numbers	431
	15.1.4 Another Way to Bound $S(T)$	432
	15.1.5 Yet Another Way to Bound $S(T)$	434
	15.2 Tails of Multiple Order Gaussian Chaos	436
	15.3 Notes and Comments	452
16.	Convergence of Orthogonal Series; Majorizing Measures	453
	16.1 A Kind of Prologue: Chaining in a Metric Space and Pisier's	
	Bound	453
	16.2 Introduction to Orthogonal Series; Paszkiewicz's Theorem	455
	16.3 Recovering the Classical Results	457
	16.4 Approach to Paszkiewicz's Theorem; Bednorz's Theorem	460
	16.5 Chaining, I	465
	16.6 Proof of Bednorz's Theorem	468
	16.7 Permutations	477
	16.8 Chaining, II	486
	16.9 Chaining, III	500
	16.10Notes and Comments	502
17.	Shor's Matching Theorem	505
	17.1 Introduction	505
	17.2 The Discrepancy Theorem	506
	17.3 Lethal Weakness of the Approach	512
19	The Illtimate Matching Theorem in Dimension Three	517
10.	18.1 Introduction	$517 \\ 517$
	18.2 Regularization of α	521
	18.3 Discretization	$521 \\ 523$
	18.4 Discrepancy Bound	$520 \\ 524$
	18.5 Geometry	528
	18.6 Probability. I	535
	18.7 Haar Basis Expansion	541
	18.8 Probability, II	546
	v /	-

	Contents	IX
	18.9 Final Effort	548
19.	$\begin{array}{l} \textbf{Applications to Banach Space Theory} \\ 19.1 \ \mbox{Cotype of Operators} \\ 19.1.1 \ \mbox{Basic Definitions} \\ 19.1.2 \ \ \mbox{Operators From } \ell_N^\infty \\ 19.1.3 \ \ \mbox{Computing the Cotype-2 Constant with Few Vectors} \\ 19.2 \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	555 555 557 559 566 566 569 574 578 578 578 578 587 587 593
А.	Discrepancy for Convex SetsA.1 IntroductionA.2 Elements of Proof of the Upper BoundA.3 The Lower Bound	601 601 601 602
В.	Some Deterministic ArgumentsB.1Hall's Matching TheoremB.2Proof of Lemma 4.7.11B.3The Shor-Leighton Grid Matching TheoremB.4End of Proof of Theorem 17.2.1B.5Proof of Proposition 17.3.1B.6Proof of Proposition 17.2.4	607 607 609 610 615 617 619
с.	Classical View of Infinitely Divisible ProcessesC.1Infinitely Divisible Random VariablesC.2Infinitely Divisible ProcessesC.3RepresentationC.4p-Stable Processes	623 623 624 625 626
D.	Reading SuggestionsD.1Partition SchemesD.2Geometry of Metric SpacesD.3Cover TimesD.4MatchingsD.5Super-concentration in the Sense of S. ChatterjeeD.6High-Dimensional Statistics	627 627 627 628 628 628 628 629

X Contents

Е.	Research Directions	
	E.1 The Latała-Bednorz Theorem	
	E.2 The Ultimate Matching Conjecture	
	E.3 My Favorite Life-Time Problem	
	E.4 From a Set to its Convex Hull	
F.	Solutions of Selected Exercises	
G.	Comparison with the First Edition	
References		

0. Introduction

This book had a previous edition [171]. The changes between the two editions are not only cosmetic or pedagogical, and the degree of improvement in the mathematics themselves is almost embarrassing at times. Besides significant simplifications in the arguments, several of the main conjectures of [171] have been solved and a new direction came to fruition. It would have been more appropriate to publish this text as a brand new book, but the improvements occurred gradually and the bureaucratic constraints of the editor did not allow a change at a late stage without further delay and uncertainty.

We first explain in broad terms the contents of this book, and then we detail some of the changes from [171].

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

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

INEQUALITIES.

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

2 0. Introduction

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

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

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

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

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

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

This book is in a sense a continuation of the monograph [75], or at least of part of it. I made no attempt to cover again all the relevant material of

3

[75], but familiarity with [75] is certainly not a prerequisite, and maybe not even helpful. The way certain results are presented there is arguably obsolete, and, more importantly, many of the problems considered in [75] (in particular, limit theorems) are no longer the focus of much interest.

One of my main goals is to communicate as much as possible of my experience from working on stochastic processes, and I have covered most of my results in this area. A number of these results were proved many years ago. I still like them, but some seem to be waiting for their first reader. The odds of these results meeting this first reader while staying buried in the original papers seemed nil, but they might increase in the present book form. In order to present a somewhat coherent body of work, I have also included rather recent results by others in the same general direction.¹ I find these results deep and very beautiful. They are sometimes difficult to access for the nonspecialist. Explaining them here in a unified and often simplified presentation could serve a useful purpose. Still, the choice of topics is highly personal and does not represent a systematic effort to cover all the important directions. I can only hope that the book contains enough state-of-art knowledge about sufficiently many fundamental questions to be useful.

Let me now try to outline the progress since the previous edition.² While attempting to explain better my results to others, I ended up understanding them much better myself. The material of the previous edition was reduced by about 100 pages due to better proofs.³ More importantly, reexamination of the material resulted in new methods, and a new direction came to fruition, that of

DECOMPOSITION THEOREMS.

The basic idea is that there are two fundamentally different way to control the size of a sum $\sum_{i \leq N} X_i$. One may take advantage of cancellations between terms, or one may bound the sum by the sum of the absolute values. One may also *interpolate* between the two methods, which in that case means writing a decomposition $X_i = X'_i + X''_i$ and controlling the size of the sum $\sum_{i \leq N} X'_i$ by taking advantage of the cancellations between terms, but controlling the sum $\sum_{i \leq N} X''_i$ by the sum of the absolute values. The same schoolboy idea, in the setting of stochastic processes, is that a process can be bounded on the one hand using chaining, and on the other hand can often be bounded by cruder methods, involving replacing certain sums by the sums of the absolute values. The amazing fact is that many processes can by controlled by interpolating between these two methods, that is can be decomposed into the sum two pieces, each of which can be controlled by one of these methods.

¹ With one single exception I did not include results by others proved after the first edition of this book.

 $^{^2}$ A precise comparison between the two editions may be found in Appendix G.

³ A limited quantity of material of secondary importance was also removed. The current edition is not shorter than the previous one because many details have been added, as well as an entire chapter on the new results, and a sketch of proof for many exercises.

4 0. Introduction

Such is the nature of the landmark result of Bednorz and Latała [23], the proof of the Bernoulli conjecture, which is the towering result of this book. Several conjectures of [171] in the same general directions have been solved⁴ concerning in particular empirical processes and random series of functions.

Despite the considerable progress represented by the solution of these conjectures, a number of seemingly important questions remain open, and one of my main goals is to popularize these. Opinions differ as to what constitutes a really important problem, but I like those I explain here because they deal with fundamental structures. These problems might be challenging. At least, I tried my best to make progress on them, but they have seen little progress and received little attention.

I would like to express my infinite gratitude to Shahar Mendelson. While he was donating his time to help another of my projects, it became clear through our interactions that, while I had produced great efforts towards the quality of the mathematics contained in my books, I certainly had not put enough efforts in the exposition of this material. I concluded that there should be real room for improvement in the text of [171], and this is why I started to revise it, and this led to the major advances presented here.

While preparing the current text I have been helped by a number of people. I would like to thank some of them here (and to appologize to all those whom I do not mention). Ramon van Handel suggested a few almost embarrassing simplifications. Hengrui Luo and Zhenyuan Zhang suggested literately hundreds of improvements, and Rafał Meller's comments had a great impact too. Further luck had it that, almost at the last minute, my text attracted the attention of Kevin Tanguy whose efforts resulted in a higher level of detail and a gentler pace of exposition. In particular his and Zhang's efforts gave me the energy to make a fresh attempt at explaining and detailing the proof of the Bernoulli conjecture obtained by Bednorz and Latala in [23]. This proof is the most stupendously beautiful piece of mathematics I have met in my entire life. I wish the power of this result and the beauty of this proof become better understood.

I dedicate this work to the memory of Xavier Fernique. Fernique was a deeply original thinker. His groundbreaking contributions to the theory of Gaussian processes were viewed as exotic by mainstream probabilists, and he never got the recognition he deserved. I owe a great debt to Fernique: it is his work on Gaussian processes which made my own work possible, first on Gaussian processes, and then on all the situations beyond this case. This work occupied many of my most fruitful years. A large part of it is presented in the present volume. It would not have existed without Fernique's breakthroughs.

⁴ After another crucial contribution of Bednorz and Martynek [25].

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

1.1 Philosophy

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

The first advice I received from my advisor Gustave Choquet was as follows: always consider a problem under the minimum structure in which it makes sense. This advice has literally shaped my mathematical life. It will probably be as fruitful in the future as it has been in the past. By following it, one is naturally led to study problems with a kind of minimal and intrinsic structure. Not so many structures are really basic, and one may hope that these will remain of interest for a very long time. This book is devoted to the study of such structures which arise when one tries to estimate the suprema of stochastic processes.

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

1.2 What is Chaining?

A stochastic process is a collection of random variables (r.v.s) $(X_t)_{t \in T}$ indexed by a set T. To study it, Kolmogorov invented chaining, the main tool of this book. The fundamental idea of chaining is to replace the index set T

by a sequence of finite approximations T_n , and to study the r.v.s X_t through successive approximations $X_{\pi_n(t)}$ where $\pi_n(t) \in T_n$. The first approximation consists of a single point t_0 so $T_0 = \{t_0\}$. The fundamental relation is then

$$X_t - X_{t_0} = \sum_{n \ge 1} (X_{\pi_n(t)} - X_{\pi_{n-1}(t)}) .$$
(1.1)

When T is finite, the only case we really need, the sum on the right is finite. This relation gives its name to the method: the chain of approximations $\pi_n(t)$ links t_0 and t. To control the differences $X_t - X_{t_0}$ it suffices then to control all the differences $|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}|$.

1.3 The Kolmogorov Conditions

Kolmogorov stated the "Kolmogorov conditions", which robustly ensure the good behavior of a stochastic process indexed by a subset of \mathbb{R}^m . These conditions are studied in any advanced probability course. If you have taken such a course, this section will refresh your memory about these conditions, and the next few sections will present the natural generalization of the chaining method in an abstract metric space, as it was understood in, say, 1970. Learning in detail about these historical developments now makes sense only if you have already heard of them, because the modern chaining method, which is presented in Chapter 2 is in a sense far simpler than the classical method. For this reason, the material up to Section 1.4 included is directed towards a reader who is already fluent in probability theory. If, on the other hand, you have never heard of these things, and if you find this material too difficult, you should start directly with Chapter 2, which is written at a far greater level of detail and assumes minimal familiarity with even basic probability theory.

We say that a process $(X_t)_{t\in T}$, where $T = [0,1]^m$, satisfies the Kolmogorov conditions if

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

where d(s,t) denotes the Euclidean distance and $p > 0, \alpha > m$. Here E denotes mathematical expectation. In our notation the operator E applies to whatever expression is placed behind it, so that $E|Y|^p$ stands for $E(|Y|^p)$ and not for $(E|Y|)^p$. This convention is in force throughout the book.

Let us apply the idea of chaining to processes satisfying the Kolmogorov conditions. The most obvious candidate for the approximating set T_n is the set G_n of points x in $[0,1]^m$ such that the coordinates of $2^n x$ are positive integers.¹ Thus card $G_n = 2^{nm}$. It is completely natural to choose

¹ There is no other reason for using the points x in $[0, 1]^m$ such that the coordinates of $2^n x$ are positive integers rather than the points x in $[0, 1]^m$ with the same property than the fact that there are 2^{nm} such points rather than the typographically unpleasant number $(2^n + 1)^m$.

 $\overline{7}$

 $\pi_n(u) \in G_n$ as close to u as possible, so that $d(u, \pi_n(u)) \leq \sqrt{m}2^{-n}$, and $d(\pi_n(u), \pi_{n-1}(u)) \leq d(\pi_n(u), u) + d(u, \pi_{n-1}(u)) \leq 3\sqrt{m}2^{-n}$.

For $n \ge 1$ let us then define

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

Given $s = (s_1, \ldots, s_m) \in G_n$, the number of points $t = (t_1, \ldots, t_m) \in G_n$ with $d(s,t) \leq 3\sqrt{m2^{-n}}$ is bounded independently of s and n because $|t_i - s_i| \leq d(s,t)$ for each $i \leq m$, so that we have the crucial property

$$\operatorname{card} U_n \le K(m)2^{nm} , \qquad (1.4)$$

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

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

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

$$|X_{\pi_n(u)} - X_{\pi_{n-1}(u)}| \le Y_n .$$
(1.6)

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

$$\sup_{t \in G \ ; \ d(s,t) \le 2^{-k}} |X_s - X_t| \le 3 \sum_{n \ge k} Y_n \ . \tag{1.7}$$

To prove this consider $n \ge k$ such that $s, t \in G_n$, so that $s = \pi_n(s)$ and $t = \pi_n(t)$. Assuming $d(s,t) \le 2^{-k}$, we have

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

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

s

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

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

$$X_u - X_{\pi_k(u)} = X_{\pi_n(u)} - X_{\pi_k(u)} = \sum_{k \le \ell < n} X_{\pi_{\ell+1}(u)} - X_{\pi_\ell(u)} ,$$

and since $|X_{\pi_{\ell+1}(u)} - X_{\pi_{\ell}(u)}| \leq Y_{\ell+1}$ we obtain $|X_u - X_{\pi_k(u)}| \leq \sum_{\ell \geq k} Y_{\ell+1}$. To obtain (1.7) we then use the previous inequalities and the identity

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

Let us now draw some consequences of (1.7). For a finite family of numbers $V_i \ge 0$, we have

$$(\max_i V_i)^p \le \sum_i V_i^p , \qquad (1.8)$$

and thus

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

since $\mathsf{E}|X_s - X_t|^p \leq K(m, \alpha)2^{-n\alpha}$ for $(s,t) \in U_n$ by (1.2) and using (1.4). To proceed one needs to distinguish whether or not $p \geq 1$. For specificity we assume $p \geq 1$. Since, as we just proved, $||Y_n||_p := (\mathsf{E}|Y_n|^p)^{1/p} \leq K(m, p, \alpha)2^{n(m-\alpha)/p}$, the triangle inequality in L^p yields²

$$\left\|\sum_{n \ge k} Y_n\right\|_p \le \sum_{n \ge k} K(m, p, \alpha) 2^{n(m-\alpha)/p} \le K(m, p, \alpha) 2^{k(m-\alpha)/p} .$$
(1.9)

Combining with (1.7) we then obtain

$$\left\| \sup_{s,t \in G; d(s,t) \le 2^{-k}} |X_s - X_t| \right\|_p \le K(m, p, \alpha) 2^{k(m-\alpha)/p} , \qquad (1.10)$$

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

$$\mathsf{E}\sup_{s,t\in G}\frac{|X_s - X_t|^p}{d(s,t)^\beta} < \infty . \tag{1.11}$$

Exercise 1.3.1. Prove (1.11). Hint: prove that

$$\sum_{k\geq 0} \mathsf{E} \sup_{s,t\in G; d(s,t)\leq 2^{-k}} 2^{k\beta} |X_s - X_t|^p < \infty .$$
(1.12)

Thus, chaining not only proves that the process $(X_t)_{t\in T}$ has a continuous version, it also provides the very good estimate (1.10). One reason for which everything is so easy in this case is that the size of the terms $X_{\pi_{n+1}(u)} - X_{\pi_n(u)}$ decreases like a geometric series.

Let us then pause for a moment and reflect on what we have been doing.

- The Euclidean metric structure of T is not really intrinsic to the problem. Far more intrinsic is the (quasi) distance on T given by $\delta(s,t) = ||X_s - X_t||_p$. The condition (1.2), which we may now write as $\delta(s,t) \leq d(s,t)^{\alpha/p}$ simply enforces a kind of "smallness condition" on the metric space (T, δ) .
- The use of the bound (1.6) is rather pessimistic, as it bounds each of the increments along the chain by the worst possible case among each increment.

 $^{^2\,}$ There of course the two occurrences of the constant $K(m,p,\alpha)$ are not the same.

These two remarks contain in germ much of the future progress we will make. Following the first remark, we will learn, starting with the next section, to look at problems in a more intrinsic manner. And our sharp chaining methods will avoid the crude bound of each increment by the worst possible case.

There are many variations on the previous ideas. The next two exercises explore one.

Exercise 1.3.2. Consider a convex function $\varphi \ge 0$ with $\varphi(0) = 0$. Prove that for r.v.s $V_i \ge 0$ one has

$$\mathsf{E}\max_{i} V_{i} \le \varphi^{-1} \left(\sum_{i} \mathsf{E}\varphi(V_{i}) \right) \,. \tag{1.13}$$

Exercise 1.3.3. Consider the function φ as above, and consider positive numbers c_n, d_n . Assume that the process $(X_t)_{t \in T}$ satisfies

$$\forall n \ge 0 , \forall s, t \in T , d(s,t) \le 3\sqrt{m}2^{-n} \Rightarrow \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{c_n}\Big) \le d_n .$$
 (1.14)

Prove that

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

The series in (1.15) has no reason to converge like a geometric series, so we already are being more sophisticated than in the case of the Kolmogorov conditions.³

1.4 Chaining in a Metric Space: Dudley's Bound

Suppose now that we want to study the uniform convergence on [0,1] of a random Fourier series $X_t = \sum_{k\geq 1} a_k g_k \cos(2\pi i k t)$ where a_k are numbers and (g_k) are independent standard Gaussian r.v.s. The Euclidean structure of [0,1] is not intrinsic to the problem. Far more relevant is the distance d given by

$$d(s,t)^{2} = \mathsf{E}(X_{s} - X_{t})^{2} = \sum_{k} a_{k}^{2} (\cos(2i\pi ks) - \cos(2i\pi kt))^{2} .$$
(1.16)

This simple idea took a very long time to emerge. Once one thinks about the distance d, then in turn the fact that the index set T is [0, 1] is no longer very

³ In the left-hand side of (1.15) we would like to do better than controlling the expectation, but one really needs some regularity of the function φ for this. It suffices here to say that when $\varphi(x) = |x|^p$ for $p \ge 1$ we may replace the expectation by the norm of L^p , proceeding exactly as we did in the case of the Kolmogorov conditions.

relevant because this particular structure does not connect very well with the distance d. One is then led to consider Gaussian processes indexed by an abstract set T.⁴ We say that $(X_t)_{t\in T}$ is a *Gaussian process* when the family $(X_t)_{t\in T}$ is jointly Gaussian centered.⁵ Then, just as in (1.16), the process induces a canonical distance d on T given by $d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2}$. We will express that Gaussian r.v.s have small tails by the inequality

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

where $\varphi(x) = \exp(x^2/4) - 1$. This inequality holds because if g is a standard Gaussian r.v. then $\mathsf{E}\exp(g^2/4) \leq 2.^6$

To perform chaining for such a process, in the absence of further structure on our metric space (T, d), how do we choose the approximating sets T_n ? Thinking back to the Kolmogorov conditions it is very natural to introduce the following definition.

Definition 1.4.1. For $\epsilon > 0$ the covering number $N(T, d, \epsilon)$ of a metric space (T, d) is the smallest integer N such that T can be covered by N balls of radius ϵ .⁷

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

Let us denote by $\Delta(T) = \sup_{s,t\in T} d(s,t)$ the diameter of T, and observe that $N(T, d, \Delta(T)) = 1$. We construct our approximating sets T_n as follows. Consider the largest integer n_0 with $\Delta(T) \leq 2^{-n_0}$. For $n \geq n_0$ consider a set $T_n \subset T$ with card $T_n = N(T, d, 2^{-n})$ such that each point of T is within distance 2^{-n} of a point of T_n .⁸ In particular T_0 consists of a single point.

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

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

so that

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

⁴ Let us stress the point. Even though the index set is a subset of \mathbb{R}^m we have no chance to really understand the process unless we forget this irrelevant structure.

⁵ Centered means that $\mathsf{E}X_t = 0$ for each t.

⁶ Starting with the next chapter we will control the r.v.s $|X_s - X_t|$ through their tail properties, and (1.17) is just another way to present the same situation.

⁷ Here our balls are closed balls. One could also use open balls in this definition. There seems to be no universal agreement about this. For our purpose it makes no difference whatsoever.

⁸ We do not require that $T_n \subset T_{n+1}$. In Section 1.3 it does happen that $G_n \subset G_{n+1}$ but this was not really used in the arguments.

This crude bound is hard to improve in general and should be compared to (1.4). We now apply (1.13) to the r.v.s $V_i = |X_s - X_t|/(3 \cdot 2^{-n})$ for $i = (s,t) \in U_n$. Since $\mathsf{E}\varphi(V_i) \leq 1$ we obtain that the r.v.

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

satisfies

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

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

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

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

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

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

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

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

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

Exercise 1.4.3. Prove that the previous bound gives the correct uniform modulus of continuity for Brownian motion on [0, 1]: for $\delta \leq 1$,

$$\mathsf{E} \sup_{|s-t| \le \delta} |B_s - B_t| \le L\sqrt{\delta \log(2/\delta)}$$

The message of Chapter 2 is simple:

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

1.5 Overall Plan of the Book

A specific feature of the index set $T = [0, 1]^m$ (provided with the Euclidean distance) occurring in the Kolmogorov conditions is that it is really "*m*-dimensional" and "the same around each point". This is not the case for index sets which occur in a great many natural situations. If one had to summarize in one sentence the content of the upper bounds presented in this book, it would be that they develop methods which are optimal even when this feature does not occur.

The main tools are built in Parts I and II. Part I is devoted to the most important situation we consider in the book, the study of Gaussian processes, and we learn the basic concepts on how to measure the "size" of a metric space. The effectiveness of the corresponding tools is then demonstrated by proving classical results on matchings.

The goal of Part II is to extend the results of the Gaussian case to other more general processes. This program of building the proper tools to go beyond the Gaussian case was started by the author soon after he obtained his results on Gaussian processes (which are presented in Chapter 2). It is a significant endeavor which requires a number of new concepts. The most important of these is the idea of families of distances. We can no longer entirely describe the situation using a single distance on the index set (as is the case for Gaussian processes). In some sense *this program has been completed*. Most of the results which were dreamed by the author⁹ between 1985 and 1990 are now proved in Chapter 11.

Part III explores situations which belong to the same circle of ideas but in diverse directions.

1.6 Does this Book Contain any Ideas?

At this stage it is not really possible to precisely describe any of the new ideas which will be presented, but if the following statements are not crystal-clear to you, you may have something to learn from this book.

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

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

Idea 3. Ellipsoids are much smaller than one would think, because they (and more generally, sufficiently convex bodies) are thin around the edges.

⁹ Including some which sounded like crazily optimistic conjectures!

This explains the funny fractional powers of logarithms in certain matching theorems.

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

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

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

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

1.7 Overview by Chapters

1.7.1 Gaussian Processes and the Generic Chaining

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

The most important question considered in this book is the boundedness of Gaussian processes. The key object is the metric space (T, d) where T is the index set and d the intrinsic distance (0.1). As investigated in Section 2.11 this metric space is far from being arbitrary: it is isometric to a subset of a Hilbert space. It is, however, a deadly trap to try to use this specific property of the metric space (T, d). The proper approach is to just think of it as a general metric space.

After reviewing some elementary facts, in Section 2.4 we explain the basic idea of the "generic chaining", one of the key ideas of this work. Chaining is a succession of steps that provide successive approximations of the index space (T, d). In the Kolmogorov chaining, for each n the difference between the n-th and the (n + 1)-th approximation of the process, which we call here "the variation of the process during the n-th chaining step", is "controlled uniformly over all possible chains". Generic chaining allows that the variation of the process during the n-th chaining step "may depend on which chain we follow". Once the argument is properly organized, it is not any more complicated than the classical argument. It is in fact exactly the same. Yet, while



Fig. 1.1. Dependence chart between chapters. Very marginal dependence is not indicated.

Dudley's classical bound is not always sharp, the bound obtained through the generic chaining is optimal. Entropy numbers are reviewed in Section 2.5.

It is technically convenient to formulate the generic chaining bound using special sequences of partitions of the metric space (T, d), that we shall call admissible sequences throughout the book. The key to make the generic chaining bound useful is then to be able to construct admissible sequences. These admissible sequences measure an aspect of the "size" of the metric space and are introduced in Section 2.7. In Section 2.8 we introduce another method to measure the "size" of the metric space, through the behavior of certain "functionals", which are simply numbers attached to each subset of the entire space. The fundamental fact is that the two measures of the size of the metric space one obtains through either admissible sequences or through functionals are equivalent in full generality. This is proved in Section 2.8 for the easy part (that the admissible sequence approach provides a larger measure of size than the functional approach) and in Section 2.9 for the converse. This converse is, in effect, an algorithm to construct sequences of partitions in a metric space given a functional. Functionals are of considerable use throughout the book. In Section 2.10 we prove that the generic bound can be reversed for Gaussian processes, therefore providing a characterization of their sampleboundedness. Generic chaining entirely explains the size of Gaussian processes, and the dream of Section 2.12 is that a similar situation will occur for many processes.

In Section 2.11 we explain why a Gaussian process in a sense *is* nothing but a subset of Hilbert space. Remarkably, a number of basic questions remain unanswered, such as how to relate through geometry the size of a subset of Hilbert space seen as a Gaussian process with the corresponding size of its convex hull.

Dudley's bound fails to explain the size of the Gaussian processes indexed by ellipsoids in Hilbert space. This is investigated in Section 2.13. Ellipsoids will play a basic role in Chapter 4.

1.7.2 Trees and Other Measures of Size

We describe different notions of trees, and show how one can measure the "size" of a metric space by the size of the largest trees it contains, in a way which is equivalent to the measures of size introduced in Chapter 2. This idea played an important part in the history of Gaussian processes. Its appeal is mostly that trees are easy to visualize. Building a large tree in a metric space is an efficient method to bound its size from below. We then learn a method of Fernique to measure the size of a metric space through certain properties of the probability measures on it. It will be amenable to vast generalizations.

1.7.3 Matching Theorems

Chapter 4 makes the point that the generic chaining (or some equivalent form of it) is already required to really understand the irregularities occurring in the distribution of N points $(X_i)_{i \leq N}$ independently and uniformly distributed in the unit square. These irregularities are measured by the "cost" of pairing (=matching) these points with N fixed points that are very uniformly spread, for various notions of cost.

These optimal results involve mysterious powers of $\log N$. We are able to trace them back to the geometry of ellipsoids in Hilbert space, so we start the chapter with an investigation of these ellipsoids in Section 4.1. The philosophy of the main result, the Ellipsoid Theorem, is that an ellipsoid is in some sense somewhat smaller than it appears at first. This is due to convexity: an ellipsoid gets "thinner" when one gets away from its center. The Ellipsoid Theorem is a special case of a more general result (with the same proof) about the structure of sufficiently convex bodies, one that will have important applications in Chapter 19.

In Section 4.3 we provide general background on matchings. In Section 4.5 we investigate the case where the cost of a matching is measured

by the average distance between paired points. We prove the result of Ajtai, Komlós, Tusnády, that the expected cost of an optimal matching is at most $L\sqrt{\log N}/\sqrt{N}$ where L is a number. The factor $1/\sqrt{N}$ is simply a scaling factor, but the fractional power of log is optimal as shown in Section 4.6. In Section 4.7 we investigate the case where the cost of a matching is measured instead by the maximal distance between paired points. We prove the theorem of Leighton and Shor that the expected cost of a matching is at most $L(\log N)^{3/4}/\sqrt{N}$, and the power of log is shown to be optimal in Section 4.8.

With the exception of Section 4.1, the results of Chapter 4 are not connected to any subsequent material before Chapter 17.

1.7.4 Warming up With *p*-stable Processes

With this chapter we start the program of vastly extending the results of Chapter 2 concerning Gaussian processes. We outline several of the fruitful methods on the class of *p*-stable processes, based on their property of being conditionally Gaussian.

1.7.5 Bernoulli Processes

Random signs are obviously important r.v.s, and occur frequently in connection with "symmetrization procedures", a very useful tool. In a Bernoulli process the individual random variables X_t are linear combinations of independent random signs. Each Bernoulli process is associated with a Gaussian process in a canonical manner, when one replaces the random signs by independent standard Gaussian r.v.s. The Bernoulli process has better tails than the corresponding Gaussian process (it is "subgaussian") and is bounded whenever the corresponding Gaussian process is bounded. There is, however, a completely different reason for which a Bernoulli process might be bounded; namely, that the sum of the absolute values of the coefficients of the random signs remain bounded independently of the index t. A natural question is then to decide whether these two extreme situations are the only fundamental reasons why a Bernoulli process can be bounded, in the sense that a suitable "mixture" of them occurs in every bounded Bernoulli process. This was the "Bernoulli Conjecture" (to be stated formally on page 167), which has been so brilliantly solved by W. Bednorz and R. Latała.

It is a long road to the solution of the Bernoulli conjecture, and we start to build the main tools bearing on Bernoulli processes. A linear combination of independent random signs looks like a Gaussian r.v. when the coefficients of the random signs are small. We can expect that a Bernoulli process will look like a Gaussian process when these coefficients are suitably small. This is a fundamental idea: the key to understanding Bernoulli processes is to reduce to situations where these coefficients are small. The Bernoulli conjecture, on which the author worked so many years, greatly influenced the way he looked at various processes. In the case of empirical processes, this is explained in Section 6.8.

1.7.6 Random Fourier Series and Trigonometric Sums

The basic example of a random Fourier series is

$$X_t = \sum_{k \ge 1} \xi_k \exp(2\pi \mathrm{i}kt) , \qquad (1.20)$$

where $i^2 = -1$, where $t \in [0, 1]$ and the r.v.s ξ_k are independent symmetric. In this chapter we provide a final answer to the question of the convergence of such series.

The fundamental case where $\xi_k = a_k g_k$ for numbers a_k and independent Gaussian r.v.s (g_k) is of great historical importance. There is, however, another motivation for the study of such series. The generic chaining and related methods are well adapted to the case of a "non-homogeneous index space". The study of certain of the processes we will consider in the next chapters is already subtle even in the absence of the extra difficulty due to this lack of homogeneity. The setting of random Fourier series allows us to put aside the issue of lack of homogeneity and to concentrate on the other difficulties, and played a great part in the development of the theory. It provides an ideal setting to understand a basic fact: many processes can be exactly controlled, not by using one or two distances, but by using an entire family of distances. This concept of "family of distances" will play a major role later. It is also while analyzing the lower bounds discovered in the setting of random Fourier series that the author discovered the method which allows to extend these bounds to general random series as explained in Chapter 11. In this chapter we also meet our first "decomposition theorem": there are two distinct reasons which explain the size of a random trigonometric sum. First, there can be a lot of cancellation between the terms. Second, it may happen that the sum of the absolute values of the terms is already small. We show that every random trigonometric sum is the sum of two such pieces, one of each type.

1.7.7 Partition Scheme for Families of Distances

Once one has survived the initial surprise of the new idea that many processes are naturally associated to an entire family of distances, it is very pleasant to realize that the tools of Section 2.9 can be extended to this setting with essentially the same proof. This is the purpose of Section 8.1.

In Section 8.3 we apply these tools to the situation of "canonical processes" where the r.v.s X_t are linear combinations of independent copies of symmetric r.v.s with density proportional to $\exp(-|x|^{\alpha})$ where $\alpha \geq 1$ (and to considerably more general situations as discovered by R. Latała). In these

situations, the size of the process can be completely described from the geometry of the index space, a far reaching extension of the Gaussian case.

1.7.8 Peaky parts of Functions

We learn how to measure the size of sets of functions on a measured space using an appropriate family of distances. We show that when we control this size, for each function of the set we can distinguish its "peaky part" in a coherent way over the whole set of functions which then has in a sense a simple structure, as it is built from simpler pieces.

1.7.9 Proof of the Bernoulli Conjecture

Having learned how to manipulate "families of distances" we are now better prepared to prove the Bernoulli conjecture. This is the (overwhelmingly important) Latała-Bednorz theorem. The challenging proof occupies most of Chapter 10.¹⁰ In the last section we investigate how to get lower bounds on Bernoulli processes using "witnessing measures".

1.7.10 Random Series of Functions

For a large class of random series of functions, we prove in full generality that chaining explains all the part of the boundedness of these processes created by cancellations, in the spirit of the Bernoulli conjecture. This covers both the cases of empirical processes and of the closely related class of selector processes. Our main tool is to reduce to processes which are conditionally Bernoulli processes and to use the Latała-Bednorz theorem and its consequences.

1.7.11 Infinitely Divisible Processes

The infinitely divisible processes we study are indexed by a general set, and are to Lévy processes what a general Gaussian process (index by an arbitrary index set) is to Brownian motion (a Gaussian process indexed by \mathbb{R} with stationary increments). We extend to these processes our results on random series of functions: chaining explains all the part of the boundedness of these processes which is due to cancellations. The results are described in complete detail with all definitions in Section 12.3.

¹⁰ It is a good research program to discover a more intuitive approach to this result.

1.7.12 Unfulfilled Dreams

Having proved in several general settings that "chaining explains all the part of the boundedness which is due to cancellation", we concentrate on the problem of describing the "part of the boundedness which owes nothing to cancellation". We propose sweeping conjectures. The underlying hope behind these conjectures is that, ultimately, a bound for a selector process always arises from the use of the 'union bound' $P(\bigcup_n A_n) \leq \sum_n P(A_n)$ in a simple situation, the use of basic principles such as linearity and positivity, or combinations of these.

1.7.13 Empirical Processes

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

We demonstrate again the power of the chaining scheme of Section 9.4 by providing a sharper version of Ossiander's bracketing theorem with a very simple proof. We then illustrate various techniques by presenting proofs of two deep recent results.

1.7.14 Gaussian Chaos

Our satisfactory understanding of the properties of Gaussian processes should bring information about processes that are, in various senses, related to Gaussian processes. Such is the case of an order two Gaussian chaos (which is essentially a family of second degree polynomials of Gaussian random variables). It seems at present a hopelessly difficult task to give lower and upper bounds of the same order for these processes, but in Section 15.1 we obtain a number of results in this direction. Chaos processes are very instructive because there exist other methods than chaining to control their size (a situation which we do not expect to occur for processes defined as sums of a random series).

In Section 15.2 we study the tails of a single multiple-order Gaussian chaos, and present (yet another) deep result of R. Latała which provides a rather complete description of the size of these tails.

1.7.15 Convergence of Orthogonal Series; Majorizing Measures

The old problem of characterizing the sequences (a_m) such that for each orthonormal sequence (φ_m) the series $\sum_{m\geq 1} a_m \varphi_m$ converges a.s. was solved by A. Paszkiewicz. Using a more abstract point of view, we present a very much simplified proof of his results (due essentially to W. Bednorz). This leads us to the question of discussing when a certain condition on the "increments" of a process implies its boundedness. When the increment condition is of "polynomial type", this is more difficult than in the case of Gaussian processes,

and requires the notion of "majorizing measure". We present several elegant results of this theory, in their seemingly final forms recently obtained by W. Bednorz.

1.7.16 Shor's Matching Theorem

This chapter continues Chapter 4. We prove a deep improvement of the Ajtai, Komlós, Tusnády theorem due to P. Shor. Unfortunately, due mostly to our lack of geometrical understanding, the best conceivable matching theorem, which would encompass this result as well as those of Chapter 4, and much more, remains as a challenging problem, "the ultimate matching conjecture" (a conjecture which is solved in the next chapter in dimension ≥ 3).

1.7.17 The Ultimate Matching Theorem in Dimension Three

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

1.7.18 Applications to Banach Space Theory

Chapter 19 gives applications to Banach space theory. As interest in this theory has decreased in recent years, we have not reproduced many of the results of [171], and we urge the interested reader to consult this earlier edition. We have kept only the results which make direct use of results presented elsewhere in the book (rather than including results based on the methods of the book). In Section 19.1.2, we study the cotype of operators from ℓ_N^{∞} into a Banach space. In Section 19.1.3, we prove a comparison principle between Rademacher (=Bernoulli) and Gaussian averages of vectors in a finite dimensional Banach space, and we use it to compute the Rademacher cotype-2 of a finite dimensional space using only a few vectors. In Section 19.2.1 we discover how to classify the elements of the unit ball of L^1 "according to the size of the level sets". In Section 19.2.3 we explain, given a 1-unconditional sequence $(e_i)_{i \leq N}$ in a Banach space E how to "compute" the quantity $\mathsf{E} \| \sum_i g_i e_i \|$ when g_i are independent Gaussian r.v.s, a further variation on the fundamental theme of the interplay between the L^1, L^2 and L^{∞} norms. In Section 19.3.1 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 19.3.2 we use these results to deduce the celebrated results of J. Bourgain on the Λ_p problem. Recent results of Gilles Pisier on Sidon sets conclude this chapter in Section 19.4.

Part I

The Generic Chaining

2. Gaussian Processes and the Generic Chaining

2.1 Overview

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

2.2 Measuring the Size of the Supremum

In this section we consider a metric space (T, d) and a process $(X_t)_{t \in T}$. Unless explicitly specified otherwise (and even when we forget to repeat it) we will *always* assume that the process is centered, i.e.

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

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

When T is uncountable it is not obvious what the quantity $\mathsf{E} \sup_{t \in T} X_t$ means.¹ We *define* it by the following formula:

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

where the right-hand side makes sense as soon as each r.v. X_t is integrable. This will be the case in almost all the situations considered in this book.

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

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

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

 $^{^{-1}}$ Such questions are treated in detail e.g. in [75] pages 42-43.

24 2. Gaussian Processes and the Generic Chaining

Proof. We note that

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

and we take expectations.²

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

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

The previous exercise is easy, but this need not be always the case. The author has never taught this material in a classroom, and cannot really evaluate the level of difficulty of the exercises for a beginner. So please do not feel discouraged if most of the exercises feel like research problems.³ A sketch of a solution is provided for almost every exercise. For the exercises which are too difficult, understanding this very concise sketch is in itself a good exercise. Just try to peek at the solution one line at a time.

In this book, we often state inequalities about the supremum of a symmetric process using the quantity $\mathsf{E} \sup_{t \in T} X_t$ simply because this quantity looks typographically more elegant than the equivalent⁴ quantity $\mathsf{E} \sup_{s,t \in T} |X_s - X_t|$. It is good to remember that when $X_{t_0} = 0$ for some $t_0 \in T$, (2.3) shows that there is not so much difference between $\mathsf{E} \sup_{t \in T} X_t$ and $\mathsf{E} \sup_{t \in T} |X_t|$.

We actually often need to control the tails of the r.v. $\sup_{s,t\in T} |X_s - X_t|$, not only its first moment. Emphasis is given to the first moment because this is the difficult part, and once this is achieved, control of higher moments is often provided by the same arguments.

2.3 The Union Bound and Other Basic Facts

From now on we assume that the process $(X_t)_{t \in T}$ satisfies the increment condition:

$$\forall u > 0 , \mathsf{P}(|X_s - X_t| \ge u) \le 2 \exp\left(-\frac{u^2}{2d(s,t)^2}\right),$$
 (2.4)

² To be really rigorous, we should first consider the case where T is finite and then appeal to (2.2), but it is better to skip this kind of tedious detail.

³ I had feedback from talented readers who felt that way. Consequently, I did not shy away to state as "exercises" rather non-trivial material complementing the text while being fully aware that one has to have achieved a rather complete understanding of the concepts as well as a mastery of the techniques to solve them.

⁴ Equivalent does not mean equal, we have been dropping a factor 2 here. Generally speaking the methods of this book are not appropriate to find sharp numerical constants and all the crucial inequalities are "sharp within a multiplicative constant".

where d is a distance on T. In particular this is the case when $(X_t)_{t\in T}$ is a Gaussian process and $d(s,t)^2 = \mathsf{E}(X_s - X_t)^2$. Our goal is to find bounds on $\mathsf{E}\sup_{t\in T} X_t$ depending on the structure of the metric space (T,d). We will assume that T is finite, which, as shown by (2.2), does not decrease generality.

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

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

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

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

Let us note that since the function $u \mapsto \mathsf{P}(Y \ge u)$ is non-increasing, for any u > 0 we have the following

$$\mathsf{E}Y \ge u\mathsf{P}(Y \ge u) \ . \tag{2.7}$$

In particular $\mathsf{P}(Y \ge u) \le \mathsf{E}Y/u$, a very important fact known as Markov's inequality. Arguments such as the following one will be of constant use.

Exercise 2.3.1. Consider a r.v. $Y \ge 0$ and a > 0. Prove that $\mathsf{P}(Y \le a\mathsf{E}Y) \ge 1 - 1/a$.

Let us stress the content of this result. It will be used when Y is a kind a random error, of very small expectation, $\mathsf{E}Y = b^2$ where b is small. Then most of time Y is small: $\mathsf{P}(Y \le b) \ge 1 - b$.

According to (2.6) it is natural to look for bounds of

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

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

$$\mathsf{P}\Big(\sup_{t\in T} (X_t - X_{t_0}) \ge u\Big) \le \sum_{t\in T} \mathsf{P}(X_t - X_{t_0} \ge u) .$$
(2.9)

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

$$\Delta(T) = \sup_{t_1, t_2 \in T} d(t_1, t_2) .$$
(2.10)

When we need to make clear which distance we use in the definition of the diameter, we will write $\Delta(T, d)$ rather than $\Delta(T)$. Consequently (2.4) and (2.9) imply
$$\mathsf{P}\left(\sup_{t\in T} (X_t - X_{t_0}) \ge u\right) \le 2\operatorname{card} T \exp\left(-\frac{u^2}{2\Delta(T)^2}\right).$$
(2.11)

Let us now record a simple yet important computation, that will allow us to use the information (2.11).

Lemma 2.3.2. Consider a r.v. $Y \ge 0$ which satisfies

$$\forall u > 0, \ \mathsf{P}(Y \ge u) \le A \exp\left(-\frac{u^2}{B^2}\right)$$
 (2.12)

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

$$\mathsf{E}Y \le LB\sqrt{\log A} \ . \tag{2.13}$$

Here, as in the entire book, L denotes a universal constant.⁵ We make the convention that this constant *is not necessarily* the same on each occurrence (even in the same equation). This should be remembered at all times. One of the benefits of the convention (as opposed to writing explicit constants) is to make clear that one is not interested in getting sharp constants. Getting sharp constants might be useful for certain applications, but it is a different game.⁶ The convention is very convenient, but one needs to get used to it. Now is the time for this, so we urge the reader to pay the greatest attention to the next exercise.

Exercise 2.3.3. (a) Prove that for $x, y \in \mathbb{R}^+$ we have $xy - Lx^3 \leq Ly^{3/2}$. (Please understand this statement as: given a number L_1 , there exists a number L_2 such that for all $x, y \in \mathbb{R}$ we have $xy - L_1x^{1/3} \leq L_2y^{3/2}$.)

(b) Consider a function $p(u) \leq 1$ for $u \geq 0$. Assume that for u > L we have $p(u) \leq L \exp(-u^2/L)$. Prove that for all u > 0 we have $p(Lu) \leq 2 \exp(-u^2)$. (Of course, this has to be understood as follows: assume that for a certain number L_1 , for $u > L_1$ we have $p(u) \leq L_1 \exp(-u^2/L_1)$. Prove that there exists a number L_2 such that for all u > 0 we have $p(L_2u) \leq \exp(-u^2)$.) (c) Consider an integer $N \geq 1$. Prove that

$$N^{L} \exp(-(\log N)^{3/2}/L) \le L \exp(-(\log N)^{3/2}/L)$$
.

Proof of Lemma 2.3.2. We use (2.6) and we observe that since $\mathsf{P}(Y \ge u) \le 1$, for any number u_0 we have

⁵ When meeting an unknown notation such as this previous L, the reader might try to look at the *index*, where some of the most common notation is recorded.

⁶ Our methods here are not appropriate for this.

2.3 The Union Bound and Other Basic Facts 27

$$\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y \ge u) \mathrm{d}u = \int_0^{u_0} \mathsf{P}(Y \ge u) \mathrm{d}u + \int_{u_0}^\infty \mathsf{P}(Y \ge u) \mathrm{d}u$$
$$\leq u_0 + \int_{u_0}^\infty A \exp\left(-\frac{u^2}{B^2}\right) \mathrm{d}u$$
$$\leq u_0 + \frac{1}{u_0} \int_{u_0}^\infty u A \exp\left(-\frac{u^2}{B^2}\right) \mathrm{d}u$$
$$= u_0 + \frac{AB^2}{2u_0} \exp\left(-\frac{u_0^2}{B^2}\right). \tag{2.14}$$

The choice of $u_0 = B\sqrt{\log A}$ gives the bound

$$B\sqrt{\log A} + \frac{B}{2\sqrt{\log A}} \le LB\sqrt{\log A}$$

since $A \geq 2$.

Next, recalling that the process $(X_t)_{t \in T}$ is assumed to satisfy (2.4) throughout the section, we claim that

$$\mathsf{E}\sup_{t\in T} X_t \le L\Delta(T)\sqrt{\log\operatorname{card} T} \ . \tag{2.15}$$

Indeed, this is obvious if card T = 1. If card $T \ge 2$ it follows from (2.11) that (2.12) holds for $Y = \sup_{t \in T} (X_t - X_{t_0})$ with A = 2 card T and $B = \Delta(T)$, and the result follows from (2.13) since $\log(2 \operatorname{card} T) \le 2 \log \operatorname{card} T$ and $\mathsf{E}Y = \mathsf{E} \sup_{t \in T} X_t$.

The following special case is fundamental.

Lemma 2.3.4. If $(g_k)_{k\geq 1}$ are standard Gaussian r.v.s then

$$\mathsf{E}\sup_{k\leq N}g_k\leq L\sqrt{\log N}\;.\tag{2.16}$$

Exercise 2.3.5. (a) Prove that (2.16) holds for any r.v.s g_k which satisfy

$$\mathsf{P}(g_k \ge u) \le 2\exp\left(-\frac{u^2}{2}\right) \tag{2.17}$$

for u > 0.

(b) For $N \ge 2$ construct N centered r.v.s $(g_k)_{k\le N}$ satisfying (2.17), and taking only the values $0, \pm \sqrt{\log N}$ and for which $\mathsf{E} \sup_{k\le N} g_k \ge \sqrt{\log N}/L$. (You are not yet asked to make these r.v.s independent.)

(c) After learning (2.18) below, solve (b) with the further requirement that the r.v.s g_k are independent. If this is too hard, look at Exercise 2.3.7, (b) below.

This is taking us a bit ahead, but an equally fundamental fact is that when the r.v.s (g_k) are jointly Gaussian, and "significantly different from each other"

i.e. $\mathsf{E}(g_k - g_\ell)^2 \geq a^2 > 0$ for $k \neq \ell$, the bound (2.16) can be reversed, i.e. $\mathsf{E}\sup_{k\leq N}g_k\geq a\sqrt{\log N}/L$, a fact known as Sudakov's minoration. Sudakov's minoration is a non-trivial fact, and to understand it, it should be really useful to solve Exercise 2.3.7 below. However, before that let us point out a simple fact, that will be used many times.

Exercise 2.3.6. Consider independent events $(A_k)_{k\geq 1}$. Prove that

$$\mathsf{P}\Big(\bigcup_{k\leq N} A_k\Big) \geq 1 - \exp\Big(-\sum_{k\leq N} \mathsf{P}(A_k)\Big) . \tag{2.18}$$

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

Exercise 2.3.7. (a) Consider independent r.v.s $Y_k \ge 0$ and u > 0 with

$$\sum_{k \le N} \mathsf{P}(Y_k \ge u) \ge 1 .$$
(2.19)

Prove that

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

Hint: use (2.18) to prove that $\mathsf{P}(\sup_{k \leq N} Y_k \geq u) \geq 1/L$. (b) We assume (2.19), but now Y_k need not be ≥ 0 . Prove that

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

Hint: observe that for each event Ω we have $\mathsf{E1}_{\Omega} \sup_k Y_k \ge -\mathsf{E}|Y_1|$. (c) Prove that if $(g_k)_{k\ge 1}$ are independent standard Gaussian r.v.s then $\mathsf{E} \sup_{k\le N} g_k \ge \sqrt{\log N}/L$.

Before we go back to our main story, we consider in detail the consequences of an "exponential decay of tails" such as in (2.12). This is the point of the next exercise.

Exercise 2.3.8. (a) Assume that for a certain B > 0 the r.v. $Y \ge 0$ satisfies

$$\forall u > 0, \ \mathsf{P}(Y \ge u) \le 2 \exp\left(-\frac{u}{B}\right).$$
 (2.20)

Prove that

$$\mathsf{E}\exp\left(\frac{Y}{2B}\right) \le L \ . \tag{2.21}$$

Prove that for x, a > 0 one has $(x/a)^a \leq \exp x$. Use this for a = p and x = Y/2B to deduce from (2.21) that for $p \geq 1$ one has

$$(\mathsf{E}Y^p)^{1/p} \le LpB \ . \tag{2.22}$$

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

$$\forall u > 0 , \mathsf{P}(Y \ge u) \le 2 \exp\left(-\frac{u^2}{B^2}\right),$$
 (2.23)

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

$$(\mathsf{E}Y^p)^{1/p} \le LB\sqrt{p} \ . \tag{2.24}$$

(c) Consider a r.v. $Y \ge 0$ and a number B > 0. Assuming that for $p \ge 1$ we have $(\mathsf{E}Y^p)^{1/p} \le Bp$ prove that for u > 0 we have $\mathsf{P}(Y > u) \le 2\exp(-u/(LB))$. Assuming that for each $p \ge 1$ we have $(\mathsf{E}Y^p)^{1/p} \le B\sqrt{p}$ prove that for u > 0 we have $\mathsf{P}(Y > u) \le 2\exp(-u^2/(LB^2))$.

In words, (2.22) states that "as p increases, the L^p norm of an exponentially integrable r.v. does not grow faster than p," and (2.24) asserts that if the square of the r.v. is exponentially integrable, then its L^p norm does not grow faster than \sqrt{p} . These two statements are closely related. More generally it is very classical to relate the size of the tails of a r.v. with the rate of growth of its L^p norm. This is not explicitly used in the sequel, but is good to know as background information. As the following shows, (2.24) provides the correct rate of growth in the case of Gaussian r.v.s.

Exercise 2.3.9. If g is a standard Gaussian r.v. it follows from (2.24) that for $p \ge 1$ one has $(\mathsf{E}|g|^p)^{1/p} \le L\sqrt{p}$. Prove one has also

$$(\mathsf{E}|g|^p)^{1/p} \ge \frac{\sqrt{p}}{L} \ . \tag{2.25}$$

One knows how to compute exactly $\mathsf{E}[g]^p$, from which one can deduce (2.25). You are however asked to provide a proof in the spirit of this work by deducing (2.25) solely from the information that, say, for u > 0 we have (choosing on purpose crude constants) $\mathsf{P}([g] \ge u) \ge \exp(-u^2/3)/100$.

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

2.4 The Generic Chaining

We go back to our main story. The bound (2.9) (and hence (2.15)) will be effective if the variables $X_t - X_{t_0}$ are rather uncorrelated (and if there are not too many of them). But it will be a disaster if many of the variables $(X_t)_{t \in T}$ are nearly identical. Thus it seems a good idea to gather those variables X_t which are nearly identical. To do this, we consider a subset T_1 of T, and

for t in T we consider a point $\pi_1(t)$ in T_1 , which we think of as a (first) approximation of t. The elements of T which correspond to the same point $\pi_1(t)$ are, at this level of approximation, considered as identical. We then write

$$X_t - X_{t_0} = X_t - X_{\pi_1(t)} + X_{\pi_1(t)} - X_{t_0} .$$
(2.26)

The idea is that it will be effective to use (2.9) for the variables $X_{\pi_1(t)} - X_{t_0}$, because there are not too many of them and, if we have done a good job at finding $\pi_1(t)$, they are rather different from each other (at least in some global sense). 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 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

$$X_t - X_{t_0} = \sum_{n \ge 1} \left(X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \right), \qquad (2.27)$$

which holds provided we arrange that $\pi_n(t) = t$ for n large enough, in which case the series is actually a finite sum. Relation (2.27) decomposes the increments of the process $X_t - X_{t_0}$ along the "chain" $(\pi_n(t))_{n\geq 0}$ (and this is why this method is called "chaining").

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

$$\operatorname{card} T_n \le N_n \tag{2.28}$$

where

$$N_0 = 1; N_n = 2^{2^n} \text{ if } n \ge 1.$$
 (2.29)

The notation (2.29) will be used throughout the book. It is at this stage that the procedure to control T_n differs from the traditional one, and it is the crucial point of the generic chaining method.

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

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

which makes it very convenient to work with this sequence.

Since $\pi_n(t)$ approximates t, it is natural to assume that⁷

⁷ The notation := below stresses that this is a definition, so that you should not worry that your memory failed and that you did not see this before.

2.4 The Generic Chaining 31

$$d(t, \pi_n(t)) = d(t, T_n) := \inf_{s \in T_n} d(t, s) .$$
(2.30)

For u > 0, (2.4) implies

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

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

card
$$T_n \cdot \text{card } T_{n-1} \le N_n N_{n-1} \le N_{n+1} = 2^{2^{n+1}}$$

We define the (favorable) event $\Omega_{u,n}$ by

$$\forall t , |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \le u 2^{n/2} d(\pi_n(t), \pi_{n-1}(t)) .$$
(2.31)

and we define $\Omega_u = \bigcap_{n \ge 1} \Omega_{u,n}$. Then

$$p(u) := \mathsf{P}(\Omega_u^c) \le \sum_{n \ge 1} \mathsf{P}(\Omega_{u,n}^c) \le \sum_{n \ge 1} 2 \cdot 2^{2^{n+1}} \exp(-u^2 2^{n-1}).$$
(2.32)

Here again, at the crucial step, we have used the union bound: $\mathsf{P}(\Omega_u^c) \leq \sum_{n\geq 1}\mathsf{P}(\Omega_{u,n}^c)$. When Ω_u occurs, (2.27) yields

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

so that

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

where

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

Thus

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

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

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

from which it follows that

$$p(u) \le L \exp\left(-rac{u^2}{2}
ight)$$
 .

We observe here that since $p(u) \leq 1$ the previous inequality holds not only for $u \geq 3$ but also for u > 0, because $1 \leq \exp(9/2) \exp(-u^2/2)$ for $u \leq 3$. This type of argument (i.e. changing the universal constant in front of the exponential, cf. Exercise (2.3.3)(b)) will be used repeatedly. Therefore

$$\mathsf{P}\left(\sup_{t\in T} |X_t - X_{t_0}| > uS\right) \le L \exp\left(-\frac{u^2}{2}\right).$$
(2.33)

In particular (2.33) implies

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

The triangle inequality yields

$$d(\pi_n(t), \pi_{n-1}(t)) \le d(t, \pi_n(t)) + d(t, \pi_{n-1}(t)) = d(t, T_n) + d(t, T_{n-1})$$

so that (making the change of variable n = n' + 1 in the second sum below)

$$S \leq \sup_{t \in T} \sum_{n \geq 1} 2^{n/2} d(t, T_n) + \sup_{t \in T} \sum_{n \geq 1} 2^{n/2} d(t, T_{n-1}) \leq 3 \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} d(t, T_n) ,$$

and we have proved the fundamental bound

$$\mathsf{E}\sup_{t\in T} X_t \le L \sup_{t\in T} \sum_{n\ge 0} 2^{n/2} d(t, T_n) \,. \tag{2.34}$$

Now, how do we construct the sets T_n ? It is obvious that we should try to make the right-hand side of (2.34) small, but this is obvious only because we have used an approach which naturally leads to this bound. In the next section, we investigate how this was traditionally done. Before this, we urge the reader to fully understand the next exercise. It will be crucial to understand a typical case where the traditional methods are not effective.

Exercise 2.4.1. Consider a countable metric space, $T = \{t_1, t_2, \ldots\}$. Assume that for each $i \geq 2$ we have $d(t_1, t_i) \leq 1/\sqrt{\log i}$. Prove that if $T_n = \{t_1, t_2, \ldots, t_{N_n}\}$ then for each $t \in T$ we have $\sum_{n>0} 2^{n/2} d(t, T_n) \leq L$.

We end this section by reviewing at a high level the scheme of the previous proof (which will be used again and again). The goal is to bound $\mathsf{E}Y$ where Y is a r.v. ≥ 0 (here $Y = \sup_t (X_t - X_{t_0})$.) The method consists of two steps:

- Given a parameter $u \ge 0$ one identifies a "good set" Ω_u , where some undesirable events do not happen. As u becomes large $\mathsf{P}(\Omega_u^c)$ becomes small.
- When Ω_u occurs we bound Y, say $Y \leq f(u)$ where f is an increasing function on \mathbb{R}^+ .

One then obtains the bound

$$\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y \ge u) \mathrm{d}u \le f(0) + \int_{f(0)}^\infty \mathsf{P}(Y \ge u) \mathrm{d}u$$
$$= f(0) + \int_0^\infty f'(u) \mathsf{P}(Y \ge f(u)) \mathrm{d}u , \quad (2.35)$$

where we have used a change of variable in the last equality. Now, since $Y \leq f(u)$ on Ω_u , we have $\mathsf{P}(Y \geq f(u)) \leq \mathsf{P}(\Omega_u^c)$ and finally

$$\mathsf{E}Y \leq f(0) + \int_0^\infty f'(u) \mathsf{P}(\Omega_u^c) \mathrm{d}u$$

In practice, we will always have $\mathsf{P}(\Omega_u^c) \leq L \exp(-u/L)$ and $f(u) = A + u^{\alpha}B$, yielding the bound $\mathsf{E}Y \leq A + K(\alpha)B$.

2.5 Entropy Numbers

For a number of years, chaining was systematically performed (as in Section 1.4) by choosing the sets T_n so that $\sup_{t \in T} d(t, T_n)$ is as small as possible for card $T_n \leq N_n$. We define

$$e_n(T) = e_n(T, d) = \inf_{T_n \subset T, \text{card } T_n \le N_n} \sup_{t \in T} d(t, T_n), \qquad (2.36)$$

where the infimum is taken over all subsets T_n of T with card $T_n \leq N_n$. (Since here T is finite, the infimum is actually a minimum.) We call the numbers $e_n(T)$ the entropy numbers.

Let us recall that in a metric space a (closed) ball is a set of the type $B(t,r) = \{s \in T; d(s,t) \leq r\}$. Balls are basic sets in a metric space and will be of constant use. It should be obvious to reformulate (2.36) as follows: $e_n(T)$ is the infimum of the set of numbers $r \geq 0$ such that T can be covered by $\leq N_n$ balls of radius $\leq r$ (the set T_n in (2.36) being the set of centers of these balls).

The definition (2.36) is not consistent with the conventions of Operator Theory, which uses e_{2^n} to denote what we call e_n .⁸ When T is infinite, the numbers $e_n(T)$ are also defined by (2.36) but are not always finite (e.g. when $T = \mathbb{R}$).

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

$$\frac{\Delta(T)}{2} \le e_0(T) \le \Delta(T) . \tag{2.37}$$

Recalling that T is finite, let us then choose for each n a subset T_n of T with card $T_n \leq N_n$ and $e_n(T) = \sup_{t \in T} d(t, T_n)$. Since $d(t, T_n) \leq e_n(T)$ for each t, (2.34) implies the following.

Proposition 2.5.1 (Dudley's entropy bound [38]). Under the increment condition (2.4), it holds that

$$\mathsf{E} \sup_{t \in T} X_t \le L \sum_{n \ge 0} 2^{n/2} e_n(T) .$$
(2.38)

⁸ We can't help it if Operator Theory gets it wrong.

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

Lemma 2.5.2. If U is a subset of T, we have $e_n(U) \leq 2e_n(T)$.

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

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

Exercise 2.5.3. Prove that the factor 2 in the inequality $e_n(U) \leq 2e_n(T)$ cannot be improved even if n = 0.

Dudley's entropy bound is usually formulated using the covering numbers of Definition 1.4.1. These relate to the entropy numbers by the formula

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

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

$$\begin{split} \epsilon < e_n(T) \Rightarrow N(T,d,\epsilon) > N_n \\ \Rightarrow N(T,d,\epsilon) \geq 1 + N_n \; . \end{split}$$

Therefore

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

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

$$\sqrt{\log 2} \sum_{n \ge 0} 2^{n/2} (e_n(T) - e_{n+1}(T)) \le \int_0^{e_0(T)} \sqrt{\log N(T, d, \epsilon)} \,\mathrm{d}\epsilon \;.$$
(2.39)

Now,

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

so (2.39) yields

2.5 Entropy Numbers 35

$$\sum_{n\geq 0} 2^{n/2} e_n(T) \leq L \int_0^{e_0(T)} \sqrt{\log N(T, d, \epsilon)} \,\mathrm{d}\epsilon \,. \tag{2.40}$$

Hence Dudley's bound now appears in the familiar form

$$\mathsf{E}\sup_{t\in T} X_t \le L \int_0^\infty \sqrt{\log N(T, d, \epsilon)} \,\mathrm{d}\epsilon \;. \tag{2.41}$$

Here, since $\log 1 = 0$, the integral takes place in fact over $0 \le \epsilon \le e_0(T)$. The right-hand side is often called *Dudley's entropy integral*.

Exercise 2.5.4. Prove that

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

showing that (2.38) is not an improvement over (2.41).

Exercise 2.5.5. Assume that for each $0 < \epsilon < A$ and some $\alpha > 0$ we have $\log N(T, d, \epsilon) \leq (A/\epsilon)^{\alpha}$. Prove that $e_n(T) \leq K(\alpha)A2^{-n/\alpha}$.

Here $K(\alpha)$ is a number depending only on α .⁹ This, and similar notation are used throughout the book. It is understood that such numbers *need not* be the same on every occurrence, and it would help to remember this at all times. The difference between the notations K and L is that L is a universal constant, i.e. a number that does not depend on anything, while K might depend on some parameters, such as α here.

When writing a bound such as (2.41), the immediate question is: how sharp is it? The word "sharp" is commonly used, even though people do not agree on what it means exactly. Let us say that a bound of the type $A \leq LB$ can be reversed if it is true that $B \leq LA$. We are not concerned with the value of the universal constants.¹⁰ Inequalities which can be reversed are our best possible goal. Then, in any circumstance, A and B are of the same order.

We give now a simple (and classical) example that illustrates well the difference between Dudley's bound (2.38) and the bound (2.34) and which shows in particular that *Dudley's bound cannot be reversed*. Consider an independent sequence $(g_i)_{i\geq 1}$ of standard Gaussian r.v.s. Set $X_1 = 0$ and for $i \geq 2$ set

$$X_i = \frac{g_i}{\sqrt{\log i}} \ . \tag{2.42}$$

Consider an integer $s \geq 3$ and the process $(X_i)_{1 \leq i \leq N_s}$ so the index set is $T = \{1, 2, \ldots, N_s\}$. The distance d associated to the process, given by $d(i, j)^2 = \mathsf{E}(X_s - X_t)^2$, satisfies for $i, j \geq 2, i \neq j$,

⁹ It just happens that in this particular case $K(\alpha) = 1$ works, but we typically do not care about the precise dependence of $K(\alpha)$ on α .

¹⁰ Not that these values are unimportant, but our methods are not appropriate for this.

$$\frac{1}{\sqrt{\log(\min(i,j))}} \le d(i,j) \le \frac{2}{\sqrt{\log(\min(i,j))}} .$$
(2.43)

Consider $1 \leq n \leq s-1$ and $T_n \subset T$ with card $T_n = N_n$. There exists $i \leq N_n+1$ with $i \notin T_n$. Then (2.43) implies that $d(i,j) \geq 2^{-n/2}/L$ for $j \in T_n$. This proves that the balls of radius $2^{-n/2}/L$ centered on T_n do not cover T, so that $e_n(T) \geq 2^{-n/2}/L$. Therefore

$$\sum_{n} 2^{n/2} e_n(T) \ge \frac{s-1}{L} .$$
 (2.44)

In the reverse direction, since for $i \ge 1$ we have $d(1,i) \le 1/\sqrt{\log i}$, Exercise 2.4.1 proves that the bound (2.34) is $\le L$. Thus the bound (2.38) is worse than the bound (2.34) by a factor about s.

Exercise 2.5.6. Prove that when T is finite, the bound (2.41) cannot be worse than (2.34) by a factor greater than about log log card T. This shows that the previous example is in a sense extremal. Hint: use $2^{n/2}e_n(T) \leq L \sup_{t \in T} \sum_{n>0} 2^{n/2} d(t, T_n)$ and $e_n(T) = 0$ if $N_n \geq \text{card } T$.

How does one estimate covering numbers (or, equivalently, entropy numbers)? Let us first stress a trivial but nonetheless fundamental fact.

Lemma 2.5.7. Consider a number $\epsilon > 0$ and a subset W of T maximal with respect to the property

$$s, t \in W \Rightarrow d(s, t) > \epsilon$$
.

Then $N(T, d, \epsilon) \leq \operatorname{card} W$.

Proof. Since W is maximum the balls of radius a centered at the points of W cover T.

Exercise 2.5.8. Consider a probability measure μ on T, a number $\epsilon > 0$ and a number a. Let $U = \{t \in T; \mu(B(t, \epsilon) \ge a)\}$. Prove that $N(U, d, 2\epsilon) \le 1/a$.

The next exercise introduces the reader to "volume estimates", a simple yet fundamental method for this purpose. It deserves to be fully understood. If this exercise is too hard, you can find all the details below in the proof of Lemma 2.13.7.

Exercise 2.5.9. (a) If (T, d) is a metric space, define the packing number $N^*(T, d, \epsilon)$ as the largest integer N such that T contains N points with mutual distances $\geq \epsilon$. Prove that $N(T, d, \epsilon) \leq N^*(T, d, \epsilon)$. Prove that if $\epsilon' > 2\epsilon$ then $N^*(T, d, \epsilon') \leq N(T, d, \epsilon)$.

(b) Consider a distance d on \mathbb{R}^k which arises from a norm $\|\cdot\|$, $d(x,y) = \|x - y\|$, and denote by B the unit ball of center 0. Let us denote by $\operatorname{Vol}(A)$ the k-dimensional volume of a subset A of \mathbb{R}^k . By comparing volumes, prove that for any subset A of \mathbb{R}^k ,

2.5 Entropy Numbers 37

$$N(A, d, \epsilon) \ge \frac{\operatorname{Vol}(A)}{\operatorname{Vol}(\epsilon B)}$$
(2.45)

and

$$N(A, d, 2\epsilon) \le N^*(A, d, 2\epsilon) \le \frac{\operatorname{Vol}(A + \epsilon B)}{\operatorname{Vol}(\epsilon B)} .$$
(2.46)

(c) Conclude that

$$\left(\frac{1}{\epsilon}\right)^k \le N(B, d, \epsilon) \le \left(\frac{2+\epsilon}{\epsilon}\right)^k$$
 (2.47)

(d) Use (c) to find estimates of $e_n(B)$ of the correct order for each value of n. Hint: $e_n(B)$ is about $2^{-2^n/k}$. This decreases very fast as n increases. Estimate Dudley's bound for B provided with the distance d.

(e) Prove that if T is a subset of \mathbb{R}^k and if n_0 is any integer then for $n \ge n_0$ one has $e_{n+1}(T) \le L2^{-2^n/k}e_{n_0}(T)$. Hint: cover T by N_{n_0} balls of radius $2e_{n_0}(T)$ and cover each of these by balls of smaller radius using (d).

(f) This part provides a generalization of (2.45) and (2.46) to a more abstract setting, but with the same proofs. Consider a metric space (T, d) and a positive measure μ on T such that all balls of a given radius have the same measure, $\mu(B(t, \epsilon)) = \varphi(\epsilon)$ for each $\epsilon > 0$ and each $t \in T$. For a subset Aof T and $\epsilon > 0$ let $A_{\epsilon} = \{t \in T; d(t, A) \le \epsilon\}$, where $d(t, A) = \inf_{s \in A} d(t, s)$. Prove that

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

There are many simple situations where Dudley's bound is not of the correct order. We gave a first example on page 35. We give such another example in Exercise 2.5.11 below. There the set T is particularly appealing: it is a simplex in \mathbb{R}^m . Yet other examples based on fundamental geometry (ellipsoids in \mathbb{R}^k) are explained in Section 2.13.

The result of the following exercise is very useful in all kinds of examples.

Exercise 2.5.10. Consider two integers k, m with $k \leq m/4$. Assume for simplicity that k is even.

(a) Prove that

$$\sum_{0 \le \ell \le k/2} \binom{m}{\ell} \le 2 \left(\frac{2k}{m}\right)^{k/2} \binom{m}{k} .$$
(2.48)

(b) Denote by \mathcal{I} the class of subsets of $\{1, \ldots, m\}$ of cardinality k. Prove that you can find in \mathcal{I} a family \mathcal{F} such for $I, J \in \mathcal{F}$ one has $\operatorname{card}(I \setminus J) \cup (J \setminus I) \geq k/2$ and $\operatorname{card} \mathcal{F} \geq (m/(2k))^{k/2}/2$. Hint: Use (a) and part (f) of Exercise 2.5.9 for μ the counting measure on \mathcal{I} . Warning: this is not so easy.

Exercise 2.5.11. Consider an integer m and an i.i.d. standard Gaussian sequence $(g_i)_{i \leq m}$. For $t = (t_i)_{i \leq m} \in \mathbb{R}^m$, let $X_t = \sum_{i \leq m} t_i g_i$. This is called the

canonical Gaussian process on \mathbb{R}^m . Its associated distance is the Euclidean distance on \mathbb{R}^m . It will be much used later. Consider the set

$$T = \left\{ (t_i)_{i \le m} \in \mathbb{R}^m \; ; \; t_i \ge 0 \; , \; \sum_{i \le m} t_i = 1 \right\} \; , \tag{2.49}$$

the convex hull of the canonical basis. By (2.16) we have $\operatorname{\mathsf{E}sup}_{t\in T} X_t = \operatorname{\mathsf{E}sup}_{i\leq m} g_i \leq L\sqrt{\log m}$. Prove, however, that the right-hand side of (2.41) is $\geq (\log m)^{3/2}/L$. (Hint: For an integer $k \leq m$ consider the subset T_k of T consisting of sequences $t = (t_i)_{i\leq m} \in T$ for which $t_i \in \{0, 1/k\}$, so that $t \in T_k$ is determined by the set $I = \{i \leq m ; t_i = 1/k\}$ and card I = k. Using Exercise 2.5.10 prove that $\log N(T_k, d, 1/(L\sqrt{k})) \geq k \log(em/k)/L$ and conclude.¹¹) Thus in this case Dudley's bound is off by a multiplicative factor of about $\log m$. Exercise 2.7.9 below will show that in \mathbb{R}^m the situation cannot be worse than this.

2.6 Rolling Up our Sleeves: Chaining in the Simplex

The bound (2.34) seems to be genuinely better than the bound (2.38) because when going from (2.34) to (2.38) we have used the somewhat brutal inequality

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

The method leading to the bound (2.34) is probably the most important idea of this work. The fact that it appears now so naturally does not reflect the history of the subject, but rather that the proper approach is being used. When using this bound, we will choose the sets T_n in order to minimize the right-hand side of (2.34) instead of choosing them as in (2.36). As we will demonstrate later, this provides essentially the best possible bound for $\mathsf{E} \sup_{t \in T} X_t$. It is remarkable that despite the fact that this result holds in complete generality, it is a non-trivial task to find sets T_n witnessing this, even in very simple situations. In the present situation we perform this task by an explicit construction for the set T of (2.49).

Proposition 2.6.1. There exists sets $T_n \subset \mathbb{R}^m$ with card $T_n \leq N_n$ such that

$$\sup_{t\in T}\sum_{n\geq 0} 2^{n/2} d(t,T_n) \leq L\sqrt{\log m} \quad (=L\mathsf{E}\sup_{t\in T} X_t).$$

Of course here d is the Euclidean distance in \mathbb{R}^m . The reader may try to find these sets herself before reading the rest of this section, as there seems to be

¹¹ In case you wonder why e occurs in this formula, it is just to take care of the case where k is nearly m. This term is not needed here, but is important in upper bounds of the same nature that we will use below.

no better way to get convinced of the depth of the present theory. The sets T_n are not subsets of T. Please figure out by yourself how to correct this.¹²

Lemma 2.6.2. For each $t \in T$ we can find a sequence $(p(n,t))_{n\geq 0}$ of integers $0 \leq p(n,t) \leq 2n$ with the following properties:

$$\sum_{n\geq 0} 2^{n-p(n,t)} \leq L , \qquad (2.50)$$

$$\forall n \ge 0, \ p(n+1,t) \le p(n,t) + 2,$$
(2.51)

$$\operatorname{card}\{i \le m \; ; \; t_i \ge 2^{-p(n,t)}\} < 2^n \; .$$
 (2.52)

Proof. There is no loss of generality to assume that the sequence $(t_i)_{i\leq m}$ is non-increasing. We set $t_i = 0$ for i > m. Then for any $n \ge 1$ and $2^{n-1} < i \le 2^n$ we have $t_i \ge t_{2^n}$, so that $2^{n-1}t_{2^n} \le \sum_{2^{n-1} < i \le 2^n} t_i$. By summation over $n \ge 1$ we obtain $\sum_{n\ge 1} 2^n t_{2^n} \le 2$, and thus $\sum_{n\ge 0} 2^n t_{2^n} \le 3$. For $n \ge 0$ consider the largest integer $q(n,t) \le 2n$ such that $2^{-q(n,t)} > t_{2^n}$. Thus $2^{-q(n,t)-1} \le t_{2^n}$ when q < 2n. In any case $2^{-q(n,t)} \le 2t_{2^n} + 2^{-2n}$ and thus $\sum_{n\ge 0} 2^{n-q(n,t)} \le L$. Also if $t_i > 2^{-q(n,t)} > t_{2^n}$ then $i < 2^n$. In particular card $\{i \le m; t_i > 2^{-q(t,n)}\} < 2^n$. Finally we define

$$p(n,t) = \minig\{q(k,t) + 2(n-k); 0 \le k \le nig\}$$
 .

Taking k = n shows that $p(n,t) \le q(n,t) \le 2n$, implying (2.52). If $k \le n$ is such that p(n,t) = q(k,t) + 2(n-k) then $p(n+1,t) \le q(k,t) + 2(n+1-k) = p(n,t) + 2$, proving (2.51). Also, since $2^{n-p(n,t)} \le \sum_{k \le n} 2^{n-2(n-k)-q(k,t)} = \sum_{k \le n} 2^{k-n+(k-q(k,t))}$ we have

$$\sum_{n \ge 0} 2^{n - p(n, t)} \le \sum_{k \ge 0} 2^{k - q(k, t)} \sum_{n \ge k} 2^{k - n} \le L .$$

Given a set $I \subset \{1, \ldots, m\}$ and a integer p we denote by $V_{I,p}$ the set of elements $u = (u_i)_{i \leq m} \in \mathbb{R}^m$ such that $u_i = 0$ if $i \notin I$ and $u_i = r_i 2^{-p}$ if $i \in I$, where r_i is an integer $0 \leq r_i \leq 3$. Then card $V_{I,p} \leq 4^{\operatorname{card} I}$. For $n \geq 1$ we denote by V_n the union of all the sets $V_{I,p}$ for card $I \leq 2^n$ and $0 \leq p \leq 2n$. Crudely we have card $V_n \leq m^{L2^n}$. We set $V_0 = \{0\}$ and for $n \geq 1$ we denote by U_n the set of all sums $\sum_{0 \leq k \leq n} x_k$ where $x_k \in V_k$. Then card $U_n \leq m^{L2^n}$.

Lemma 2.6.3. Consider $t \in T$ and the sequence $(p(n,t))_{n\geq 0}$ constructed in Lemma 2.6.2. Then for each n we can write t = u(n) + v(n) where $u(n) \in U_n$ and where $v(n) = (v(n)_i)_{i\leq m}$ satisfies $0 \leq v(n)_i \leq \min(t_i, 2^{-p(n,t)})$.

Proof. The proof is by induction over n. For n = 0 we set u(0) = 0, v(0) = t. For the induction from n to n + 1 consider the set $I = \{i \leq m; v(n)_i > i\}$

 $^{^{12}}$ The argument can be found in Section 2.14.

 $2^{-p(n+1,t)}$ }. Since $v(n)_i \leq t_i$ it follows from (2.52) that card $I < 2^{n+1}$. For each $i \in I$ let r_i be the largest integer with $r_i 2^{-p(n+1,t)} < v(n)_i$ so that $v(n)_i - r_i 2^{-p(n+1,t)} \leq 2^{-p(n+1,t)}$. Since $v(n)_i \leq 2^{-p(n,t)}$ by induction and since $p(n+1,t) \leq p(n,t) + 2$ by (2.51) we have $r_i \leq 3$. Define $u = (u_i)_{i \leq m} \in \mathbb{R}^m$ by $u_i = r_i 2^{-p(n+1,t)}$ if $i \in I$ and $u_i = 0$ otherwise. Then $u \in V_{I,p(n+1,t)} \subset V_n$. Thus x = u(n+1) + v(n+1) where $u(n+1) := u(n) + u \in U_{n+1}$ and v(n+1) := v(n) - u satisfies $v(n+1)_i \leq \min(t_i, 2^{-p(n+1,t)})$.

Lemma 2.6.4. For each $t \in T$ we have $\sum_{n\geq 0} 2^{n/2} d(t, U_n) \leq L$.

Proof. Consider the sequence $(v(n))_{n\geq 0}$ constructed in Lemma 2.6.3, so that $d(t, U_n) \leq \|v(n)\|_2$ since t = u(n) + v(n). Let $I_n = \{i \leq m; t_i > 2^{-p(n,t)}\}$ so that by (2.52) we have card $I_n < 2^n$. For $n \geq 1$ set $J_n = I_n \setminus I_{n-1}$ so that for $i \in I_n$ we have $t_i < 2^{-p(n-1,t)}$. Then $\|v(n)\|_2^2 = \sum_{i \leq m} v(n)_i^2 = \sum_{i \in I_n} v(n)_i^2 + \sum_{k \geq n} \sum_{i \in J_k} v(n)_i^2$. Since $v(n)_i \leq 2^{-p(n,t)}$ and card $I_n \leq 2^n$ the first sum is $\leq 2^{n/2-p(n,t)}$. Since $v(n)_i \leq t_i \leq 2^{-p(k-1,t)}$ for $i \in J_k$ and card $J_k \leq 2^k$ we have $\sum_{i \in J_k} v(n)_i^2 \leq 2^{k/2-p(k-1,t)}$. Thus $\|v(n)\|_2 \leq \sum_{k \geq n} 2^{k/2-p(k-1,t)}$ and

$$\sum_{n\geq 1} 2^{n/2} \|v(n)\|_2 \leq \sum_{n\geq 1} \sum_{k\geq n} 2^{n/2+k/2-p(k-1,t)} = \sum_{k\geq 1} 2^{k/2-p(k-1,t)} \sum_{n\leq k} 2^{n/2}$$
$$\leq L \sum_{k\geq 1} 2^{k-p(k-1,t)} \leq L , \quad (2.53)$$

where we have used (2.50) in the last inequality.

Proof of Proposition 2.6.1. Consider the smallest integer k_0 with $m \leq N_{k_0}$ so that $2^{k_0/2} \leq L\sqrt{\log m}$. Observe also that $m^{2^n} \leq (2^{2^{k_0}})^{2^n} = 2^{2^{k_0+n}} = N_{k_0+n}$. Thus card $U_n \leq m^{L2^n} \leq N_{k_0+n+k_1}$ where k_1 is a universal constant. For $n \geq k_0 + k_1 + 1$ we set $T_n = U_{n-n_0-k_1}$, so that card $T_n \leq N_n$. For $n \leq k_0 + k_1$ we set $T_n = \{0\}$. Finally, given $t \in T$ (and keeping in mind that k_1 is a universal constant) we have

$$\sum_{n \ge 0} 2^{n/2} d(t, T_n) \le L 2^{k_0/2} + \sum_{n \ge k_0 + k_1 + 1} 2^{n/2} d(t, T_n)$$

and, using Lemma 2.6.4 in the last inequality,

$$\sum_{n \ge k_0 + k_1 + 1} 2^{n/2} d(t, T_n) = \sum_{n \ge k_0 + k_1 + 1} 2^{n/2} d(t, U_{n-k_0 - k_1})$$
$$= \sum_{n \ge 1} 2^{(n+k_0 + k_1)/2} d(t, U_n) \le L 2^{k_0/2} . \quad \Box$$

2.7 Admissible Sequences of Partitions

The idea behind the bound (2.34) admits a technically more convenient formulation. 13

Definition 2.7.1. Given a set T an admissible sequence is an increasing sequence $(\mathcal{A}_n)_{n\geq 0}$ of partitions of T such that $\operatorname{card} \mathcal{A}_n \leq N_n$, i.e. $\operatorname{card} \mathcal{A}_0 = 1$ and $\operatorname{card} \mathcal{A}_n \leq 2^{2^n}$ for $n \geq 1$.

By an *increasing* sequence of partitions we mean that every set of \mathcal{A}_{n+1} is contained in a set of \mathcal{A}_n . Admissible sequences of partitions will be constructed recursively, by breaking each element C of \mathcal{A}_n into at most N_n pieces, obtaining then a partition \mathcal{A}_{n+1} of T consisting of at most $N_n^2 \leq N_{n+1}$ pieces.

Throughout the book we denote by $A_n(t)$ the unique element of \mathcal{A}_n which contains t. The double exponential in the definition of N_n (see (2.29)) occurs simply since for our purposes the proper measure of the "size" of a partition \mathcal{A} is log card \mathcal{A} . This double exponential ensures that "the size of the partition \mathcal{A}_n doubles at every step". This offers a number of technical advantages which will become clear gradually.

Theorem 2.7.2. (The generic chaining bound). Under the increment condition (2.4) (and if $\mathsf{E}X_t = 0$ for each t), then for each admissible sequence (\mathcal{A}_n) we have

$$\mathsf{E}\sup_{t\in T} X_t \le L\sup_{t\in T} \sum_{n\ge 0} 2^{n/2} \mathcal{\Delta}(A_n(t)) .$$
(2.54)

Here as always, $\Delta(A_n(t))$ denotes the diameter of $A_n(t)$ for d. One could think that (2.54) could be much worse than (2.34), but it will turn out that this is not the case when the sequence (\mathcal{A}_n) is appropriately chosen.

Proof. We may assume T to be finite. We construct a subset T_n of T by taking exactly one point in each set A of \mathcal{A}_n . Then for $t \in T$ and $n \geq 0$, we have $d(t,T_n) \leq \Delta(\mathcal{A}_n(t))$ and the result follows from (2.34).

Definition 2.7.3. Given $\alpha > 0$, and a metric space (T, d) (that need not be finite) we define

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

where the infimum is taken over all admissible sequences.

It is useful to observe that since $A_0(t) = T$ we have $\gamma_{\alpha}(T, d) \geq \Delta(T)$. The most important cases by far are $\alpha = 2$ and $\alpha = 1$. For the time being we need only the case $\alpha = 2$. The case $\alpha = 1$ is first met in Theorem 4.5.13, although more general functionals occur first in Definition 4.5.

 $^{^{13}}$ We will demonstrate why this is the case only later, in Theorem 4.5.13.

Exercise 2.7.4. Prove that if $d \leq Bd'$ then $\gamma_2(T, d) \leq B\gamma_2(T, d')$.

Exercise 2.7.5. Prove that $\gamma_{\alpha}(T,d) \leq K(\alpha)\Delta(T)(\log \operatorname{card} T)^{1/\alpha}$ when T is finite. Hint: Ensure that $\Delta(A_n(t)) = 0$ if $N_n \geq \operatorname{card} T$.

A large part of our arguments will take place in abstract metric spaces, and this may represent an obstacle to the reader who has never thought about this. Therefore, we cannot recommend too highly the following exercise.

Exercise 2.7.6. (a) Consider a metric space (T, d) and assume that for each $n \ge 0$ you are given a covering \mathcal{B}_n of T with card $\mathcal{B}_n \le N_n$. Prove that you can construct an admissible sequence (\mathcal{A}_n) of partitions of T with the following property:

$$\forall n \ge 1 , \forall A \in \mathcal{A}_n , \exists B \in \mathcal{B}_{n-1} , A \subset B .$$
(2.55)

(b) Prove that for any metric space (T, d) we have

$$\gamma_2(T,d) \le L \sum_{n\ge 0} 2^{n/2} e_n(T)$$
 (2.56)

The following exercise explains one of the reasons admissible sequences of sets are so convenient: given two such sequences we can construct a third sequence which merges the good properties of the two sequences.

Exercise 2.7.7. Consider a set T and two admissible sequences (\mathcal{B}_n) and (\mathcal{C}_n) . Prove that there is an admissible sequence (\mathcal{A}_n) such that

$$\forall n \ge 1 , \forall A \in \mathcal{A}_n , \exists B \in \mathcal{B}_{n-1} , A \subset B , \exists C \in \mathcal{C}_{n-1}, A \subset C .$$

The following simple property should be clear in the reader's mind.

Exercise 2.7.8. (a) Prove that for $n \ge 0$ we have

$$2^{n/2}e_n(T) \le L\gamma_2(T,d) . (2.57)$$

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

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

The reader should compare (2.57) with (2.56).

Exercise 2.7.9. Use (2.57) and Exercise 2.5.9 (e) to prove that if $T \subset \mathbb{R}^m$ then

$$\sum_{n \ge 0} 2^{n/2} e_n(T) \le L \log(m+1)\gamma_2(T,d) .$$
(2.58)

In words, Dudley's bound is never off by more than a factor of about $\log(m+1)$ in \mathbb{R}^m .¹⁴

¹⁴ And we have shown in Exercise 2.5.6 that it is never off by a factor more that about $\log \log \operatorname{card} T$ either.

Exercise 2.7.10. Prove that the estimate (2.58) is essentially optimal. Warning: this requires some skill.

Combining Theorem 2.7.2 with Definition 2.7.3 yields the following very important result.

Theorem 2.7.11. Under (2.4) and (2.1) we have

$$\mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T,d) \ . \tag{2.59}$$

To make (2.59) of interest we must be able to control $\gamma_2(T, d)$, i.e. we must learn how to construct admissible sequences, a topic we shall first address in Section 2.9.

Exercise 2.7.12. When the process $(X_t)_{t \in T}$ satisfies (2.4) but is no longer assumed to be centered prove that

$$\mathsf{E}\sup_{s,t\in T}|X_s - X_t| \le L\gamma_2(T,d) .$$
(2.60)

We now turn to the control of the tails of the process, which will follow by a small variation of the same argument.

Theorem 2.7.13. Under (2.4) and (2.1) we have

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge L\gamma_2(T,d) + Lu\Delta(T)\Big) \le L\exp(-u^2).$$
(2.61)

Proof. We use the notation of the proof of (2.34). We may assume $u \ge 1$. Let us consider the smallest integer $k \ge 0$ such that $u^2 \le 2^k$ so that $2^k \le 2u^2$. Consider the event Ω_1 defined by

$$\forall t \in T_k , |X_t - X_{t_0}| \le 4u\Delta(T) , \qquad (2.62)$$

so that by the union bound,

$$P(\Omega_1^c) \le 2^{2^k} \cdot 2\exp(-8u^2) \le 2\exp(2u^2 - 8u^2) \le \exp(-u^2) .$$
 (2.63)

Consider the event Ω_2 given by

$$\forall n \ge k \;,\; \forall t \in T \;,\; |X_{\pi_n(t)} - X_{\pi_{n+1}(t)}| \le 2^{n/2+2} \Delta(A_n(t)) \;,$$
 (2.64)

so that by the union bound again,

$$\mathsf{P}(\Omega_2^c) \le \sum_{n \ge k} 2^{2^{n+1}} \cdot 2\exp(-2^{n+3}) \le \sum_{n \ge k} 2\exp(-2^{n+2}) \le 4\exp(-u^2) , \quad (2.65)$$

using e.g. in the last inequality that $2^{n+2} \ge 2^{k+2} + n - k \ge 4u^2 + n - k$. Consequently, $\mathsf{P}(\Omega_1 \cup \Omega_2) \ge 1 - 5 \exp(-u^2)$, and on this event we have

$$|X_t - X_{\pi_k(t)}| \le \sum_{n \ge k} 2^{n/2+2} \Delta(A_n(t)) \le L\gamma_2(T, d) ,$$

so that $|X_t - X_{t_0}| \le |X_t - X_{\pi_k(t)}| + |X_{\pi_k(t)} - X_{t_0}| \le L\gamma_2(T, d) + Lu\Delta(T)$. \Box

Let us note in particular that, using (2.24),

$$\left(\mathsf{E}\sup_{s,t\in T} |X_s - X_t|^p\right)^{1/p} \le L\sqrt{p}\gamma_2(T,d) .$$
(2.66)

Needless to say that we will look for extensions of Theorem 2.7.13. We can prove right away a particularly elegant result (due independently to R. Latała and S. Mendelson). Let us consider a process $(X_t)_{t \in T}$, which is assumed to be centered but need not to be symmetric. For $n \ge 1$ consider the distance δ_n on T given by $\delta_n(s,t) = ||X_s - X_t||_{2^n}$. Denote by $\Delta_n(A)$ the diameter of a subset A of T for the distance δ_n .

Theorem 2.7.14. Consider an admissible sequence $(\mathcal{A}_n)_{n\geq 0}$ of partitions of T. Then

$$\mathsf{E}\sup_{s,t\in T} |X_s - X_t| \le L \sup_{t\in T} \sum_{n\ge 0} \Delta_n(A_n(t)) .$$
(2.67)

Moreover, given u > 0 and the largest integer k with $2^k \leq u^2$ we have

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge L\Delta_k(T) + \sup_{t\in T} \sum_{n\ge 0} \Delta_n(A_n(t)) \Big) \le L \exp(-u^2) . \quad (2.68)$$

Proof. The increment condition (2.4) will be replaced by the following. For a r.v. Y and $p \ge 1$ we have

$$\mathsf{P}(|Y| \ge u) \le \mathsf{P}(|Y|^p \ge u^p) \le \left(\frac{\|Y\|_p}{u}\right)^p.$$
(2.69)

Let us then consider the points $\pi_n(t)$ as usual. For $u \ge 1$ let us consider the event Ω_u defined by¹⁵

$$\forall n \ge 1 , |X_{\pi_n(t)} - X_{\pi_{n+1}(t)}| \le u \Delta_n(A_n(t)) ,$$
 (2.70)

so that by the union bound and (2.69) for $u \ge 4$ we have

$$\mathsf{P}(\Omega_u^c) \le \sum_{n\ge 1} \left(\frac{2}{u}\right)^{2^n} \le \sum_{k\ge 2} \left(\frac{2}{u}\right)^k \le \frac{L}{u^2} \,. \tag{2.71}$$

On Ω_u summation of the inequalities (2.70) for $n \ge 1$ yields $\sup_{t \in T} |X_t - X_{\pi_1(t)}| \le Lu \sum_{n>1} \Delta_n(A_n(t))$. Combining with (2.71) we obtain

$$\mathsf{E}\sup_{t\in T} |X_t - X_{\pi_1(t)}| \le L \sum_{n\ge 1} \Delta_n(A_n(t)) \;.$$

Since $\operatorname{\mathsf{E}sup}_{t\in T} |X_{\pi_1(t)} - X_{\pi_0(t)}| \leq L\Delta_0(T)$ we have $\operatorname{\mathsf{E}sup}_{t\in T} |X_t - X_{\pi_0(t)}| \leq L\sum_{n\geq 0} \Delta_n(A_n(t))$ and (2.67) follows. The proof of (2.68) is nearly identical to the proof of (2.61) and is left to the reader. \Box

 $^{^{15}}$ We are following here the general method outlined at the end of Section 2.4.

2.8 Functionals

Given a metric space (T, d) how do we calculate $\gamma_2(T, d)$? Of course there is no free lunch. The quantity $\gamma_2(T, d)$ reflects a highly non-trivial geometric characteristic of the metric space. This geometry must be understood in order to compute $\gamma_2(T, d)$. There are unsolved problems in this book (such as Conjecture 17.1.4) which boil down to estimating $\gamma_2(T, d)$ for a certain metric space.

In this section we introduce *functionals*, which are an efficient way to bring up the geometry of a metric space and to build competent admissible sequences, providing upper bounds for $\gamma_2(T, d)$. We will say that a map Fis a *functional* on a set T if, to each subset H of T it associates a number $F(H) \ge 0$, and if it is increasing, i.e.

$$H \subset H' \subset T \Rightarrow F(H) \le F(H'). \tag{2.72}$$

Intuitively a functional is a measure of "size" for the subsets of T. It allows to identify which subsets of T are "large" for our purposes. A first example is given by $F(H) = \Delta(H)$. In the same direction, a fundamental example of a functional is

$$F(H) = \gamma_2(H, d)$$
. (2.73)

A second example, equally important, is the quantity

$$F(H) = \mathsf{E}\sup_{t \in H} X_t \tag{2.74}$$

where $(X_t)_{t \in T}$ is a given process indexed by T and satisfying (2.4).

For our purposes the relevant property of functionals is by no means intuitively obvious yet (but we shall soon see that the functional (2.73) does enjoy this property). Let us first try to explain it in words: if a set is the union of many small pieces far enough from each other, then this set is significantly larger (as measured by the functional) than the *smallest* of its pieces. "Significantly larger" depends on the scale of the pieces, and on their number. This property will be called a "growth condition".

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

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

This factor 2 is a nuisance. It is qualitatively the same to say that a set is contained in a ball of small radius or has small diameter, but quantitatively we have to account for this factor 2. In countless constructions we will produce sets A which are "small" because they are contained in a ball of small radius r. Either we keep track of this property, which is cumbersome, or we control the size of A through its diameter and we deal with this inelegant factor 2. We have chosen here the second method.¹⁶

¹⁶ The opposite choice was made in [171].

What do we mean by "small pieces far from each other"? There is a scale a > 0 at which this happens, and a parameter $r \ge 8$ which gives us some room. The pieces are small at that scale: they are contained in balls with radius 2a/r.¹⁷ The balls are far from each other: any two centers of such balls are at mutual distance $\ge a$. The reason why we require $r \ge 8$ is that we want the following: two points taken in different balls with radius 2a/r whose centers are at distance $\ge a$ cannot be too close to each other. This would not be true for, say, r = 4, so we give ourselves some room, and take $r \ge 8$. Here is the formal definition.

Definition 2.8.1. Given a > 0 and an integer $r \ge 8$ we say that subsets H_1, \ldots, H_m of T are (a, r)-separated if

$$\forall \ell \le m \,, \, H_\ell \subset B(t_\ell, 2a/r) \,, \tag{2.75}$$

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

$$\forall \ell, \, \ell' \le m, \, \ell \ne \ell' \Rightarrow a \le d(t_\ell, t_{\ell'}) \le 2ar \,. \tag{2.76}$$

A secondary feature of this definition is that the small pieces H_{ℓ} are not only well separated (on a scale *a*), but they are in the "same region of *T*" (on the larger scale *ra*). This is the content of the last inequality in condition (2.76).

Exercise 2.8.2. Find interesting examples of metric spaces for which there are no points t_1, \ldots, t_m as in (2.76), for all large enough values of m.

Now, what does "the union of the pieces is significantly larger than the *smallest* of these pieces" mean? This is an "additive property", not a multiplicative one. In this first version of the growth condition, it means that the size of this union is larger than the size of the smallest piece by a quantity $a\sqrt{\log N}$ where N is the number of pieces.¹⁸ Well, sometimes it will only be larger by a quantity of say $a\sqrt{\log N}/100$. This is how the parameter c^* below comes into the picture. One could also multiply the functionals by a suitable constant (i.e. $1/c^*$) to always reduce to the case $c^* = 1$ but this is a matter of taste.

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

Definition 2.8.3. We say that the functional F satisfies the growth condition with parameters $r \ge 8$ and $c^* > 0$ if for any integer $n \ge 1$ and any

¹⁷ This coefficient 2 is motivated by the considerations of the previous paragraph. ¹⁸ We remind the reader that the function $\sqrt{\log y}$ arises from the fact that it is the inverse of the function $\exp(x^2)$.

a > 0 the following holds true, where $m = N_n$. For each collection of subsets H_1, \ldots, H_m of T that are (a, r)-separated we have

$$F\left(\bigcup_{\ell \le m} H_\ell\right) \ge c^* a 2^{n/2} + \min_{\ell \le m} F(H_\ell) .$$
(2.77)

This definition is motivated by the fundamental fact that when $(X_t)_{t\in T}$ is a Gaussian process, the functional (2.74) satisfies a form of the growth condition, see Proposition 2.10.8 below.

The following illustrates how we might use the first part of (2.76).

Exercise 2.8.4. Let (T, d) be isometric to a subset of \mathbb{R}^k provided with the distance induced by a norm. Prove that in order to check that a functional satisfies the growth condition of Definition 2.8.3, it suffices to consider the values of n for which $N_{n+1} \leq (1+2r)^k$. Hint: it follows from (2.47) that for larger values of n and $m = N_n$ there are no points t_1, \ldots, t_m as in (2.76).

You may find it hard to give simple examples of functionals which satisfy the growth condition (2.77). It will become gradually apparent that this condition imposes strong restrictions on the metric space (T, d) and in particular a control from above of the quantity $\gamma_2(T, d)$. It bears repeating that $\gamma_2(T, d)$ reflects the geometry of the space (T, d). Once this geometry is understood, it is usually possible to guess a good choice for the functional F. Many examples will be given in subsequent chapters.

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

Proposition 2.8.5. Assume $r \ge 16$. Then the functional $F(H) = \gamma_2(H, d)$ satisfies the growth condition with parameters r and $c^* = 1/8$.

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

$$\gamma_2(H,d) \ge \frac{1}{8}a2^{n/2} + \min_{\ell \le m} \gamma_2(H_\ell,d)$$
 (2.78)

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

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

Picking for $\ell \in I_n$ an arbitrary element $A \in \mathcal{A}_{n-1}$ with $A \subset H_\ell$ defines a oneto-one map from I_n to \mathcal{A}_{n-1} . Thus card $I_n \leq \text{card } \mathcal{A}_{n-1} \leq N_{n-1} < m = N_n$. Hence there exists $\ell_0 \notin I_n$. Next, we prove that for $t \in H_{\ell_0}$ we have

$$\Delta(A_{n-1}(t)) \ge \Delta(A_{n-1}(t) \cap H_{\ell_0}) + \frac{1}{4}a.$$
(2.79)

Since $\ell_0 \notin I_n$, we have $A_{n-1}(t) \notin H_{\ell_0}$, so that since $A_{n-1}(t) \subset H$, the set $A_{n-1}(t)$ must intersect a set $H_{\ell} \neq H_{\ell_0}$, and consequently it intersects the ball

 $B(t_{\ell}, a/8)$. Since $t \in H_{\ell_0}$ we have $d(t, B(t_{\ell}, a/8)) \ge a/2$. Since $t \in A_{n-1}(t)$ this implies that $\Delta(A_{n-1}(t)) \ge a/2$. This proves (2.79) since $\Delta(A_{n-1}(t) \cap H_{\ell_0}) \le \Delta(H_{\ell_0}) \le a/4$.

Now, since for each $k \ge 0$ we have $\Delta(A_k(t)) \ge \Delta(A_k(t) \cap H_{\ell_0})$, we have

$$\sum_{k\geq 0} 2^{k/2} (\Delta(A_k(t)) - \Delta(A_k(t) \cap H_{\ell_0})))$$

$$\geq 2^{(n-1)/2} (\Delta(A_{n-1}(t) - \Delta(A_{n-1}(t) \cap H_{\ell_0})))$$

$$\geq \frac{1}{4} a 2^{(n-1)/2} ,$$

where we have used (2.79) in the last inequality, and, consequently,

$$\sum_{k\geq 0} 2^{k/2} \Delta(A_k(t)) \geq \frac{1}{4} a 2^{(n-1)/2} + \sum_{k\geq 0} 2^{k/2} \Delta(A_k(t) \cap H_{\ell_0}) .$$
 (2.80)

Next, consider the admissible sequence (\mathcal{A}'_n) of H_{ℓ_0} given by $\mathcal{A}'_n = \{A \cap H_{\ell_0}; A \in \mathcal{A}_n\}$. We have by definition

$$\sup_{t \in H_{\ell_0}} \sum_{k \ge 0} 2^{k/2} \Delta(A_k(t) \cap H_{\ell_0}) \ge \gamma_2(H_{\ell_0}, d)$$

Hence, taking the supremum over t in H_{ℓ_0} in (2.80) we get

$$\sup_{t \in H_{\ell_0}} \sum_{k \ge 0} 2^{k/2} \Delta(A_k(t)) \ge \frac{1}{4} a 2^{(n-1)/2} + \gamma_2(H_{\ell_0}, d) \ge \frac{1}{8} a 2^{n/2} + \min_{\ell \le m} \gamma_2(H_\ell, d) + \frac{1}{4} a 2^{n/2} + \frac{1}{4} a 2^$$

Since the admissible sequence (\mathcal{A}_n) is arbitrary, we have proved (2.78). \Box

2.9 Partitioning Schemes

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

Theorem 2.9.1. Assume that there exists on T a functional F which satisfies the growth condition of Definition 2.8.3 with parameters r and c^* . Then¹⁹

$$\gamma_2(T,d) \le \frac{Lr}{c^*} F(T) + Lr\Delta(T) .$$
(2.81)

This theorem and its generalizations form the backbone of this book. The essence of this theorem is that it produces (by actually constructing them) a sequence of partitions that witnesses the inequality (2.81). For this reason, it could be called "the fundamental partitioning theorem."

¹⁹ It is certain that as r grows, we must obtain a weaker result. The dependence of the right-hand side of (2.81) on r is not optimal. It may be improved with further work.

Exercise 2.9.2. Consider a metric space T consisting of exactly two points. Prove that the functional given by F(H) = 0 for each $H \subset T$ satisfies the growth condition of Definition 2.8.3 for r = 8 and any $c^* > 0$. Explain why we cannot replace (2.81) by the inequality $\gamma_2(T, d) \leq LrF(T)/c^*$.

Let us first stress the following trivial fact (connected to Exercise 2.5.9 (a)). It will be used many times. The last statement of (a) is particularly useful.

Lemma 2.9.3. (a) Consider an integer N. If we cannot cover T by at most N-1 balls of radius a then there exist points $(t_{\ell})_{\ell \leq N}$ with $d(t_{\ell}, t_{\ell'}) \geq a$ for $\ell \neq \ell'$. In particular if $e_n(T) > a$ we can find points $(t_{\ell})_{\ell \leq N_n}$ with $d(t_{\ell}, t_{\ell'}) \geq a$ for $\ell \neq \ell'$.

(b) Assume that any sequence $(t_{\ell})_{\ell \leq m}$ with $d(t_{\ell}, t_{\ell'}) \geq a$ for $\ell \neq \ell'$ satisfies $m \leq N$. Then T can be covered by N balls of radius a.

(c) Consider points $(t_{\ell})_{\ell \leq N_n+1}$ such that $d(t_{\ell}, t_{\ell'}) \geq a$ for $\ell \neq \ell'$. Then $e_n(T) \geq a/2$.

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

(b) You can view this either as a reformulation of (a) or argue directly that when m is taken as large as possible the balls $B(t_{\ell}, a)$ cover T.

(c) If T is covered by sets $(B_{\ell'})_{\ell' \leq N_n}$, by the pigeon hole principle at least two of the points t_{ℓ} must fall into one of these sets, which therefore cannot be a ball of radius < a/2.

The admissible sequence of partitions witnessing (2.81) will be constructed by recursive application of the following basic principle.

Lemma 2.9.4. Under the conditions of Theorem 2.9.1 consider $B \subset T$ with $\Delta(B) \leq 2r^{-j}$ for a certain $j \in \mathbb{Z}$ and consider any $n \geq 0$. Let $m = N_n$. Then we can find a partition $(A_\ell)_{\ell \leq m}$ of B into sets which have either of the following properties:

$$\Delta(A_\ell) \le 2r^{-j-1} , \qquad (2.82)$$

or else

$$t \in A_{\ell} \Rightarrow F(B \cap B(t, 2r^{-j-2})) \le F(B) - c^* 2^{n/2} r^{-j-1}$$
. (2.83)

In words, the piece of the partitions have two further properties. Either (case (2.82)) we have reduced the bound on their diameter from $2r^{-j}$ for B to $2r^{-j-1}$, or else we have no new information on the diameter, but we have gathered the information (2.83).

Proof. Consider the set

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

Consider points $(t_{\ell})_{\ell \leq m'}$ in C such that $d(t_{\ell}, t_{\ell'}) \geq r^{-j-1}$ for $\ell \neq \ell'$. We prove that m' < m. For otherwise, using (2.77) for $a = r^{-j-1}$ and for the sets $H_{\ell} := B \cap B(t_{\ell}, 2r^{-j-2})$ shows that

$$F(B) \ge F\left(\bigcup_{\ell \le m} H_\ell\right) \ge c^* r^{-j-1} 2^{n/2} + \min_{\ell \le m} F(H_\ell) > F(B)$$

This contradiction proves that m' < m. Consequently, using Lemma 2.9.3 (b) for N = m - 1 we may cover C by m' < m balls $(B_\ell)_{\ell \le m'}$ of radius $\le r^{-j-1}$. We then set $A_\ell = C \cap (B_\ell \setminus \bigcup_{\ell' < \ell} B_{\ell'})$ for $\ell \le m'$, $A_\ell = \emptyset$ for $m' < \ell < m$ and $A_m = B \setminus C$.

So, in picturesque terms, Lemma 2.9.4 produces many small pieces and (possibly) a large one (on which one has further information).

Before we start the proof of Theorem 2.9.1 we need the following technical fact which will be used many times: the sum of a geometric series is basically of the size of either its first or its last term.

Lemma 2.9.5. Consider numbers $(a_n)_{n\geq 0}$, $a_n > 0$, and assume $\sup_n a_n < \infty$. Consider $\alpha > 1$ and define

$$I = \left\{ n \ge 0 \; ; \; \forall k \ge 0 \; , k \ne n \; , \; a_k < a_n \alpha^{|n-k|} \right\} \; . \tag{2.84}$$

Then $I \neq \emptyset$ and we have

$$\sum_{k\geq 0} a_k \leq \frac{2\alpha}{\alpha - 1} \sum_{n \in I} a_n .$$
(2.85)

Proof. Let us write $k \prec n$ when $a_k \leq a_n \alpha^{-|n-k|}$. This relation is a partial order: if $k \prec n$ and $n \prec p$ then $a_k \leq a_p \alpha^{-|n-k|-|n-p|} \leq a_p \alpha^{-|k-p|}$, so that $k \prec p$. We can then restate the definition of I:

$$I = ig\{ n \geq 0 \; ; \; orall k \geq 0 \; , n \prec k \Rightarrow n = k ig\} \; .$$

In words, I is the set of elements n of \mathbb{N} that are maximal for the partial order \prec .

Next, we prove that for each k in \mathbb{N} there exists $n \in I$ with $k \prec n$. Indeed otherwise we can recursively construct an infinite sequence $n_1 = n \prec n_2 \prec \cdots$ and this is absurd because $a_{n_{\ell+1}} \ge \alpha a_{n_{\ell}}$ and we assume that the sequence (a_n) is bounded.

Thus for each k in \mathbb{N} there exists $n \in I$ with $k \prec n$. Then $a_k \leq a_n \alpha^{-|n-k|}$, and therefore

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

Proof of Theorem 2.9.1. There is no question that this proof is the most demanding up to this point. The result is however absolutely central, on its own and also because several of our main results will follow the same overall scheme of proof.

We have to construct an admissible sequence of partitions which witnesses the inequality (2.81). The construction of this sequence is as simple as it could be: we recursively use Lemma 2.9.4. More precisely, we construct an admissible sequence of partitions \mathcal{A}_n and for $A \in \mathcal{A}_n$ we construct an integer $j_n(A) \in \mathbb{Z}$ with

$$\Delta(A) \le 2r^{-j_n(A)} . \tag{2.86}$$

We start with $\mathcal{A}_0 = \{T\}$ and $j_0(T)$ the largest integer $j_0 \in \mathbb{Z}$ with $\Delta(T) \leq 2r^{-j_0}$, so that $2r^{-j_0} \leq r\Delta(T)$. Having constructed \mathcal{A}_n we construct \mathcal{A}_{n+1} as follows. For each $B \in \mathcal{A}_n$, we use Lemma 2.9.4 with $j = j_n(B)$ to split B into sets $(A_\ell)_{\ell \leq N_n}$. If A_ℓ satisfies (2.82) we set $j_{n+1}(A_\ell) = j_n(B) + 1$ and otherwise (since we have no new information on the diameter) we set $j_{n+1}(A_\ell) = j_n(B)$. Thus, in words, $j_{n+1}(A_\ell) = j_n(B) + 1$ if A_ℓ is a small piece of B and $j_{n+1}(A_\ell) = j_n(B)$ if A_ℓ is the large piece of B.

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

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

$$t \in A \Rightarrow F(B \cap B(t, 2r^{-j_{n+1}(A)-2})) \le F(B) - c^* 2^{n/2} r^{-j_{n+1}(A)-1}$$
. (2.87)

Now we start the hard part of the proof, proving that the sequence of partitions we just constructed witnesses (2.81). For this we fix $t \in T$. We want to prove that

$$\sum_{k\geq 0} 2^{n/2} \Delta(A_n(t)) \leq \frac{Lr}{c^*} F(T) + Lr \Delta(T) \; .$$

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

$$\sum_{n\geq 0} a(n) \leq \frac{Lr}{c^*} F(T) + Lr \Delta(T) . \qquad (2.88)$$

First, we prove a side result, that for $n \ge 0$ we have

$$a(n) \le \frac{Lr}{c^*} F(T) + L\Delta(T) .$$
(2.89)

If $n \ge 1$ and j(n-1) = j(n) then using (2.87) for n-1 rather than n yields (2.89). Next, if $n \ge 1$ and j(n-1) = j(n) - 1 then $a(n) = \sqrt{2}r^{-1}a(n-1) \le a(n-1)$ since $r \ge 8$, and iterating this relation until we reach an integer n' with either j(n'-1) = j(n') or n' = 0 proves (2.89) since $a(0) \le L\Delta(T)$.

In particular the sequence (a(n)) is bounded. Consider then the set I as provided by Lemma 2.9.5 for $\alpha = \sqrt{2}$ and $a_n = a(n)$, that is

$$I = \left\{ n \ge 0 \; ; \; \forall k \ge 0 \; , n \ne k \; , \; a(k) < a(n) 2^{|k-n|/2} \right\}$$

Recalling that $a(0) = r^{-j_0} \leq r \Delta(T)/2$, it suffices to prove that

$$\sum_{n \in I \setminus \{0\}} a(n) \le \frac{Lr}{c^*} F(T) . \tag{2.90}$$

For $n \in I$, $n \ge 1$ we have $a(n+1) < \sqrt{2}a(n)$ and $a(n-1) < \sqrt{2}a(n)$. Since $a(n+1) = \sqrt{2}r^{j(n)-j(n+1)}a(n)$ this implies

$$j(n+1) = j(n) + 1$$
; $j(n-1) = j(n)$. (2.91)

Proving (2.90) is the difficult part. Assuming first that I is infinite (the important case), let us enumerate the elements of $I \setminus \{0\}$ as $n_1 < n_2 < \ldots$ so that (2.91) implies

$$j(n_k + 1) = j(n_k) + 1$$
; $j(n_k - 1) = j(n_k)$. (2.92)

In words, n_k is at the end of a sequence of partitions steps in which $A_{\ell+1}(t)$ was the large piece of $A_{\ell}(t)$, and $A_{n_k+1}(t)$ is a small piece of $A_{n_k}(t)$. Let us note that as a consequence of (2.92) we have

$$j_{n_{k+1}} \ge j_{n_k+1} \ge j_{n_k} + 1$$
.

The key to the proof is to show that for $k \ge 1$ we have

$$a(n_k) \le \frac{Lr}{c^*} (f(n_k - 1) - f(n_{k+2}))$$
(2.93)

where $f(n) = F(A_n(t))$. Now the sequence (f(n)) is decreasing because $A_n(t) \subset A_{n-1}(t)$, and f(0) = F(T). When $k \ge 2$ then $f(n_k - 1) \le f(n_{k-1})$, so that (2.93) implies

$$a(n_k) \le \frac{Lr}{c^*} (f(n_{k-1}) - f(n_{k+2}))$$
 (2.94)

Summation of the inequalities (2.94) for $k \ge 2$ then yields

$$\sum_{k \ge 2} a(n_k) \le \frac{Lr}{c^*} F(T) , \qquad (2.95)$$

and combining with (2.93) for k = 1 proves (2.90) and concludes the proof of the theorem when I is infinite.

We now prove (2.93). Since $n_k \ge 1$ we may define $n^* := n_k - 1$. By (2.92) we have $j(n_k - 1) = j(n_k)$, i.e. $j(n^*) = j(n^* + 1)$. We may then use (2.87) for $B = A_{n^*}(t)$, $A = A_{n_k}(t) = A_{n^*+1}(t)$ to obtain that

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

Recalling that $n^* = n_k - 1$ this means

$$F(B \cap B(t, 2r^{-j_{n_k}(A)-2})) \le F(B) - c^* 2^{(n_k-1)/2} r^{-j_{n_k}(A)-1} , \qquad (2.96)$$

so that

$$a(n_k) \le \frac{Lr}{c^*} \left(F(B) - F(B \cap B(t, 2r^{-j_{n_k}(A) - 2})) \right).$$
(2.97)

Furthermore, by (2.92),

$$j(n_{k+2}) \ge j(n_{k+1}) + 1 \ge j(n_k) + 2$$
. (2.98)

Since $j(n_{k+2}) = j_{n_{k+2}}(A_{n_{k+2}}(t)), (2.86)$ implies $\Delta(A_{n_{k+2}}(t)) \le 2r^{-j(n_{k+2})} \le$ $2r^{-j(n_k)-2}$ so that

$$A_{n_{k+2}}(t) \subset B \cap B(t, 2r^{-j(n_k)-2}) , \qquad (2.99)$$

and thus $f(n_{k+2}) = F(A_{n_{k+2}}(t)) \leq F(B \cap B(t, 2r^{-j_{n_k}(A)-2}))$. Combining with (2.97) and since $F(B) = f(n_k - 1)$ we have proved (2.93).

Assuming now that I is finite, it has a largest element $n_{\bar{k}}$. We use the previous argument to control $a(n_k)$ when $k+2 \leq \overline{k}$, and for $k = \overline{k} - 1$ and $k = \overline{k}$ we simply use (2.89).

It is important for the sequel that you fully master the previous argument.

Exercise 2.9.6. We say that a sequence $(F_n)_{n\geq 0}$ of functionals on (T,d)satisfies the growth condition with parameters $r \ge 8$ and $c^* > 0$ if

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

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

$$F_n\left(\bigcup_{\ell \le m} H_\ell\right) \ge c^* a 2^{n/2} + \min_{\ell \le m} F_{n+1}(H_\ell) .$$
 (2.100)

Prove that then

$$\gamma_2(T,d) \le \frac{Lr}{c^*} F_0(T) + Lr\Delta(T)$$
 (2.101)

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

Proposition 2.9.7. Consider a metric space (T, d), and for $n \ge 0$, consider subsets T_n of T with card $T_0 = 1$ and card $T_n \leq N_n$ for $n \geq 1$. Consider a number S and let

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

Then $\gamma_2(U, d) \leq LS$.

Proof. For $H \subset U$ we define $F(H) = \inf \sup_{t \in H} \sum_{n \geq 0} 2^{n/2} d(t, V_n)$ where the infimum is taken over all choices of $V_n \subset T$ with card $V_n \leq N_n$. It is important here not to assume that $V_n \subset H$ to ensure that F is increasing. We then prove that F satisfies the growth condition by an argument very similar to that of Proposition 2.8.5. The proof follows from Theorem 2.9.1 since $\Delta(U, d) \leq 2S$, as each point of U is within distance S of the unique point of T_0 .

A slightly different partitioning scheme has recently been discovered by R. van Handel [59], and we describe a variant of it now. We consider a metric space (T, d) and an integer $r \ge 8$. We assume that for $j \in \mathbb{Z}$ we are given a function $s_j(t) \ge 0$ on T.

Theorem 2.9.8. Assume that the following holds.

For each subset A of T, for each
$$j \in \mathbb{Z}$$
 with $\Delta(A) \leq 2r^{-j}$ and for each $n \geq 1$, then either $e_n(A) \leq r^{-j-1}$ or else there exists $t \in A$ with $s_j(t) > 2^{n/2}r^{-j-1}$. (2.102)

Then

$$\gamma_2(T,d) \le Lr\Big(\Delta(T) + \sup_{t \in T} \sum_{j \in \mathbb{Z}} s_j(t)\Big) .$$
(2.103)

We will show later how to construct functions $s_j(t)$ satisfying (2.102) using a functional which satisfies the growth condition.²⁰

The right-hand side of (2.103) is the supremum over t of a sum of terms. It need not always be the same terms which will contribute the most for different values of t, and the bound is definitely better than if the supremum and the summation were exchanged.

Proof of Theorem 2.9.8. Consider the largest $j_0 \in \mathbb{Z}$ with $\Delta(T) \leq 2r^{-j_0}$, so that $2r^{-j_0} \leq r\Delta(T)$. We construct by induction an increasing sequence of partitions \mathcal{A}_n with card $\mathcal{A}_n \leq N_n$, and for $A \in \mathcal{A}_n$ we construct an integer $j_n(A) \in \mathbb{Z}$ with $\Delta(A) \leq 2r^{-j_n(A)}$. We start with $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$ and $j_0(T) = j_1(T) = j_0$.

Once \mathcal{A}_n has been constructed $(n \geq 1)$, we further split every element $B \in \mathcal{A}_n$. The idea is to first split B into sets which are basically level sets for the function $s_j(t)$ in order to achieve the crucial relation (2.107) below, and then to further split each of these sets according to its metric entropy. More precisely, we may assume that $S = \sup_{t \in T} \sum_{j \in \mathbb{Z}} s_j(t) < \infty$, for there is nothing to prove otherwise. Let us set $j = j_n(B)$ and define the sets A_k for $1 \leq k \leq n$ by setting for k < n

$$A_k = \{ t \in B \; ; \; 2^{-k}S < s_j(t) \le 2^{-k+1}S \} \; , \tag{2.104}$$

 $^{^{20}}$ See [59] for other constructions.

and

$$A_n = \{ t \in B ; \ s_j(t) \le 2^{-n+1}S \} .$$
(2.105)

The purpose of this construction is to ensure the following:

$$k \le n ; t, t' \in A_k \Rightarrow s_j(t') \le 2(s_j(t) + 2^{-n}S) .$$
 (2.106)

This is obvious since $s_j(t') \leq 2s_j(t)$ for k < n and $s_j(t') \leq 2^{-n+1}S$ if k = n. For each set A_k , $k \leq n$ we use the following procedure.

- If $e_{n-1}(A_k) \leq r^{-j-1}$ we may cover A_k by at most N_{n-1} balls of radius $2r^{-j-1}$, so we may split A_k into N_{n-1} pieces of diameter $\leq 4r^{-j-1}$. We decide that each of these pieces A is an element of \mathcal{A}_{n+1} , for which we set $j_{n+1}(A) = j + 1$. Thus $\Delta(A) \leq 4r^{-j-1} = 4r^{-j_{n+1}(A)}$.
- Otherwise we decide that $A_k \in \mathcal{A}_{n+1}$ and we set $j_{n+1}(A_k) = j$. Thus $\Delta(A_k) \leq 2r^{-j} = 2r^{-j_{n+1}(A)}$. From (2.102) there exists $t' \in A_k$ for which $s_j(t') \geq 2^{(n-1)/2}r^{-j-1}$. Then by (2.106) we have

$$\forall t \in A_k \; ; \; 2^{(n-1)/2} r^{-j-1} \le 2(s_j(t) + 2^{-n}S) \; .$$
 (2.107)

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

- either $j_{n+1}(A) = j_n(B) + 1$
- or else $j_{n+1}(A) = j_n(B)$ and, from (2.107)

$$\forall t \in A \; ; \; 2^{(n-1)/2} r^{-j_{n+1}(A)-1} \le 2(s_{j_n(B)}(t) + 2^{-n}S) \; . \tag{2.108}$$

This completes the construction. Now for $n \ge 1$ we have $n \le N_{n-1}$ so that card $\mathcal{A}_{n+1} \le nN_{n-1}N_n \le N_{n+1}$ and the sequence (\mathcal{A}_n) is admissible. Next, we fix $t \in T$. We set $j_n = j_n(\mathcal{A}_n(t))$, and we observe that by construction $j_n \le j_{n+1} \le j_n + 1$. Since $\Delta(\mathcal{A}_n(t)) \le 4r^{-j_n(t)}$ we have $2^{n/2}\Delta(\mathcal{A}_n(t)) \le 4a(n)$ where $a(n) := 2^{n/2}r^{-j_n(t)}$. To complete the argument we prove that

$$\sum_{n\geq 0} a(n) \leq Lr(S + \Delta(T)) . \tag{2.109}$$

For this consider the set I provided by Lemma 2.9.5 for $\alpha = \sqrt{2}$, so that since $r^{-j_0} \leq 2r\Delta(T)$ it suffices to prove that

$$\sum_{n \in I \setminus \{0\}} a(n) \le LrS . \tag{2.110}$$

For $n \in I \setminus \{0\}$, it holds that $j_{n-1} = j_n < j_{n+1}$ (since otherwise this contradicts the definition of I). In particular, the integers j_n for $n \in I$ are all different so that $\sum_{n\geq 0} s_{j_n}(t) \leq S$. Using (2.108) for n-1 instead of n yields $2^{(n-2)/2}r^{-j_n-1} \leq 2(s_{j_{n-1}}(t)+2^{-n+1}S)$. Since $j_{n-1}=j_n$ we get

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

and summing these relations we obtain the desired result.

The following connects Theorems 2.9.1 and 2.9.8.

γ

Proposition 2.9.9. Assume that the functional F satisfies the growth condition with parameters r and c^* . Then the functions

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

satisfy (2.102).

Proof. Consider a subset A of T, $j \in \mathbb{Z}$ with $\Delta(A) \leq 2r^{-j}$ and $n \geq 1$. Let $m = N_n$. If $e_n(A) > r^{-j-1}$ then by Lemma 2.9.3 we may find $(t_\ell)_{\ell \leq m}$ in A with $d(t_\ell, t_{\ell'}) \geq r^{-j-1}$ for $\ell \neq \ell'$. Consider the set $H_\ell = B(t_\ell, 2r^{-j-2})$ so that by (2.77) used for $a = r^{-j-1}$ it holds that

$$F\left(\bigcup_{\ell \le m} H_{\ell}\right) \ge c^* r^{-j-1} 2^{n/2} + \min_{\ell \le m} F(H_{\ell}) .$$
 (2.111)

Let us now consider $\ell_0 \leq m$ such that $F(H_{\ell_0})$ achieves the minimum in the right-hand side, so that $\min_{\ell \leq m} F(H_\ell) = F(B(t_{\ell_0}, 2r^{-j-2}))$. The crude inequality $2r^{-j-2} + 2r^{-j} \leq 2r^{-j+1}$ implies that $H_\ell \subset B(t_{\ell_0}, 2r^{-j+1})$ for each ℓ , so that $F(\bigcup_{\ell \leq m} H_\ell) \leq F(B(t_{\ell_0}, 2r^{-j+1}))$. Then (2.111) implies

$$F(B(t_{\ell_0}, 2r^{-j+1})) \ge c^* r^{-j-1} 2^{n/2} + F(B(t_{\ell_0}, 2r^{-j-2}))$$
i.e. $s_j(t_{\ell_0}) \ge 2^{n/2} r^{-j-1}$.

Despite the fact that the proof of Theorem 2.9.8 is a few lines shorter than the proof of Theorem 2.9.1, in the various generalizations of this principle we will mostly follow the scheme of proof of Theorem 2.9.1. The reason for this choice is simple: it should help the reader that our various partition theorems follow a common pattern. The most difficult partition theorem we present is Theorem 6.2.8 (the Latała-Bednorz theorem), which is one of the highlights of this work, and it is not clear at this point whether the method of Theorem 2.9.8 can be adapted to the proof of this theorem.

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

Lemma 2.9.10. Consider $\alpha > 0$, an integer $\tau \ge 0$ and an increasing sequence of partitions $(\mathcal{B}_n)_{n>0}$ with card $\mathcal{B}_n \le N_{n+\tau}$. Let

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

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

$$\sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} \Delta(A_n(t)) \leq 2^{\tau/\alpha} (S + K(\alpha) \Delta(T)) .$$
 (2.112)

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

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

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

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

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

Exercise 2.9.11. Prove that (2.112) might fail if one replaces the right-hand side by $K(\alpha, \tau)S$. Hint: S does not control $\Delta(T)$.

2.10 Gaussian Processes: The Majorizing Measure Theorem

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

$$d(s,t) = \left(\mathsf{E}(X_s - X_t)^2\right)^{1/2}.$$
 (2.113)

Recall the functional γ_2 of Definition 2.7.3.

Theorem 2.10.1. (The Majorizing Measure Theorem.) For a universal constant L it holds that

$$\frac{1}{L}\gamma_2(T,d) \le \mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T,d) .$$
(2.114)

The reason for the name is explained in Section 3.1. We will meditate on this statement in Section 2.12. We will spend much time trying to generalize this theorem to other classes of processes. To link the statements of these generalizations with that of (2.114) it may be good to reformulate the lower bound $\gamma_2(T, d) \leq \mathsf{LE}\sup_{t \in T} X_t$ in the following general terms:

The control from above of $\mathsf{E}\sup_{t\in T} X_t$ implies the existence of a "small" sequence of admissible partitions of T.

The right-hand side inequality in (2.114) is Theorem 2.7.11. To prove the lower bound we will use Theorem 2.9.1 and the functional

$$F(H) = \mathsf{E}\sup_{t \in H} X_t := \sup_{H^* \subset H, H^* \text{finite}} \mathsf{E}\sup_{t \in H^*} X_t . \tag{2.115}$$

For this we need to prove that this functional satisfies the growth condition with c^* a universal constant and to bound $\Delta(T)$. We strive to give a proof that relies on general principles, and lends itself to generalizations. Lemma 2.10.2. (Sudakov minoration) Assume that

$$\forall p,q \leq m \,, \quad p \neq q \quad \Rightarrow d(t_p,t_q) \geq a \,.$$

Then we have

$$\mathsf{E}\sup_{p\leq m} X_{t_p} \geq \frac{a}{L_1}\sqrt{\log m} \ . \tag{2.116}$$

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

The proof of the Sudakov minoration is given just after Lemma 15.2.7.

Exercise 2.10.3. Prove that Lemma 2.10.2 is equivalent to the following statement. If $(X_t)_{t \in T}$ is a Gaussian process, and d is the canonical distance, then

$$e_n(T,d) \le L2^{-n/2} \mathsf{E} \sup_{t \in T} X_t$$
 (2.117)

Compare with Exercise 2.7.8.

To understand the relevance of Sudakov minoration, let us consider the case where $\mathsf{E}X_{t_p}^2 \leq 100a^2$ (say) for each *p*. Then (2.116) means that the bound (2.15) is of the correct order in this situation.

Exercise 2.10.4. Prove (2.116) when the r.v.s X_{t_p} are independent. That is, assume that these variables are Gaussian independent centered. Hint: use the method of Exercise 2.3.7 (b).

Exercise 2.10.5. A natural approach ("the second moment method") to prove that $\mathsf{P}(\sup_{p \leq m} X_{t_p} \geq u)$ is at least 1/L for a certain value of u is as follows. Consider the r.v. $Y = \sum_p \mathbf{1}_{\{X_{t_p} \geq u\}}$, prove that $\mathsf{E}Y^2 \leq L(\mathsf{E}Y)^2$, and then use the Paley-Zygmund inequality (6.15) below to prove that $\sup_{p \leq m} X_{t_p} \geq a\sqrt{\log m}/L_1$ with probability $\geq 1/L$. Prove that this approach works when the r.v.s X_{t_ℓ} are independent, but find examples showing that this naive approach does not work in general to prove (2.116).

The following is a very important property of Gaussian processes, and one of the keys to Theorem 2.10.1. It is a facet of the theory of concentration of measure, a leading idea of modern probability theory. We refer the reader to [74] to learn about this.

Lemma 2.10.6. Consider a Gaussian process $(X_t)_{t \in U}$, where U is finite and let $\sigma = \sup_{t \in U} (\mathsf{E} X_t^2)^{1/2}$. Then for $u \ge 0$ we have

$$\mathsf{P}\Big(\Big|\sup_{t\in U} X_t - \mathsf{E}\sup_{t\in U} X_t\Big| \ge u\Big) \le 2\exp\left(-\frac{u^2}{2\sigma^2}\right).$$
(2.118)

In words, the size of the fluctuations of $\sup_{t \in U} X_t$ are governed by the size of the individual r.v.s X_t , rather than by the (typically much larger) quantity $\mathsf{E} \sup_{t \in U} X_t$. It is essential that the cardinality of U does not appear in (2.118).

Exercise 2.10.7. Find an example of a Gaussian process for which

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

whereas the fluctuations of $\sup_{t \in T} X_t$ are of order σ , e.g. the variance of $\sup_t X_t$ is about σ^2 . Hint: $T = \{(t_i)_{i \leq n}; \sum_{i \leq n} t_i^2 \leq 1\}$ and $X_t = \sum_{i \leq n} t_i g_i$ where g_i are independent standard Gaussian r.v.s

Proposition 2.10.8. Consider points $(t_{\ell})_{\ell \leq m}$ in 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

$$\mathsf{E}\sup_{t\in H} X_t \ge \frac{a}{L_1}\sqrt{\log m} - L_2\sigma\sqrt{\log m} + \min_{\ell \le m} \mathsf{E}\sup_{t\in H_\ell} X_t \ . \tag{2.119}$$

When $\sigma \leq a/(2L_1L_2)$, (2.119) implies

$$\mathsf{E}\sup_{t\in H} X_t \ge \frac{a}{2L_1}\sqrt{\log m} + \min_{\ell \le m} \mathsf{E}\sup_{t\in H_\ell} X_t , \qquad (2.120)$$

which can be seen as a generalization of Sudakov's minoration (2.116) by taking $H_{\ell} = \{t_{\ell}\}$. When $m = N_n$, (2.120) proves that the functional $F(H) = \mathsf{E} \sup_{t \in T} X_t$ satisfies the growth condition (2.77).

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

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

For $t \in H_{\ell}$ we set $Z_t = X_t - X_{t_{\ell}}$. Since $H_{\ell} \subset B(t_{\ell}, \sigma)$ we have $\mathsf{E}Z_t^2 = d(t, t_{\ell})^2 \leq \sigma^2$ and, for $u \geq 0$ equation (2.118) used for the process $(Z_t)_{t \in H_{\ell}}$ implies

$$\mathsf{P}(|Y_{\ell} - \mathsf{E}Y_{\ell}| \ge u) \le 2\exp\left(-\frac{u^2}{2\sigma^2}\right).$$
(2.121)

Thus if $V = \max_{\ell \leq m} |Y_{\ell} - \mathsf{E}Y_{\ell}|$ then combining (2.121) and the union bound, we get

$$\mathsf{P}(V \ge u) \le 2m \exp\left(-\frac{u^2}{2\sigma^2}\right),\tag{2.122}$$

and (2.13) implies

$$\mathsf{E}V \le L_2 \sigma \sqrt{\log m} \ . \tag{2.123}$$

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

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

and thus

59

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

so that

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

Taking expectations we obtain

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

and we use (2.116) and (2.123).

Exercise 2.10.9. Prove that (2.120) might fail if one allows $\sigma = a$. Hint: the intersection of the balls $B(t_{\ell}, a)$ might contain a ball with positive radius.

Exercise 2.10.10. Consider subsets $(H_{\ell})_{\ell \leq m}$ of B(0, a) and $H = \bigcup_{\ell \leq m} H_{\ell}$. Prove that

$$\mathsf{E}\sup_{t\in H} X_t \le La\sqrt{\log m} + \max_{\ell \le m} \mathsf{E}\sup_{t\in H_\ell} X_t .$$
(2.124)

Try to find improvements on this bound. Hint: peek at (19.61) below.

Proof of Theorem 2.10.1. We fix $r = \max(8, 4L_1L_2)$, so that $2a/r \le a/2L_1L_2$. The growth condition for the functional F of (2.115) follows from (2.120), which implies that (2.77) holds for $c^* = 1/L$. Theorem 2.9.1 implies

$$\gamma_2(T,d) \leq L\mathsf{E}\sup_{t\in T} X_t + L\Delta(T) \;.$$

To control the term $\Delta(T)$ we write that for $t_1, t_2 \in H$,

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

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

The proof of Theorem 2.10.1 displays an interesting feature. This theorem aims at understanding $\mathsf{E} \sup_{t \in T} X_t$, and for this we use functionals that are based on precisely this quantity. This is not a circular argument. The content of Theorem 2.10.1 is that there is simply no other way to bound a Gaussian process than to control the quantity $\gamma_2(T, d)$. The miracle of this theorem is that it relates in complete generality two quantities, namely $\mathsf{E} \sup_{t \in T} X_t$ and $\gamma_2(T, d)$ which are both very hard to estimate. Still, in concrete situations, to estimate these quantities, we must in some way gain understanding of the underlying geometry.

The following is a noteworthy consequence of Theorem 2.10.1.

Theorem 2.10.11. 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 increment condition

$$\forall u > 0, \ \forall s, t \in T, \ \mathsf{P}(|Y_s - Y_t| \ge u) \le 2 \exp\left(-\frac{u^2}{d(s,t)^2}\right),$$
 (2.125)

where d is the distance (2.113) associated to the process X_t . Then we have

$$\mathsf{E}\sup_{s,t\in T}|Y_s - Y_t| \le L\mathsf{E}\sup_{t\in T} X_t \ . \tag{2.126}$$

Processes satisfying the condition (2.125) are sometimes called sub-Gaussian. We will see many examples later, see (6.2) below.

Proof. We combine (2.60) with the left-hand side of (2.114).

Let us also note the following consequence of (2.126) and Lemma 2.2.1.²¹

Corollary 2.10.12. Consider two Gaussian processes $(X_t)_{t \in T}$ and $(Y_t)_{t \in T}$. Assume that

$$\forall s, t \in T$$
 , $\mathsf{E}(Y_s - Y_t)^2 \le \mathsf{E}(X_s - X_t)^2$.

Then

$$\mathsf{E}\sup_{t\in T} Y_t \le L\mathsf{E}\sup_{t\in T} X_t \ . \tag{2.127}$$

2.11 Gaussian Processes as Subsets of a Hilbert Space

In this section we learn to think of a Gaussian process as a subset of a Hilbert space. This will reveal our lack of understanding of basic geometric questions.

First, consider a Gaussian process $(Y_t)_{t\in T}$ and assume (the only case which is of interest to us) that there is a countable set $T' \subset T$ which is dense in T. We view each Y_t as a point in the Hilbert space $L^2(\Omega, \mathsf{P})$ where (Ω, P) is the basic probability space. The closed linear span of the r.v.s $(Y_t)_{t\in T}$ in $L^2(\Omega, \mathsf{P})$ is a separable Hilbert space, and the map $t \mapsto Y_t$ is an isometry from (T, d) to its image (by the very definition of the distance d). In this manner we associate a subset of a Hilbert space to each Gaussian process.

Conversely, consider a separable Hilbert space, which we may assume to be $\ell^2 = \ell^2(\mathbb{N})$. Consider an independent sequence $(g_i)_{i\geq 1}$ of standard Gaussian r.v.s We can then define the Gaussian process $(X_t)_{t\in\ell^2}$, where

$$X_t = \sum_{i>1} t_i g_i \tag{2.128}$$

(the series converges in $L^2(\Omega)$). Thus

 $^{^{21}}$ Please do not miss the comments at the end of Section 2.16.
$$\mathsf{E}X_t^2 = \sum_{i \ge 1} t_i^2 = \|t\|^2 . \tag{2.129}$$

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 by (2.129).

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

Theorem 2.11.1. For a subset T of ℓ^2 , we have

$$\gamma_2(\operatorname{conv} T) \le L\gamma_2(T) . \tag{2.130}$$

Of course we also have $\gamma_2(T) \leq \gamma_2(\operatorname{conv} T)$ since $T \subset \operatorname{conv} T$.

Proof. To prove (2.130) we observe that since $X_{a_1t_1+a_2t_2} = a_1X_{t_1} + a_2X_{t_2}$ we have

$$\sup_{\in \operatorname{conv} T} X_t = \sup_{t \in T} X_t . \tag{2.131}$$

We then use (2.114) to write

$$\frac{1}{L}\gamma_2(\operatorname{conv} T) \le \mathsf{E}\sup_{t\in\operatorname{conv} T} X_t = \mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T) \ . \qquad \Box$$

A basic problem is that it is absolutely not obvious how to construct an admissible sequence of partitions on conv T witnessing (2.130).

Research problem 2.11.2. Give a geometrical proof of (2.130).

What we mean by geometrical proof is a proof that does not use Gaussian processes but only the geometry of Hilbert space. The difficulty of the problem is that the structure of an admissible sequence which witnesses that $\gamma_2(\operatorname{conv} T) \leq L\gamma_2(T)$ must depend on the "geometry" of the set T. A really satisfactory argument would give a proof that holds in Banach spaces more general than Hilbert space, for example by providing a positive answer to the following, where the concept of q-smooth Banach space is explained in [80].

Research problem 2.11.3. Consider a 2-smooth Banach space, and the distance d induced by its norm. Is it true that for each subset T of its unit ball one has $\gamma_2(\operatorname{conv} T, d) \leq K\sqrt{\log \operatorname{card} T}$? More generally, is it true that for each finite subset T one has $\gamma_2(\operatorname{conv} T, d) \leq K\gamma_2(T, d)$? (Here K may depend on the Banach space, but not on T.)

Research problem 2.11.4. Still more generally, is it true that for a finite subset T of a q-smooth Banach space, one has $\gamma_q(\operatorname{conv} T) \leq K \gamma_q(T)$?

Even when the Banach space is ℓ^p , I do not know the answer to these problems (unless p = 2!). (The Banach space ℓ^p is 2-smooth for $p \ge 2$ and q-smooth for p < 2, where 1/p + 1/q = 1.) One concrete case is when the set T consists of the first N vectors of the unit basis of ℓ^p . It is possible to show in this case that $\gamma_q(\operatorname{conv} T) \le K(p)(\log N)^{1/q}$, where 1/p + 1/q = 1. We leave this as a challenge to the reader. The proof here is pretty much the same as for the case p = q = 2 which was covered in Section 2.6.

Exercise 2.11.5. Prove that if $a \ge 2$ we have $\sum_{k>1} (k+1)^{-a} \le L2^{-a}$.

We recall the ℓ^2 norm $\|\cdot\|$ of (2.129). Here is a simple fact.

Proposition 2.11.6. Consider a sequence $(t_k)_{k\geq 1}$ such that.

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

Let $T = \{\pm t_k, k \ge 1\}$. Then $\mathsf{E} \sup_{t \in T} X_t \le L$ and thus $\mathsf{E} \sup_{t \in \operatorname{conv} T} X_t \le L$ by (2.131).

Proof. We have

$$\mathsf{P}\Big(\sup_{k\geq 1} |X_{t_k}| \geq u\Big) \leq \sum_{k\geq 1} \mathsf{P}(|X_{t_k}| \geq u) \leq \sum_{k\geq 1} 2\exp\left(-\frac{u^2}{2}\log(k+1)\right) \quad (2.132)$$

since X_{t_k} is Gaussian with $\mathsf{E} X_{t_k}^2 \leq 1/\log(k+1)$. For $u \geq 2$, the right-hand side of (2.132) is at most $L \exp(-u^2/L)$ by the result of Exercise 2.11.5, and as usual the conclusion follows from (2.6).

Exercise 2.11.7. Deduce Proposition 2.11.6 from (2.34). Hint: Use Exercise 2.4.1.

It is particularly frustrating not to be able to solve the following special instance of Problem 2.11.2.

Research problem 2.11.8. In the setting of Proposition 2.11.6 find a geometrical proof that $\gamma_2(\operatorname{conv} T) \leq L$.

The following shows that the situation of Proposition 2.11.6 is in a sense generic.

Theorem 2.11.9. Consider a countable set $T \subset \ell^2$, with $0 \in T$. Then we can find a sequence (t_k) with

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

and

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

Furthermore we may assume that each t_k is a multiple of the difference of two elements of T.²²

 22 This information is of secondary importance and will be used only much later.

Proof. By Theorem 2.10.1 we can find an admissible sequence (\mathcal{A}_n) of T with

$$\forall t \in T, \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \le L \mathsf{E} \sup_{t \in T} X_t := S.$$
(2.133)

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 $T = \bigcup_{n \ge 0} T_n$ and that $T_0 = \{0\}$. (To do this, we simply enumerate the elements of T as $(v_n)_{n\ge 1}$ with $v_0 = 0$ and we ensure that v_n is in T_n .) For $n \ge 1$ consider the set U_n that consists of all the points

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

where $t \in T_n, v \in T_{n-1}$, and $t \neq v$. Thus each element of U_n has norm $2^{-n/2}$, and U_n has at most $N_n N_{n-1} \leq N_{n+1}$ elements. Let $U = \bigcup_{n\geq 1} U_n$. Then since $\sum_{\ell\leq n} N_{\ell+1} \leq N_{n+2}$, U contains at most N_{n+2} elements of norm $\geq 2^{-n/2}$. We enumerate U as $\{z_k; k = 1, \ldots\}$ where the sequence $(||z_k||)$ is non-increasing, so that $||z_k|| < 2^{-n/2}$ for $k > N_{n+2}$. Let us now prove that $||z_k|| \leq L/\sqrt{\log(k+1)}$. If $k \leq N_2$ this holds because $||z_k|| \leq 1$. Assume then that $k > N_2$ and let $n \geq 0$ be the largest integer with $k > N_{n+2}$. Then by definition of n we have $k \leq N_{n+3}$ and thus $2^{-n/2} \leq L/\sqrt{\log k}$. But then $||z_k|| \leq 2^{-n/2} \leq L/\sqrt{\log k}$, proving the required inequality.

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

$$t = \sum_{1 \le n \le m} \pi_n(t) - \pi_{n-1}(t) = \sum_{1 \le n \le m} a_n(t) u_n(t) , \qquad (2.134)$$

with $a_n(t) = 2^{n/2} \|\pi_n(t) - \pi_{n-1}(t)\|$ and

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

Since

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

and since $u_n(t) \in U_n \subset U$ we see from (2.134) that

$$t = \sum_{1 \le n \le m} a_n(t) u_n(t) + (2S - \sum_{1 \le n \le m} a_n(t)) \times 0 \in 2S \operatorname{conv}(U \cup \{0\}) .$$

Thus $T \subset 2S \operatorname{conv}(U \cup \{0\}) = \operatorname{conv}(2SU \cup \{0\})$, and it suffices to take $t_k = 2Sz_k$.

Exercise 2.11.10. What is the purpose of the condition $0 \in T$?

Exercise 2.11.11. Prove that if $T \subset \ell^2$ and $0 \in T$, then (even when T is not countable) we can find a sequence (t_k) in ℓ^2 , with $||t_k|| \sqrt{\log(k+1)} \leq L \mathsf{E} \sup_{t \in T} X_t$ for all k and

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

where $\overline{\text{conv}}$ denotes the closed convex hull. (Hint: do the obvious thing, apply Theorem 2.11.9 to a dense countable subset of *T*.) Denoting now by $\text{conv}^*(A)$ the set of infinite sums $\sum_i \alpha_i a_i$ where $\sum_i |\alpha_i| = 1$ and $a_i \in A$, prove that one can also achieve

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

Exercise 2.11.12. Consider a set $T \subset \ell^2$ with $0 \in T \subset B(0, \delta)$. Prove that we can find a sequence (t_k) in ℓ^2 , with the following properties:

$$\forall k \ge 1 , \ \|t_k\| \sqrt{\log(k+1)} \le L \mathsf{E} \sup_{t \in T} X_t ,$$
 (2.135)

$$|t_k|| \le L\delta , \qquad (2.136)$$

$$T \subset \overline{\operatorname{conv}}\{t_k \; ; \; k \ge 1\} \;, \tag{2.137}$$

where $\overline{\text{conv}}$ denotes the closed convex hull. Hint: copy the proof of Theorem 2.11.9, observing that since $T \subset B(0, \delta)$ one may chose $\mathcal{A}_n = \{T\}$ and $T_n = \{0\}$ for $n \leq n_0$, where n_0 is the smallest integer for which $2^{n_0/2} \geq \delta^{-1}\mathsf{E}\sup_{t \in T} X_t$, and thus $U_n = \emptyset$ for $n \leq n_0$.

The following problems are closely related to Problem (2.11.2).

Research problem 2.11.13. Give a geometric proof of the following fact. Given subsets $(T_k)_{k \leq N}$ of a Hilbert space, and $T = \sum_{k \leq N} T_k = \{x_1 + \ldots + x_N; \forall k \leq N, x_k \in T_k\}$, prove that $\gamma_2(T) \leq L \sum_{k \leq N} \gamma_2(T_k)$.

We do not even know how to solve the following special case.

Research problem 2.11.14. Consider sets $(T_k)_{k \leq N}$ in a Hilbert space, and assume that each T_k consists of M vectors of length 1. Let $T = \sum_{k \leq N} T_k$. Give a geometrical proof of the fact that $\gamma_2(T) \leq LN\sqrt{\log M}$.

The next exercise is inspired by the paper [9] of S Artstein. It is more elaborate, and may be omitted on first reading. A Bernoulli r.v. ε is such that $P(\varepsilon = \pm 1) = 1/2$.²³

Exercise 2.11.15. Consider a subset $T \subset \mathbb{R}^n$, where \mathbb{R}^n is provided with the Euclidean distance. We assume that for some $\delta > 0$, we have

$$0 \in T \subset B(0,\delta) . \tag{2.138}$$

²³ One must distinguish Bernoulli r.v.s ε_i from positive numbers $\epsilon_k!$

Consider independent Bernoulli r.v.s $(\varepsilon_{i,p})_{i,p\geq 1}$. Given a number $q\leq n$ consider the operator $U_q: \mathbb{R}^n \to \mathbb{R}^q$ given by

$$U_q(x) = \left(\sum_{i \le n} \varepsilon_{i,p} x_i\right)_{p \le q}$$

(a) Prove that $||U_q|| \ge \sqrt{n}$.

We want to prove that despite (a) there exists a number L such that if $\mathsf{E}\sup_{t\in T}\sum_{i\leq n}g_it_i\leq \delta\sqrt{q}$, then with high probability

$$U_q(T) \subset B(0, L\delta\sqrt{q}) , \qquad (2.139)$$

whereas from (2.138) we would not expect better than $U_q(T) \subset B(0, \delta\sqrt{n})$. (b) Use the subgaussian inequality (6.1.1) to prove that if ||x|| = 1, then

$$\mathsf{E}\exp\left(\frac{1}{4}\left(\sum_{i\leq n}\varepsilon_{i,p}x_i\right)^2\right)\leq L.$$
(2.140)

(c) Use (2.140) and independence to prove that for $x \in \mathbb{R}^n$ and $v \ge 1$,

$$\mathsf{P}(\|U_q(x)\| \ge Lv\sqrt{q}\|x\|) \le \exp(-v^2q) .$$
(2.141)

(d) Use (2.141) to prove that with probability close to 1, for each of the vectors t_k of Exercise 2.11.12 one has $||U_q(t_k)|| \leq L\delta\sqrt{q}$ and conclude.

We end this section with a discussion of a question which shares some features with Problem 2.11.2, in a sense that it is a property which is obvious on one hand, but difficult to prove without using the magic of linearity.²⁴ For $k \leq N$ let us consider Gaussian processes $(X_{k,t})_{t \in T_k}$ with associated distances d_k . On the space $T = \prod_{k \leq N} T_k$ let us consider the distance d given by

$$d((t_k)_{k \le N}, (t'_k)_{k \le N}) = \left(\sum_{k \le N} d_k (t_k, t'_k)^2\right)^{1/2}.$$
 (2.142)

Proposition 2.11.16. We have

$$\gamma_2(T,d) \le L \sum_{k \le N} \gamma_2(T_k,d_k) . \tag{2.143}$$

Proof. Assuming without loss of generality that the processes $(X_k)_{k \leq N}$ are independent we can consider the Gaussian process $(X_t)_{t \in T}$ given for $t = (t_k)_{k \leq N}$ by $X_t = \sum_{k \leq N} X_{k,t_k}$. It is obvious that the distance d of (2.143) is associated to this process. It is also obvious that

$$\sup_{t\in T} X_t = \sum_{k\leq N} \sup_{t\in T_k} X_{t_k}$$

Taking expectation and combining with (2.114) concludes the proof. \Box

²⁴ There are several equally frustrating instances of this situation.

The question now is to prove (2.143) without using Gaussian processes, for example by proving it for any sequence $((T_k, d_k))_{k \leq N}$ of metric spaces. The most interesting part of that project is that it is unexpectedly hard. Is it the sign that we are still missing an important ingredient? In the next exercise we show how to prove about the simplest possible case of (2.143), which is already pretty challenging.

Exercise 2.11.17. Throughout this exercise each space T_k consists of M_k points, and the mutual distance of any two different points of T_k is $\epsilon_k > 0$. The goal is to prove the inequality

$$\int \sqrt{\log N(T, d, \epsilon)} d\epsilon \le L \sum_{k \le N} \epsilon_k \sqrt{\log M_k} .$$
(2.144)

Throughout the exercise I denotes a subset of $\{1, \ldots, N\}$ and I^c denotes its complement.

(a) Prove that if $\sum_{k \in I} \epsilon_k^2 \leq \epsilon^2$ then $N(T, d, \epsilon) \leq \prod_{k \in I^c} M_k$. (b) Show that to prove (2.144) it suffices to prove the following. Consider two sequences $(\epsilon_k)_{k\leq N}$ and $(\eta_k)_{k\leq N}$. For $\epsilon > 0$ define $S(\epsilon)$ by

$$S(\epsilon)^2 = \inf \left\{ \sum_{k \in I^c} \eta_k^2 \ ; \ \sum_{k \in I} \epsilon_k^2 \le \epsilon^2 \right\} \,,$$

where the infimum is over all choices of I. Then

$$\int_0^\infty S(\epsilon) \mathrm{d}\epsilon \le L \sum_{k \le N} \epsilon_k \eta_k .$$
(2.145)

(c) To prove (2.145), show that it suffices to prove the following. Consider a function h > 0 on a probability space. For $\epsilon > 0$ define $S(\epsilon)$ by

$$\tilde{S}(\epsilon)^2 = \inf \left\{ \int_{A^c} \frac{1}{h} \mathrm{d}\mu \; ; \; \int_A h \mathrm{d}\mu \leq \epsilon^2 \right\} \, ,$$

where the infimum is over all choices of A. Then $\int_0^\infty \tilde{S}(\epsilon) d\epsilon \leq L$. Hint: Reduce to the case where $\sum_{k < N} \epsilon_k \eta_k = 1$. Use the probability μ on $\{1, \ldots, N\}$ such

that $\mu(\{k\}) = \epsilon_k \eta_k$ and the function h given by $h(k) = \epsilon_k / \eta_k$. (d) Show that it suffices to prove that $\sum_{\ell \in \mathbb{Z}} 2^{-\ell} \tilde{S}(2^{-\ell}) \leq L$. (e) Assuming for simplicity that μ has no atoms²⁵ prove the statement given in (d). Hint: For $\ell \in \mathbb{Z}$ and $2^{-2\ell} \leq \int h d\mu$ consider the set A_ℓ of the type $A_{\ell} = \{h \leq t_{\ell}\}$ where t_{ℓ} is such that $\int_{A_{\ell}} h d\mu = 2^{-2\ell}$, so that $\tilde{S}(2^{-\ell})^2 \leq 1$ $\int_{A_{e}^{c}}(1/h)d\mu$. Warning: this is not easy.

 $^{^{25}}$ I am sure that this is true without this hypothesis, but I did not find the energy to carry out the details.

Exercise 2.11.18. This exercise continues the previous one. The spaces (T_k, d_k) are now any metric spaces, and the goal is to prove that

$$\int_0^\infty \sqrt{\log N(T, d, \epsilon)} d\epsilon \le L \sum_{k \le N} \int_0^\infty \sqrt{\log N(T_k, d_k, \epsilon)} d\epsilon .$$
 (2.146)

Proving this requires passing the main difficulty of proving (2.143), but to prove (2.143) itself, it will be convenient to use different tools, and that proof is the object of Exercise 3.1.6.

(a) Show that to prove (2.146) it suffices to prove the following. Consider decreasing functions $f_k : \mathbb{R}^+ \to \mathbb{R}^+$ and for $\epsilon > 0$ define $V(\epsilon)$ by

$$V(\epsilon)^2 = \inf\left\{\sum_{k \le N} f_k(\epsilon_k)^2 \; ; \; \sum_{k \le N} \epsilon_k^2 \le \epsilon^2\right\} \, ,$$

where the infimum is taken over all families $(\epsilon_k)_{k \leq N}$. Then

$$\int_0^\infty V(\epsilon) \mathrm{d}\epsilon \le L \sum_{k \le N} \int_0^\infty f_k(\epsilon) \mathrm{d}\epsilon \;. \tag{2.147}$$

(b) When each f_k is of the type $f_k = \eta_k \mathbf{1}_{[0,\theta_k[}$, deduce (2.146) from (2.145). (c) Convince yourself that by approximation it suffices to consider the case where each f_k is a finite sum $\sum_{\ell} 2^{-\ell} \mathbf{1}_{[0,\theta_{k\ell}[}$.

(d) In the case (c) prove (2.147) by applying the special case (b) to the family $f_{k,\ell}$ of functions given by $f_{k,\ell} := 2^{-\ell} \mathbf{1}_{[0,\theta_{k,\ell}[}$ for all relevant values of k,ℓ . Hint: this is a bit harder.

2.12 Dreams

We may reformulate the inequality (2.114)

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

of Theorem 2.10.1 by the statement

Chaining suffices to explain the size of a Gaussian process. (2.148)

We simply mean that the "natural" chaining bound for the size of a Gaussian process (i.e. the right-hand side inequality in (2.114)) is of correct order, *provided* one uses the best possible chaining. This is what the left-hand side of (2.114) shows. We may dream of removing the word "Gaussian" in that statement. The desire to achieve this lofty goal in as many situations as possible motivates much of the rest of the book.

Besides the generic chaining, we have found in Theorem 2.11.9 another optimal way to bound Gaussian processes: to put them into the convex hull of a "small" process, that is to use the inequality

$$\mathsf{E} \sup_{t \in T} X_t \le L \inf \{ S \; ; \; T \subset \operatorname{conv}\{t_k, k \ge 1\}, \|t_k\| \le S/\sqrt{\log(k+1)} \} \; .$$

Since we do not really understand the geometry of going from a set to its convex hull, it is better (for the time being) to consider this method as somewhat distinct from the generic chaining. Let us try to formulate it in a way which is suitable for generalizations. Given a countable set \mathcal{V} of r.v.s, let us define the (possibly infinite) quantity

$$S(\mathcal{V}) = \inf\left\{S > 0 \; ; \; \int_{S}^{\infty} \sum_{V \in \mathcal{V}} \mathsf{P}(|V| > u) \mathrm{d}u \le S\right\}.$$

$$(2.149)$$

Lemma 2.12.1. It holds that

$$\mathsf{E}\sup_{V\in\operatorname{conv}\mathcal{V}}|V| \le 2S(\mathcal{V}) \ . \tag{2.150}$$

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

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

Thus (2.150) provides a method to bound stochastic processes. This method may look childish, but for Gaussian processes, the following reformulation of Theorem 2.11.9 shows that it is in fact optimal.

Theorem 2.12.2. Consider a countable set T. Consider a Gaussian process $(X_t)_{t\in T}$ and assume that $X_{t_0} = 0$ for some $t_0 \in T$. Then there exists a countable set \mathcal{V} of Gaussian r.v.s, each of which is a multiple of the difference of two variables X_t , with

$$\forall t \in T \; ; \; X_t \in \operatorname{conv} \mathcal{V} \; , \tag{2.151}$$

$$S(\mathcal{V}) \le L \mathsf{E} \sup_{t \in \mathcal{T}} X_t \ . \tag{2.152}$$

To understand the need of the condition $X_{t_0} = 0$ for some t_0 think of the case where T consists of one single point. The proof of Theorem 2.12.2 is nearly obvious by using (2.132) to bound $S(\mathcal{V})$ for the set \mathcal{V} consisting of the variables X_{t_k} for the sequence (t_k) constructed in Theorem 2.11.9. We may dream of proving statements such as Theorem 2.12.2 for many classes of processes.

Also worthy of detailing is another remarkable geometric consequence of Theorem 2.11.9 in a somewhat different direction. Consider an integer N.

Considering i.i.d. standard Gaussian r.v.s, we define as usual the process $X_t = \sum_{i \leq N} g_i t_i$. We may view an element t of ℓ_N^2 as a function on ℓ_N^2 by the canonical duality, and therefore view t as a r.v. on the probability space (ℓ_N^2, μ) , where μ is the law of the sequence $(g_i)_{i \leq N}$. The processes (X_t) and (t) have the same law, hence they are really the same object viewed in two different ways. Consider a subset T of ℓ_N^2 , and assume that $T \subset \operatorname{conv}\{t_k; k \geq 1\}$. Then for any v > 0 we have

$$\left\{\sup_{t\in T} t > v\right\} \subset \bigcup_{k\geq 1} \left\{t_k > v\right\}.$$
(2.153)

The sets $\{t_k \ge v\}$ on the right are very simple: they are *half-spaces*. Assume now that for $k \ge 1$ and a certain S we have $||t_k|| \sqrt{\log(k+1)} \le S$. Then for $u \ge 2$

$$\sum_{k \ge 1} \mu(\{t_k \ge Su\}) \le \sum_{k \ge 1} \exp\left(-\frac{u^2}{2}\log(k+1)\right) \le L \exp(-u^2/L) ,$$

the very same computation as in (2.132). Theorem 2.11.9 implies that one may find such t_k for $S = L \mathsf{E} \sup_t X_t$. Therefore for $v \ge L \mathsf{E} \sup_t X_t$, the fact that the set in the left-hand side of (2.153) is small (in the sense of probability) may be witnessed by the fact that this set can be covered by a union of *simple sets* (half-spaces) the *sum* of the probabilities of which is small.

We may dream that something similar occurs in many other settings. In Chapter 13, which can be read right now, we will meet a fundamental such setting, which inspired the author's lifetime favorite problem, see Section 13.3.

2.13 A First Look at Ellipsoids

We have illustrated the gap between Dudley's bound (2.41) and the sharper bound (2.34), using the examples (2.49) and (2.42). These examples might look artificial, but here we demonstrate that the gap between Dudley's bound (2.41) and the generic chaining bound (2.34) already exists for *ellipsoids* in Hilbert space. Truly understanding ellipsoids will be fundamental in several subsequent questions, such as the matching theorems of Chapter 4. A further study of ellipsoids is proposed in Section 3.2.

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

$$\mathcal{E} = \left\{ t \in \ell^2 \; ; \; \sum_{i \ge 1} \frac{t_i^2}{a_i^2} \le 1 \right\} \,. \tag{2.154}$$

Proposition 2.13.1. When $\sum_{i>1} a_i^2 < \infty$ we have

$$\frac{1}{L} \left(\sum_{i \ge 1} a_i^2 \right)^{1/2} \le \mathsf{E} \sup_{t \in \mathcal{E}} X_t \le \left(\sum_{i \ge 1} a_i^2 \right)^{1/2}.$$
(2.155)

Proof. The Cauchy-Schwarz inequality implies

$$Y := \sup_{t \in \mathcal{E}} X_t = \sup_{t \in \mathcal{E}} \sum_{i \ge 1} t_i g_i \le \left(\sum_{i \ge 1} a_i^2 g_i^2 \right)^{1/2} .$$
(2.156)

Taking $t_i = a_i^2 g_i / (\sum_{j \ge 1} a_j^2 g_j^2)^{1/2}$ yields that actually $Y = (\sum_{i \ge 1} a_i^2 g_i^2)^{1/2}$ and thus $\mathsf{E}Y^2 = \sum_{i \ge 1} a_i^2$. The right-hand side of (2.155) follows from the Cauchy-Schwarz inequality:

$$\mathsf{E}Y \le (\mathsf{E}Y^2)^{1/2} = \left(\sum_{i \ge 1} a_i^2\right)^{1/2}.$$
 (2.157)

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

$$\mathsf{E}X_t^2 = \sum_i t_i^2 \le \max_i a_i^2 \sum_j \frac{t_j^2}{a_j^2} \le \sigma^2 .$$
 (2.158)

Then (2.118) implies²⁶

$$\begin{split} \mathsf{E}(Y-\mathsf{E}Y)^2 &\leq L\sigma^2 \leq L(\mathsf{E}Y)^2 \;,\\ \text{so that}\; \sum_{i\geq 1}a_i^2 &= \mathsf{E}Y^2 = \mathsf{E}(Y-\mathsf{E}Y)^2 + (\mathsf{E}Y)^2 \leq L(\mathsf{E}Y)^2. \end{split}$$

As a consequence of Theorem 2.10.1,

$$\gamma_2(\mathcal{E}) \le L\left(\sum_{i\ge 1} a_i^2\right)^{1/2}.$$
(2.159)

This statement is purely about the geometry of ellipsoids. The proof we gave was rather indirect, since it involved Gaussian processes. Later on, in Chapter 4 we will learn how to give "purely geometric" proofs of similar statements that will have many consequences.

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

$$2^n \le i \le 2^{n+1} \Rightarrow a_{2^n} \ge a_i \ge a_{2^{n+1}}$$

we get

²⁶ One may extend (2.118) to the case where U is infinite by a proper definition of $\sup_{t \in U} X_t$.

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

and

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

and thus $\sum_{n\geq 0} 2^n a_{2^n}^2 \leq 3 \sum_{i\geq 1} a_i^2$. So we may rewrite (2.155) as

$$\frac{1}{L} \left(\sum_{n \ge 0} 2^n a_{2^n}^2 \right)^{1/2} \le \mathsf{E} \sup_{t \in \mathcal{E}} X_t \le \left(\sum_{n \ge 0} 2^n a_{2^n}^2 \right)^{1/2} . \tag{2.160}$$

Proposition 2.13.1 describes the size of ellipsoids with respect to Gaussian processes. Our next result describes their size with respect to Dudley's entropy bound (2.38).

Proposition 2.13.2. We have

$$\frac{1}{L}\sum_{n\geq 0} 2^{n/2} a_{2^n} \le \sum_{n\geq 0} 2^{n/2} e_n(\mathcal{E}) \le L \sum_{n\geq 0} 2^{n/2} a_{2^n} .$$
(2.161)

The right-hand sides in (2.160) and (2.161) are distinctly different.²⁷ Dudley's bound fails to describe the behavior of Gaussian processes on ellipsoids. This is a simple occurrence of a general phenomenon. In some sense an ellipsoid is smaller than what one would predict just by looking at its entropy numbers $e_n(\mathcal{E})$. This idea will be investigated further in Section 4.1.

Exercise 2.13.3. Prove that for an ellipsoid \mathcal{E} of \mathbb{R}^m one has

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

and that this estimate is essentially optimal. Compare with (2.58).

The proof of (2.161) hinges on ideas which are at least 50 years old, and which relate to the methods of Exercise 2.5.9. The left-hand side is the easier part (it is also the most important for us). It follows from the next lemma, the proof of which is basically a special case of (2.45).

Lemma 2.13.4. We have $e_n(\mathcal{E}) \geq \frac{1}{2}a_{2^n}$.

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

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

²⁷ This difference may seem rather small, but, as we shall see in Chapter 4, there are natural situations where it really matters.

Since \mathcal{E}_n is the image of \mathcal{E} by a contraction²⁸(namely the "projection on the first 2^n coordinates") it holds that $e_n(\mathcal{E}_n) \leq e_n(\mathcal{E})$.

Throughout the rest of this section we 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 card $T \leq 2^{2^n}$, and $\epsilon > 0$; then

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

Since we have assumed that the sequence (a_i) is non-increasing, we have $a_i \geq a_{2^n}$ for $i \leq 2^n$ and thus $a_{2^n}B \subset \mathcal{E}_n$, so that $\operatorname{Vol}\mathcal{E}_n \geq a_{2^n}^{2^n}\operatorname{Vol}B$. Thus, whenever $2\epsilon < a_{2^n}$, we cannot have $\mathcal{E}_n \subset \bigcup_{t \in T} (\epsilon B + t)$, so that $e_n(\mathcal{E}_n) \geq a_{2^n}/2$.

We now turn to the upper bound, which relies on a special case of (2.46). We keep the notation of the proof of Lemma 2.13.4.

Lemma 2.13.5. We have

$$e_{n+3}(\mathcal{E}) \le e_{n+3}(\mathcal{E}_n) + a_{2^n}$$
 (2.162)

Proof. We observe that when $t \in \mathcal{E}$, then, using that $a_i \leq a_{2^n}$ for $i > 2^n$ in the last inequality,

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

so that $(\sum_{i>2^n} t_i^2)^{1/2} \leq a_{2^n}$ and, viewing \mathcal{E}_n as a subset of \mathcal{E} , we have $d(t, \mathcal{E}_n) \leq a_{2^n}$. Thus if for $k \geq 1$ we cover \mathcal{E}_n by N_k balls of radius ϵ , the balls with the same centers but radius $\epsilon + a_{2^n}$ cover \mathcal{E} . This proves that $e_k(\mathcal{E}_n) \leq e_k(\mathcal{E}) + a_{2^n}$ and hence (2.162).

Lemma 2.13.6. Let $\epsilon = \max_{k \leq n} a_{2^k} 2^{k-n}$. Consider a subset A of \mathcal{E}_n with the following property:

any two points of A are at mutual distance
$$\geq 2\epsilon$$
. (2.163)

Then card $A \leq N_{n+3}$.

Proof. The balls centered at the points of A, with radius ϵ , have disjoint interiors, so that the volume of their union is card $A \operatorname{Vol}(\epsilon B)$, and since these balls are entirely contained in $\mathcal{E}_n + \epsilon B$ we have

$$\operatorname{card} A \operatorname{Vol}(\epsilon B) \le \operatorname{Vol}(\mathcal{E}_n + \epsilon B)$$
. (2.164)

²⁸ Generally speaking, a map φ from a metric space (T, d) to a metric space (T', d') is called a contraction if $d'(\varphi(x), \varphi(y)) \leq d(x, y)$.

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

$$\sum_{i \le 2^n} \frac{(t_i + t'_i)^2}{c_i^2} \le \sum_{i \le 2^n} \frac{2t_i^2 + 2t'_i^2}{c_i^2} \le \sum_{i \le 2^n} \frac{1}{2} \left(\frac{t_i^2}{a_i^2} + \frac{t'_i^2}{\epsilon^2} \right) \le 1 ,$$

we have

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

Therefore

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

and comparing with (2.164) yields

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

Next it follows from the choice of ϵ that for any $k \leq n$ we have $a_{2^k} 2^{k-n} \leq \epsilon.$ Then $a_i \leq a_{2^k} \leq \epsilon 2^{n-k}$ for $2^k < i \leq 2^{k+1}$, so that

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

since $\sum_{i>0} i2^{-i} = 4$. Therefore card $A \le 2^{2^n} \cdot 2^{2^{n+2}} \le N_{n+3}$. Lemma 2.13.7. We have

$$e_{n+3}(\mathcal{E}_n) \le 2 \max_{k \le n} (a_{2^k} 2^{k-n})$$
 (2.165)

Proof. Assume now that A is as large as possible under (2.163). Then the balls centered at points of A and with radius $\leq 2\epsilon$ cover \mathcal{E}_n , for otherwise we could add a point to A. Since card $A \leq N_{n+3}$ we have $e_{n+3}(\mathcal{E}_n) \leq 2\epsilon$.

Combining (2.165) with (2.162) we obtain

Corollary 2.13.8. We have

$$e_{n+3}(\mathcal{E}) \le 3 \max_{k \le n} (a_{2^k} 2^{k-n}) .$$
 (2.166)

Proof of Proposition 2.13.2. We have, using (2.166)

$$\sum_{n\geq 3} 2^{n/2} e_n(\mathcal{E}) = \sum_{n\geq 0} 2^{(n+3)/2} e_{n+3}(\mathcal{E}) \le L \sum_{n\geq 0} 2^{n/2} \left(\sum_{k\leq n} 2^{k-n} a_{2^k} \right)$$
$$= L \sum_{k\geq 0} 2^k a_{2^k} \sum_{n\geq k} 2^{-n/2} \le L \sum_{k\geq 0} 2^{k/2} a_{2^k} .$$

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

2.14 Rolling Up our Sleeves: Chaining on Ellipsoids

Let us recall the ellipsoid \mathcal{E} of (2.154). We have proved (2.159) as a consequence of the Majorizing Measure Theorem, Theorem 2.10.1. We will later give a more geometrical proof of this result. In the present section we demonstrate the hard way that these results are deep, by explicitly constructing a chaining on the ellipsoid \mathcal{E}^{29} . This is surprisingly non-trivial.³⁰ Let us assume that the sequence (a_i) is non-increasing, and for $n \geq 0$ let $I_n = \{i; 2^n \leq i < 2^{n+1}\}$ so that card $I_n = 2^n$ and

$$\mathcal{E} \subset \mathcal{E}' = \left\{ t \in \ell^2 \; ; \; \sum_{n \ge 0} \frac{1}{a_{2^n}^2} \sum_{i \in I_n} t_i^2 \le 1 \right\} = \left\{ t \in \ell^2 \; ; \; \sum_{n \ge 0} \frac{2^n}{c_n} \sum_{i \in I_n} t_i^2 \le 1 \right\},$$
(2.167)

where $c_n = 2^n a_{2^n}$. Furthermore (as in the previous section) $\sum_{n\geq 0} c_n = \sum_{n\geq 0} 2^n a_{2^n}^2 \leq 3 \sum_{i\geq 1} a_i^2$. For such an ellipsoid \mathcal{E}' we will construct sets $U_n \subset \ell^2$ with card $U_n \leq N_{n+n_0}$ (where n_0 is a universal constant) such that

$$\forall t \in \mathcal{E}' \ , \ \sum_{n \ge 0} 2^{n/2} d(t, U_n) \le L \left(\sum_{n \ge 0} c_n\right)^{1/2} .$$
 (2.168)

Let us now deduce from this result how to perform the chaining on the ellipsoid \mathcal{E} of (2.154). As we have just seen, such an ellipsoid is contained in an ellipsoid \mathcal{E}' of the type (2.167) for which $\sum_{n\geq 0} c_n \leq L \sum_{i\geq 1} a_i^2$. Consider the sets $U_n \subset \ell^2$ as in (2.168). Consider a map $\varphi : \ell^2 \to \mathcal{E}$ such that $d(x,\varphi(x)) \leq 2d(x,\mathcal{E})$ and observe that for $t \in \mathcal{E}$ and $x \in \ell^2$ we have $d(x,\varphi(x)) \leq 2d(x,t)$ so that $d(t,\varphi(x)) \leq d(t,x) + d(x,\varphi(x)) \leq 3d(t,x)$. Consequently, $d(t,\varphi(U_n)) \leq 3d(t,U_n)$. The sets $\varphi(U_n) \subset \mathcal{E}$ satisfy card $\varphi(U_n) \leq card U_n \leq N_{n+n_0}$. We define $T_n = \{0\}$ for $n \leq n_0$ and $T_n = U_{n-n_0}$ for $n > n_0$. Thus card $T_n \leq N_n$ and (2.168) implies

$$\forall t \in \mathcal{E} \ , \ \sum_{n \ge 0} 2^{n/2} d(t, T_n) \le L \left(\sum_i a_i^2\right)^{1/2}.$$

We now prepare for the construction of the sets U_n . There is no loss of generality to assume that $\sum_{n>0} c_n = 1$.

Lemma 2.14.1. Given $t \in \mathcal{E}'$ we can find a sequence $(p(n,t))_{n\geq 0}$ of integers with the following properties.

$$\sum_{i \in I_n} t_i^2 \le 2^{-p(n,t)} , \qquad (2.169)$$

²⁹ There are obvious similarities between this section and Section 2.6. It is a good challenge to figure out by yourself how to do the chaining on ellipsoids after having studied Section 2.6.

³⁰ I am grateful to Dali Liu for having suggested to include this section

$$\sum_{n\geq 0} 2^{n/2 - p(n,t)/2} \le L , \qquad (2.170)$$

$$\forall n \ge 0, \ p(n+1,t) \le p(n,t) + 2.$$
 (2.171)

Proof. Define q(n,t) as the smallest integer $q \leq 2n$ such that $\sum_{i \in I_n} t_i^2 \leq 2^q$. Let $A = \{n \geq 0; q(n,t) < 2n\}$, so that for $n \in A$ we have $2^{-q(n,t)} \leq 2\sum_{i \in I_n} t_i^2$. Thus, since $t \in \mathcal{E}'$,

$$\sum_{n \in A} \frac{2^{n-q(n,t)}}{c_n} \le 2 \sum_{n \ge 0} \frac{2^n}{c_n} \sum_{i \in I_n} t_i^2 \le 2.$$

Since $\sum_n c_n = 1$ then $\sum_{n \in A} 2^{n/2-q(n,t)/2} \leq L$ by the Cauchy-Schwarz inequality. Since $2^{n/2-q(n,t)/2} = 2^{-n/2}$ for $n \notin A$ we have $\sum_{n \geq 0} 2^{n/2-q(n,t)/2} \leq L$. We define now $p(n,t) = \min\{q(k,t) + 2(n-k); 0 \leq k \leq n\}$ so that (2.171) holds. Also, since $2^{-p(n,t)/2} \leq \sum_{k \leq n} 2^{-q(k,t)/2-(n-k)}$ we obtain

$$\begin{split} \sum_{n\geq 0} 2^{n/2 - p(n,t)/2} &\leq \sum_{n\geq 0} \sum_{k\leq n} 2^{k/2 - q(k,t)/2} 2^{-(n-k)/2} \\ &= \sum_{k\geq 0} 2^{k/2 - q(k,t)/2} \sum_{n\geq k} 2^{-(n-k)/2} \leq L \;. \quad \Box \end{split}$$

For each $n \ge 1$ and $p \ge 0$ consider the set $B(n, p) \subset \ell^2$ which consists of the $t = (t_i)_{i\ge 1}$ such that $t_i = 0$ if $i \ge 2^n$ and $||t||_2 \le 2^{-p/2+2}$. This is a ball of dimension $2^n - 1$ and radius $2^{-p/2+2}$. Using (2.47) for $\epsilon = 1/4$ there is a set $V_{n,p} \subset B(n,p)$ with card $V_{n,p} \le L^{2^n}$ such that every point of B(n,p) is within distance $\le 2^{-p/2}$ of $V_{n,p}$. We consider the set $V_n = \bigcup_{0\le p\le 2n} V_{n,p}$ so that card $V_{n,p} \le L^{2^n}$. We set $U_0 = \{0\}$ and we consider the sets U_n consisting of the elements $x_0 + \ldots + x_n$ where $x_k \in V_k$

For $t \in \ell^2$ and $n \ge 0$ we define $t^{(n)} \in \ell^2$ by $t_i^{(n)} = t_i$ if $i < 2^n$ and $t_i^{(n)} = 0$ if $i \ge 2^n$. Note that $t^{(n)} = t$ for $t \in U_n$.

Lemma 2.14.2. For $t \in \mathcal{E}'$ consider the sequence (p(n,t)) of Lemma 2.14.1. Then for each n we can find $u(n) \in U_n$ such that $d(u(n), t^{(n)}) \leq 2^{-p(n,t)/2}$.

Proof. The proof is by induction over *n*. For *n* = 0 it suffices to take *u*(0) = 0 since *t*⁽⁰⁾ = 0. For the induction step from *n* to *n*+1 we have *t*^(*n*) = *u*(*n*)+*v*(*n*) where *u*(*n*) ∈ *U_n* and $||v(n)||_2 \le 2^{-p(n,t)/2}$, so that $t^{(n+1)} = u(n)+v'(n)$ where $v'(n) = v(n) + t^{(n+1)} - t^{(n)}$. By (2.169) $||t^{(n+1)} - t^{(n)}||_2 = (\sum_{i \in I_n} t_i^2)^{1/2} \le 2^{-p(n,t)/2}$. Thus $||v'(n)||_2 \le 2^{-p(n,t)/2+1} \le 2^{-p(n+1,t)/2+2}$ where we have used (2.171) in the second inequality. Since $v(n)_i = 0$ for $i \ge 2^n$ we have $v'(n)_i = 0$ for $i \ge 2^{n+1}$ so that $v'(n) \in B(n+1, p(n+1,t))$. Thus there is an element $w \in V_{n+1,p(n+1,t)} \subset V_{n+1}$ for which $||v'(n) - w||_2 \le 2^{-p(n+1,t)/2}$. Setting $u(n+1) := u(n) + w \in U_{n+1}$ we then have $t^{(n+1)} - u(n+1) = v'(n) - w$. □ **Corollary 2.14.3.** For $t \in \mathcal{E}'$ we have $\sum_{n \ge 0} 2^{n/2} d(t, U_n) \le L$.

Proof. Recalling (2.169) we have

$$||t - t^{(n)}||_2^2 = \sum_{k>n} \sum_{i \in I_k} t_i^2 \le \sum_{k>n} 2^{-p(k,t)}$$

so that $||t - t^{(n)}||_2 \le \sum_{k>n} 2^{-p(k,t)/2}$. Then

$$\sum_{n\geq 0} 2^{n/2} \|t - t^{(n)}\|_2 \le \sum_{n\geq 0} \sum_{k>n} 2^{-p(k,t)/2} = \sum_{k\geq 1} 2^{-p(k,t)/2} \sum_{0\leq n< k} 2^{n/2} \le L \sum_{k\geq 1} 2^{k/2-p(k,t)/2} \le L$$

using (2.170) in the last inequality. Since $d(t, U_n) \leq d(t^{(n)}, U_n) + ||t - t^{(n)}||_2$ the result follows, using also Lemma 2.14.2.

2.15 Continuity of Gaussian Processes

By far the most important result concerning continuity of Gaussian processes is Dudley's bound (1.19). However since the finiteness of the right-hand side of (1.19) is not necessary for the Gaussian process to be continuous, there are situations where this bound is not appropriate.³¹ In the present section we show that a suitable form of the generic chaining allows us to capture the exact modulus of continuity of a Gaussian process with respect to its canonical distance in full generality. Not surprisingly, the modulus of continuity is closely related to the rate at which the series $\sum_n 2^{n/2} \Delta(A_n(t))$ converges uniformly on T for a suitable admissible sequence (\mathcal{A}_n) . Our first result shows how to obtain a modulus of continuity using the generic chaining.

Lemma 2.15.1. Consider a metric space (T, d) and a process $(X_t)_{t \in T}$ which satisfies the increment condition (2.4):

$$\forall u > 0 , \mathsf{P}(|X_s - X_t| \ge u) \le 2 \exp\left(-\frac{u^2}{2d(s,t)^2}\right) .$$
 (2.4)

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

$$\sup_{t \in T} \sum_{n \ge m} 2^{n/2} d(t, T_n) \le B .$$
(2.172)

³¹ In practice however, as of today, the Gaussian processes for which continuity is important can be handled through Dudley's bound, while for those which cannot be handled through this bound (such as in Chapter 4) it is boundedness which matters. For this reason, the considerations of the present section are of purely theoretical interest and may be skipped at first reading.

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

$$\forall s, t \in T, \ d(s, t) \le \delta \Rightarrow |X_s - X_t| \le Lu(2^{m/2}\delta + B) \ . \tag{2.173}$$

Proof. We assume T finite for simplicity. For $n \ge m$ and $t \in T$ denote by $\pi_n(t)$ an element of T_n such that $d(t, \pi_n(t)) = d(t, T_n)$. Consider the event Ω_u defined by³²

$$\forall n \ge m+1 , \forall t \in T , |X_{\pi_{n-1}(t)} - X_{\pi_n(t)}| \le Lu 2^{n/2} d(\pi_{n-1}(t), \pi_n(t)) ,$$
(2.174)

and

$$\forall s', t' \in T_m, |X_{s'} - X_{t'}| \le Lu 2^{m/2} d(s', t').$$
 (2.175)

Then as in Section 2.4 we have $\mathsf{P}(\Omega_u) \geq 1 - \exp(-u^2 2^m)$. Now, when Ω_u occurs, using chaining as usual and (2.174) we get

$$\forall t \in T , |X_t - X_{\pi_m(t)}| \le LuB .$$
(2.176)

Moreover, using (2.172) in the inequality, $d(t, \pi_m(t)) = d(t, T_m) \leq B2^{-m/2}$, so that, using (2.175),

$$d(s,t) \le \delta \Rightarrow d(\pi_m(s), \pi_m(t)) \le \delta + 2B2^{-m/2}$$

$$\Rightarrow |X_{\pi_m(s)} - X_{\pi_m(t)}| \le Lu(\delta 2^{m/2} + B) .$$

Combining with (2.175) proves that $|X_s - X_t| \leq Lu(\delta 2^{m/2} + B)$ and completes the proof. \Box

Exercise 2.15.2. Deduce Dudley's bound (1.19) from Lemma 2.15.1.

We now turn to our main result, which exactly describes the modulus of continuity of a Gaussian process in term of certain admissible sequences. It implies in particular the remarkable fact (discovered by X. Fernique) that for Gaussian processes the "local modulus of continuity" (as in (2.177)) is also "global".

Theorem 2.15.3. There exists a constant L^* with the following property. Consider a Gaussian process $(X_t)_{t\in T}$, with canonical associated distance d given by (0.1). Assume that $S = \mathsf{E} \sup_{t\in T} X_t < \infty$. For $k \ge 1$ consider $\delta_k > 0$ and assume that

$$\forall t \in T ; \mathsf{E} \sup_{\{s \in T; d(s,t) \le \delta_k\}} |X_s - X_t| \le 2^{-k} S.$$
 (2.177)

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

 $^{^{32}}$ We are again following the general method outlined at the end of Section 2.4.

2.15 Continuity of Gaussian Processes 79

$$L^* S 2^{-n_k/2-k} \le \delta_k . (2.178)$$

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

$$\forall k \ge 0 \; ; \; \sup_{t \in T} \sum_{n \ge n_k} 2^{n/2} \Delta(A_n(t)) \le LS2^{-k} \; .$$
 (2.179)

Conversely, given integers n_k and an admissible sequence (\mathcal{A}_n) as in (2.179), and defining now $\delta_k^* = S2^{-n_k/2-k}$, with probability $\geq 1 - \exp(-u^2 2^{n_k})$ we have

$$\sup_{\{s,t\in T; d(s,t)\leq \delta_k^*\}} |X_s - X_t| \leq Lu 2^{-k} S.$$
(2.180)

The abstract formulation here might make it hard at first to feel the power of the statement. The numbers δ_k control the (local) modulus of continuity of the process. The numbers n_k control the uniform convergence (over t) of the series $\sum_{n\geq 0} 2^{n/2} \Delta(A_n(t))$. They relate to each other by the relation $\delta_k \sim S2^{-n_k/2-k}$. The second part of the theorem asserts that in turn the numbers n_k control the uniform modulus of continuity (2.180).

Proof. According to the Majorizing Measure Theorem and specifically (2.114), there exists a constant L^* such that for each subset U of T there exists an admissible sequence (\mathcal{A}_n) of partitions of U such that

$$\forall t \in U , \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \le \frac{L^*}{2} \mathsf{E} \sup_{s \in U} X_s .$$
 (2.181)

Assuming (2.177), by induction over k we construct an admissible sequence $(\mathcal{A}_n)_{n \leq n_k}$ such that

$$1 \le p \le k \Rightarrow \sup_{t \in T} \sum_{n_{p-1} < n \le n_p} 2^{n/2} \Delta(A_n(t)) \le L^* S 2^{-p} .$$
 (2.182)

For k = 1 the existence of the sequence $(\mathcal{A}_n)_{n < n_1}$ follows from the Majorizing Measure Theorem through (2.181) as explained, so we turn to the induction step from k to k + 1. Using (2.182) for p = k we deduce that for each $t \in T$, $2^{n_k/2} \Delta(A_{n_k}(t)) \leq L^* S 2^{-k}$, so that $\Delta(A_{n_k}(t)) \leq L^* S 2^{-n_k/2-k} \leq \delta_k$ using (2.178) in the last inequality. Consequently, for any element C of \mathcal{A}_{n_k} we have $\Delta(C) \leq \delta_k$, so that considering any point t of C we have, using (2.177) in the last inequality

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

Using the Majorizing Measure Theorem we construct for each $C \in \mathcal{A}_{n_k}$ an admissible sequence $(\mathcal{A}_{C,n})_{n>0}$ of partitions of C for which

$$\forall t \in C , \sum_{n \ge 0} 2^{n/2} \Delta(A_{C,n}(t)) \le L^* S 2^{-k-1} .$$
(2.183)

For $n_k < n \le n_{k+1}$ we simply define \mathcal{A}_n as the collection of all sets in one of the partitions $A_{C,n-1}$ where $C \in \mathcal{A}_{n_k}$, so that $\operatorname{card} \mathcal{A}_n \le N_{n-1} \operatorname{card} \mathcal{A}_{n_k} \le N_{n-1}^2 \le N_n$. Since for $t \in C$ we have $A_n(t) = A_{C,n-1}(t)$ it follows from (2.183) that for any $C \in \mathcal{A}_{n_k}$ we have

$$\sup_{t \in C} \sum_{n_k < n \le n_{k+1}} 2^{n/2} \Delta(A_n(t)) \le \sup_{t \in C} \sum_{n > n_k} 2^{n/2} \Delta(A_{C,n-1}(t)) \le L^* S 2^{-k}$$

This completes the induction and the construction of the sequence (\mathcal{A}_n) since (2.182) implies (2.179).

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

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

For each k we then use Lemma 2.15.1 for $m = n_k$ and $B = S2^{-k}$.

Key ideas to remember.³³

- The generic chaining efficiently organizes the standard chaining argument for processes whose increments have Gaussian-like tails governed by a distance d as in (2.4).
- The generic chaining applied to such processes motivates the introduction of our main measure $\gamma_2(T, d)$ of the size of a metric space (T, d). This measure involves the existence of suitable sequences of partitions.
- The fundamental problem then becomes how to construct such sequences of partitions in a metric space.
- There is a machine (called a partitioning scheme) to construct such sequences of partitions. The input to the machine is a functional, a function of the subsets of our basic metric space, which in a sense is a measure of their size. The existence of such functionals with specific growth properties is intrinsically linked to the existence of such sequences of partitions.
- The Majorizing Measure Theorem is the statement that for a Gaussian process with index set T and canonical distance d the quantity $\mathsf{E}\sup_{t\in T} X_t$ is exactly of order $\gamma_2(T, d)$. The proof relies on a partitioning scheme, used for the functional $F(A) = \mathsf{E}\sup_{t\in A} X_t$. Sudakov minoration and concentration of measure are the main tools to prove that this functional satisfies the required growth condition.
- Gaussian processes can be seen as subsets of a standard Hilbert space, but the geometric understanding that would relate the size of a set with the size of its convex hull is still lacking.
- The traditional way to organize chaining uses entropy numbers. Even for sets as basic as ellipsoids in Hilbert space, entropy numbers provide only a suboptimal description of their size.

³³ The function of this brief summary is *not* to explain the material again, but is a way for the reader to check that she did understand the main ideas. If any of the points made below is not clear to the reader, she may not be ready to proceed and may want to review the corresponding material.

2.16 Notes and Comments

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

In 1967, R. Dudley obtained the inequality (2.38). (As he pointed out, R. Dudley did not state (2.38) though he performed all the essential steps, and (2.38) totally deserves to be called Dudley's bound.) A few years later, X. Fernique proved that in the "stationary case" Dudley's inequality can be reversed [43], i.e. he proved in that case the lower bound of Theorem 2.10.1. This historically important result was central to the work of Marcus and Pisier [85], [86] who built on it to solve all the classical problems on random Fourier series. Some of their results will be presented in Chapter 7. Interestingly, now that the right approach has been found, the proof of Fernique's result is not really easier than that of Theorem 2.10.1.

Another major contribution of Fernique (building on earlier ideas of C. Preston) was an improvement of Dudley's bound based on a new tool called majorizing measures (which we will study in Subsection 3.1.3). Fernique conjectured that his bound was essentially optimal. Gilles Pisier suggested in 1983 that I should work on this conjecture. In my first attempt I proved fast that Fernique's conjecture held in the case where the metric space (T, d)is ultrametric. I learned that Fernique had already done this, so I was discouraged for a while. In the second attempt, I tried to decide whether a majorizing measure existed on ellipsoids. I had the hope that some simple density with respect to the volume measure would work. It was difficult to form any intuition, and I struggled in the dark for months. At some point I tried a combination of suitable point masses, and easily found a direct construction of the majorizing measure on ellipsoids. This made it believable that Fernique's conjecture was true, but I still tried to disprove it. Then I realized that I did not understand why a direct approach using a partitioning scheme should fail, while this understanding should be useful to construct a counterexample. Once I tried this direct approach, it was a matter of three days to prove Fernique's conjecture. Gilles Pisier made two comments about this discovery. The first one was "you are lucky", by which he meant that I was lucky that Fernique's conjecture was true, since a counter example would have been of limited interest. I am grateful to this day for his second comment: "I wish I had proved this myself, but I am very glad you did it."

Fernique's concept of majorizing measures is difficult to grasp, and was dismissed by the main body of probabilists as a mere curiosity. (I myself found it very difficult to understand.) This could be the main reason why Fernique's path-breaking work did not receive the recognition it should have. I have tried to repair this and to express my personal admiration by dedicating this book to his memory and by paying homage to his work at numerous places in this book.

In 2000, while discussing one of the open problems of this book with Keith Ball (be he blessed for his interest in it!) I discovered that one could replace majorizing measures by the totally natural variation on the usual chaining arguments that was presented here. That this was not discovered much earlier is a striking illustration of the inefficiency of my brain. For two decades it looked like majorizing measures would not be of any use anymore, but they now play again a major role again for reasons to be explained in Chapter 5.

In [150] the author presented a particularly simple proof of Theorem 2.10.1 (expressed in terms of majorizing measures since the generic chaining had not been invented yet). It is based a on partition scheme related to the one we use here. The precise relationship is discussed on page 72 of [171].

It is on purpose that I did not mention Slepian's lemma, which is the statement that (2.127) holds for L = 1. This lemma is very specific to Gaussian processes, and focusing on it seems a good way to guarantee that one will never move beyond these. One notable progress I made was to discover (ages ago) the scheme of proof of Proposition 2.10.8 that dispenses with Slepian's lemma, and that we shall use in many situations. Comparison results such as Slepian's lemma are not at the root of results such as the Majorizing Measure Theorem, but rather are (at least qualitatively) a consequence of them as in Corollary 2.10.12. This being said, Slepian's lemma is historically very important as it crystallizes the link between $\mathsf{E} \sup_{t \in T} X_t$ and the structure of the metric space (T, d).

3. Trees and Other Measures of Size

In this chapter we systematically investigate different ways to measure the size of a metric space. One of them, Fernique's functional of Section 3.3 will play a major role in the sequel, as it is the form which lends itself to vast generalizations. The concept of a tree presented in Section 3.1 is historically important: the author discovered many of the results he presents while thinking in terms of trees. We know now how to present these results and their proofs without ever mentioning trees, arguably in a more elegant fashion, so that trees are not used explicitly elsewhere in this book. However, it might be too early to dismiss this concept, at least as an instrument of discovery.

3.1 Trees

We shall describe different ways to measure the size of a metric space and show that they are all equivalent to the functional $\gamma_2(T, d)$.¹

In a nutshell, a tree is a certain structure that requires a "lot of space" to be constructed, so that a metric space needs to be large in order to contain large trees. At the simplest level, it already takes some space to construct in a set A sets B_1, \ldots, B_n which are appropriately separated from each other. This is even more so if the sets B_1, \ldots, B_n are themselves large (for example because they contain many sets far from each other). Trees are a proper formulation of the iteration of this idea. The basic use of trees is to measure the size of a metric space by the size of the largest tree (of a certain type) which it contains. Different types of trees yield different measures of size.

A tree \mathcal{T} of a metric space (T, d) is a *finite* collection of non-empty subsets of T with the following two properties.

Given
$$A, B$$
 in \mathcal{T} , if $A \cap B \neq \emptyset$, then either $A \subset B$ or else $B \subset A$. (3.1)

$$\mathcal{T}$$
 has a largest element. (3.2)

The important condition here is (3.1), and (3.2) is just for convenience. If $A, B \in \mathcal{T}$ and $B \subset A, B \neq A$, we say that B is a *child* of A if

¹ It is possible to consider more general notions corresponding to other functionals considered in the book, but for simplicity we consider only the case of γ_2 .

84 3. Trees and Other Measures of Size

$$C \in \mathcal{T}, B \subset C \subset A \Rightarrow C = B \text{ or } C = A.$$
 (3.3)

We denote by c(A) the number of children of A. Since our trees are finite, some of their sets will have no children. It is convenient to "shrink these sets to a single point", so we will consider only trees with the following property:

If
$$A \in \mathcal{T}$$
 and $c(A) = 0$, then A contains exactly one point. (3.4)

A fundamental property of trees is as follows. Consider trees $\mathcal{T}_1, \ldots, \mathcal{T}_m$ and for $1 \leq \ell \leq m$ let A_ℓ be the largest element of \mathcal{T}_ℓ . Assume that the sets A_ℓ are disjoint, and consider a set A with $\bigcup_{\ell \leq m} A_\ell \subset A \subset T$. Then the collection of subsets of T consisting of A and of $\bigcup_{\ell \leq m} \mathcal{T}_\ell$ is a tree. The proof is straightforward. This fact allows one to construct iteratively more and more complicated (and larger) trees.

An important structure in a tree is a branch. A sequence A_0, A_1, \ldots, A_k is a branch if $A_{\ell+1}$ is a child of A_ℓ , and if moreover A_0 is the largest element of \mathcal{T} while A_k has no child. Then by (3.4) the set A_k is reduced to a single point t, and A_0, \ldots, A_k are exactly those elements of \mathcal{T} which contain t. So in order to describe the branches of \mathcal{T} it is convenient to introduce the set

$$S_{\mathcal{T}} = \{t \in T ; \{t\} \in \mathcal{T}\}, \qquad (3.5)$$

which we call the *support* of \mathcal{T} . If a set A in a tree has no child, one may call it a *leaf*. Thus a leaf of a tree is reduced to one single point, and the support of a tree is the union of its leaves. By considering all the collections $\{A \in \mathcal{T}; t \in A\}$ as t varies in $S_{\mathcal{T}}$ we obtain all the branches of \mathcal{T} .

3.1.1 Separated Trees

We now quantify our desired property that the children of a given set should be far from each other in an appropriate sense. A *separated* tree is a tree \mathcal{T} such that to each A in \mathcal{T} with $c(A) \geq 1$ is associated an integer $s(A) \in \mathbb{Z}$ with the following properties. First,

If B_1 and B_2 are distinct children of A, then $d(B_1, B_2) \ge 4^{-s(A)}$. (3.6)

Here $d(B_1, B_2) = \inf\{d(x_1, x_2); x_1 \in B_1, x_2 \in B_2\}$. We observe that in (3.6) we make no restriction on the diameter of the children of A. (Such restrictions will however occur in the other notion of tree that we consider later.) Second, to rule out pathologies, we will also make the following purely technical assumption:

If B is a child of A, then
$$s(B) > s(A)$$
. (3.7)

To measure the size of a separated tree T we introduce its *depth*, i.e.

$$\rho(\mathcal{T}) := \inf_{t \in S_{\mathcal{T}}} \sum_{t \in A \in \mathcal{T}} 4^{-s(A)} \sqrt{\log c(A)} .$$
(3.8)

Here and below we make the convention that the summation does not include the term $A = \{t\}$ (for which c(A) = 0). The quantity (3.8) takes into account both the separation between the children of A (through the term $4^{-s(A)}$) and their number (through the term $\sqrt{\log c(A)}$). This will be a common feature of all our notions of sizes of trees.

We observe that in (3.8) we have the *infimum* over $t \in S_{\mathcal{T}}$. In words:

A tree is large if it is large along *every* branch.

We can then measure the size of T by

$$\sup\{\rho(\mathcal{T}) ; \mathcal{T} \text{ separated tree} \subset T\}.$$
(3.9)



Fig. 3.1. A separated tree. The children of A_0 are A_1 and B. The children of A_1 are A_2 and C. A_0, A_1, A_2, A_3 form a branch, of which A_3 is a leaf.

3.1.2 Organized trees

The notion of separated tree we just considered is but one of many possible notions of trees and it does not seem fundamental. Rather, the quantity (3.9) is used as a convenient intermediate technical step to prove the equivalence of several more important quantities. Let us now consider another notion

86 3. Trees and Other Measures of Size

of trees, which is more restrictive (and apparently much more important). An *organized* tree is a tree \mathcal{T} such that to each $A \in \mathcal{T}$ with $c(A) \geq 1$ are associated an integer $j = j(A) \in \mathbb{Z}$, and points $t_1, \ldots, t_{c(A)}$ with the properties that

$$1 \le \ell < \ell' \le c(A) \Rightarrow 4^{-j-1} \le d(t_{\ell}, t_{\ell'}) \le 4^{-j+2}$$
(3.10)

and that each ball $B(t_{\ell}, 4^{-j-2})$ contains exactly one child of A. In some sense $4^{-j(A)}$ tells you at which scale the children of A live. Please note that it may happen that $4^{-j(A)}$ is much smaller than $\Delta(A)$.

If B_1 and B_2 are distinct children of A in an organized tree, then

$$d(B_1, B_2) \ge 4^{-j(A)-2}, \qquad (3.11)$$

so that an organized tree is also a separated tree, with s(A) = j(A) + 2, but the notion of organized tree is more restrictive. (For example we have no control over the diameter of the children of A in a separated tree.)



Fig. 3.2. An organized tree. Here $j(A_2) > j(A_1)$.

We define the depth $\tau(\mathcal{T})$ of an organized tree by

$$\tau(\mathcal{T}) := \inf_{t \in S_{\mathcal{T}}} \sum_{t \in A \in \mathcal{T}} 4^{-j(A)} \sqrt{\log c(A)} .$$
(3.12)

Another way to measure the size of T is then

$$\sup\{\tau(\mathcal{T}) ; \mathcal{T} \text{ organized tree} \subset T\}.$$
(3.13)

If we simply view an organized tree \mathcal{T} as a separated tree using (3.11), then $\rho(\mathcal{T}) = \tau(\mathcal{T})/16$ (where $\rho(\mathcal{T})$ is the depth of \mathcal{T} as a separated tree). Thus we have shown the following.

Proposition 3.1.1. We have

 $\sup\{\tau(\mathcal{T}); \mathcal{T} \text{ organized tree}\} \le 16 \sup\{\rho(\mathcal{T}); \mathcal{T} \text{ separated tree}\}.$ (3.14)

The next result provides the fundamental connection between trees and the functional γ_2 .

Proposition 3.1.2. We have

$$\gamma_2(T,d) \le L \sup\{\tau(\mathcal{T}) ; \mathcal{T} \text{ organized tree}\}.$$
 (3.15)

Proof. We consider the functional

 $F(A) = \sup\{\tau(\mathcal{T}) ; \mathcal{T} \subset A, \mathcal{T} \text{ organized tree}\},\$

where we write $\mathcal{T} \subset A$ as a shorthand for " $\forall B \in \mathcal{T}, B \subset A$ ".

Next we prove that this functional satisfies the growth condition (2.77) for r = 16 whenever a is of the type 16^{-j} , for $c^* = 1/L$. For this consider $n \ge 1$ and $m = N_n$. Consider $j \in \mathbb{Z}$ and $t_1, \ldots, t_m \in T$ with

$$1 \le \ell < \ell' \le m \Rightarrow 16^{-j} \le d(t_{\ell}, t_{\ell'}) \le 2 \cdot 16^{-j+1} .$$
(3.16)

Consider sets $H_{\ell} \subset B(t_{\ell}, 2 \cdot 16^{-j-1})$ and $\alpha < \min_{\ell \le m} F(H_{\ell})$. Consider, for $\ell \le m$ an organized tree $\mathcal{T}_{\ell} \subset H_{\ell}$ with $\tau(\mathcal{T}_{\ell}) > \alpha$ and denote by A_{ℓ} its largest element. Next we claim that the tree \mathcal{T} consisting of $C = \bigcup_{\ell \le m} H_{\ell}$ (its largest element) and the union of the trees \mathcal{T}_{ℓ} , $\ell \le m$, is organized, with j(C) = 2j - 1, and A_1, \ldots, A_m as children of C (so that c(C) = m). To see this we observe that since $4^{-j(C)-1} = 16^{-j}$ we have $4^{-j(C)-1} \le d(t_{\ell}, t_{\ell'}) \le 2 \cdot 16^{-j+1} \le 4^{-j(C)+2}$, so that (3.10) holds for C. Furthermore $A_{\ell} \subset H_{\ell} \subset B(t_{\ell}, 2 \cdot 16^{-j-1}) \subset B(t_{\ell}, 4^{-j(C)-2})$, so that this ball contains exactly one child of C. The other conditions follow from the fact that the trees \mathcal{T}_{ℓ} are themselves organized. Moreover $S_{\mathcal{T}} = \bigcup_{\ell \le m} S_{\mathcal{T}_{\ell}}$.

Consider $t \in S_{\mathcal{T}}$, and let ℓ with $t \in S_{\mathcal{T}_{\ell}}$. Then $\overline{t} \in C \in \mathcal{T}$, and also $t \in A \in \mathcal{T}$ whenever $t \in A \in \mathcal{T}_{\ell}$. Thus, using also in the second line that j(C) = 2j - 1 and that c(C) = m, we obtain

$$\sum_{t \in A \in \mathcal{T}} 4^{-j(A)} \sqrt{\log c(A)} \ge 4^{-j(C)} \sqrt{\log c(C)} + \sum_{t \in A \in \mathcal{T}_{\ell}} 4^{-j(A)} \sqrt{\log c(A)}$$
$$\ge 4 \cdot 16^{-j} \sqrt{\log m} + \tau(\mathcal{T}_{\ell}) \ge \frac{1}{L} 16^{-j} 2^{n/2} + \alpha .$$

Since α is arbitrary, we have proved that

$$F\left(\bigcup_{\ell \le m} H_\ell\right) \ge \tau(\mathcal{T}) \ge \frac{1}{L} 16^{-j} 2^{n/2} + \min_{\ell \le m} F(H_\ell) .$$

This completes the proof of the growth condition (2.77).

88 3. Trees and Other Measures of Size

If one examines the proof of Theorem 2.9.1, one observes that it requires only the growth condition (2.77) to hold true when a is of the type r^{-j} , and we have just proved that this is the case (for r = 16), so that from (2.81) we have proved that $\gamma_2(T, d) \leq L(F(T) + \Delta(T))$. It remains only to prove that $\Delta(T) \leq LF(T)$. For this we simply note that if $s, t \in T$, and j_0 is the largest integer with $4^{-j_0} \geq d(s, t)$, then the tree \mathcal{T} consisting of $T, \{t\}, \{s\}$, is organized with $j(T) = j_0$ and c(T) = 2, so that $F(T) \geq \tau(\mathcal{T}) \geq 4^{-j_0} \sqrt{\log 2}$ and $4^{-j_0} \leq LF(T)$.

3.1.3 Majorizing Measures

For a probability measure μ on a metric space (T, d), with countable support,² we define for each $t \in T$ the quantity

$$I_{\mu}(t) := \int_{0}^{\infty} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}} d\epsilon = \int_{0}^{\Delta(T)} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}} d\epsilon .$$
(3.17)

The second equality follows from the fact that $\mu(B(t, \epsilon)) = 1$ when $B(t, \epsilon) = T$, so that then the integrand is 0. It is important to master the mechanism at play in the following elementary exercise.

Exercise 3.1.3. Consider a number $\Delta > 0$ and non-increasing function f: $[0, \Delta] \to \mathbb{R}^+$. Define $\epsilon_0 = \Delta$ and for $n \ge 1$ define $\epsilon_n = \inf\{\epsilon > 0; f(\epsilon) \le 2^n\}$. Prove that

$$\frac{1}{2}\sum_{n\geq 1} 2^n \epsilon_n \leq \int_0^\Delta f(\epsilon) \mathrm{d}\epsilon \leq 2\sum_{n\geq 0} 2^n \epsilon_n$$

Proposition 3.1.4. Given a metric space (T, d) we can find on T a probability measure μ , supported by a countable subset of T, and such that

$$\sup_{t \in T} I_{\mu}(t) = \sup_{t \in T} \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}} d\epsilon \le L\gamma_2(T,d) .$$
(3.18)

Any probability measure³ μ on (T, d) is called a *majorizing measure*. The reason for this somewhat unsatisfactory name is that Xavier Fernique proved that for a Gaussian process and a probability measure μ on T one has

$$\mathsf{E}\sup_{t\in T} X_t \le L \sup_{t\in T} I_{\mu}(t) , \qquad (3.19)$$

 $^{^2}$ We assume μ with countable support because we do not need a more general setting. The advantage of this hypothesis is that there are no measurability problems.

 $^{^3}$ To avoid technicalities one may assume that μ has countable support.

so that μ can be used to "majorize" the process $(X_t)_{t\in T}$.⁴ This was a major advance over Dudley's bound. The (in)famous theory of majorizing measures used the quantity

$$\inf_{\mu} \sup_{t \in T} I_{\mu}(t) \tag{3.20}$$

as a measure of the size of the metric space (T, d), where the infimum is over all choices of the probability measure μ . This method is technically quite challenging. We are going to prove that the quantity (3.20) is equivalent to $\gamma_2(T, d)$. A related idea which is still very useful is explained in Section 3.3.

Proof. Consider an admissible sequence (\mathcal{A}_n) with

$$\forall t \in T, \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \le 2\gamma_2(T, d) .$$

Let us now pick a point $t_{n,A}$ in each set $A \in \mathcal{A}_n$, for each $n \geq 0$. Since card $\mathcal{A}_n \leq N_n$, for each *n* there are at most N_n points of the type $t_{n,A}$. Attributing a mass $1/(2^n N_n)$ to each of them we obtain a total mass ≤ 1 . Thus there is a probability measure μ on *T*, supported by a countable set, and satisfying $\mu(\{t_{n,A}\}) \geq 1/(2^n N_n)$ for each $n \geq 0$ and each $A \in \mathcal{A}_n$. Then,

$$\forall n \ge 1, \, \forall A \in \mathcal{A}_n, \, \mu(A) \ge \mu(\{t_{n,A}\}) \ge \frac{1}{2^n N_n} \ge \frac{1}{N_n^2}$$

so that given $t \in T$ and $n \ge 1$,

$$\epsilon > \Delta(A_n(t)) \Rightarrow \mu(B(t,\epsilon)) \ge \frac{1}{N_n^2}$$
$$\Rightarrow \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}} \le 2^{n/2+1} . \tag{3.21}$$

Now, since μ is a probability, $\mu(B(t,\epsilon)) = 1$ for $\epsilon > \Delta(T)$, and then $\log(1/\mu(B(t,\epsilon))) = 0$. Thus

$$I_{\mu}(t) = \int_{0}^{\infty} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}} d\epsilon = \sum_{n \ge 0} \int_{\Delta(A_{n+1}(t))}^{\Delta(A_{n}(t))} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}} d\epsilon$$
$$\leq \sum_{n \ge 1} 2^{(n+1)/2+1} \Delta(A_{n}(t)) \le L\gamma_{2}(T,d)$$

using (3.21).

 4 One typically uses the name only when such the right-hand side of (3.19) is usefully small.

90 3. Trees and Other Measures of Size

Proposition 3.1.5. If μ is a probability measure on T (supported by a countable set) and \mathcal{T} is a separated tree on T, then

$$\rho(\mathcal{T}) \le L \sup_{t \in T} I_{\mu}(t) \; .$$

Combining with (3.14), (3.15) and (3.18), this completes the proof that the four "measures of the size of T" considered in this section, namely (3.9), (3.13), (3.20) and $\gamma_2(T, d)$ are indeed equivalent.

Proof. The basic observation is as follows. The sets

$$B(C, 4^{-s(A)-1}) = \{x \in T ; d(x, C) < 4^{-s(A)-1}\}$$

are disjoint as C varies over the children of A (as follows from (3.6)), so that one of them has measure $\leq c(A)^{-1}$.

We then proceed in the following manner, constructing recursively an appropriate branch of the tree. This is a typical and fundamental way to proceed when working with trees. We start with the largest element A_0 of \mathcal{T} . We then select a child A_1 of A_0 with $\mu(B(A_1, 4^{-s(A_0)-1})) \leq 1/c(A_0)$, and a child A_2 of A_1 with $\mu(B(A_2, 4^{-s(A_1)-1})) \leq 1/c(A_1)$, etc., and continue this construction as long as we can. It ends only when we reach a set of \mathcal{T} that has no child, and hence by (3.4) is reduced to a single point t which we now fix. For any set A with $t \in A \in \mathcal{T}$, by construction we have

$$\mu(B(t, 4^{-s(A)-1})) \le \frac{1}{c(A)}$$

so that

$$4^{-s(A)-2}\sqrt{\log c(A)} \le \int_{4^{-s(A)-2}}^{4^{-s(A)-1}} \sqrt{\frac{1}{\log \mu(B(t,\epsilon))}} d\epsilon , \qquad (3.22)$$

because the integrand is $\geq \sqrt{\log c(A)}$ and the length of the interval of integration is larger than $4^{-s(A)-2}$. By (3.7) the intervals $]4^{-s(A)-2}$, $4^{-s(A)-1}[$ are disjoint for different sets A with $t \in A \in \mathcal{T}$, so summation of the inequalities (3.22) yields

$$\frac{1}{16}\rho(\mathcal{T}) \le \sum_{t \in A \in \mathcal{T}} 4^{-s(A)-2} \sqrt{\log c(A)} \le \int_0^\infty \sqrt{\frac{1}{\log \mu(B(t,\epsilon))}} \mathrm{d}\epsilon = I_\mu(t) \ . \quad \Box$$

In the rest of this chapter, we will implicitly use the previous method of "selecting recursively the branch of the tree we follow" to prove lower bounds without mentioning trees.

We end this section by an exercise completing the proof of (2.143).

Exercise 3.1.6. Consider metric spaces $(T_k, d_k)_{k \leq N}$ and probability measures μ_k on T_k . Consider the product probability μ on $T = \prod_{k \leq N} T_k$ and the distance (2.142).

(a) Prove that for $t = (t_k)_{k \le N}$ we have

$$I_{\mu}(t) \leq L \sum_{k \leq N} I_{\mu_k}(t_k) \; .$$

Hint: use (2.147).(b) Complete the proof of (2.143).

3.2 Rolling Up our Sleeves: Trees in Ellipsoids

It is one thing to have proved abstract results, but quite another thing to visualize the combinatorics in concrete situations. Consider an ellipsoid \mathcal{E} as in (2.154), so that according to (2.155), $S := \sqrt{\sum_{i \leq N} a_i^2}$ measures its size. Assuming $S < \infty$ the goal of the present section is to construct explicitly an organized tree \mathcal{T} whose depth $\tau(\mathcal{T})$ witnesses the size of the ellipsoid, i.e. $\tau(\mathcal{T}) \geq S/L$. This elaborate exercise will have us confront a number of technical difficulties.

We first reduce to the case where each a_i is of the type 2^{-k} for some $k \in \mathbb{Z}$. Let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. For $k \in \mathbb{Z}$ let us set $I_k = \{i \in \mathbb{N}^*; 2^{-k} \le a_i \le 2^{-k-1}\}$, so that the sets I_k cover \mathbb{N}^* and $\sum_k 2^{-2k} \operatorname{card} I_k \ge \sum_{i\ge 1} a_i^2/4 \ge S^2/4$, while at the same time

$$\mathcal{E} \supset \mathcal{E}' = \left\{ t \in \ell^2 \; ; \; \sum_k 2^{2k} \sum_{i \in I_k} t_i^2 \le 1 \right\} \,. \tag{3.23}$$

Thus we have reduced the problem to considering only ellipsoids of the type \mathcal{E}' . Here comes a somewhat unexpected argument. We are going to replace \mathcal{E}' by a set of the type

$$\mathcal{P} = \left\{ t \in \ell^2 ; \ \forall k \ , \ 2^{2k} \sum_{i \in I_k} t_i^2 \leq \alpha_k \right\},$$

where $\sum_k \alpha_k = 1$ (a condition which ensures that $\mathcal{P} \subset \mathcal{E}'$). At first sight the set \mathcal{P} looks much smaller than \mathcal{E}' , but this is not the case. First, how should we chose α_k to ensure that \mathcal{P} is as large as possible? Considering independent Gaussian r.v.s (g_i) we have $\sup_{t \in \mathcal{P}} \sum_{i \ge 1} t_i g_i = \sum_k \sqrt{\alpha_k} 2^{-k} \sqrt{\sum_{i \in I_k} g_i^2}$, so that since $\mathsf{E}_k \sqrt{\sum_{i \in I_k} g_i^2}$ is about $\sqrt{\operatorname{card} I_k}$ by (2.155), we obtain

$$\mathsf{E}\sup_{t\in\mathcal{P}}\sum_{i\geq 1} t_i g_i \text{ is about } \sum_k \sqrt{\alpha_k} 2^{-k} \sqrt{\operatorname{card} I_k} .$$
 (3.24)

So to maximize this quantity (and, in a sense, the size of \mathcal{P}), it is a good idea to chose $\alpha_k = 2^{-2k} \operatorname{card} I_k / S'^2$ where $S'^2 = \sum_{\ell} 2^{-2\ell} \operatorname{card} I_{\ell}$. Then \mathcal{P} takes the form

92 3. Trees and Other Measures of Size

$$\mathcal{P} = \left\{ t \in \ell^2 ; \forall k , \sum_{i \in I_k} t_i^2 \le \frac{2^{-4k}}{S'^2} \operatorname{card} I_k \right\}.$$

This set is very simple: geometrically, \mathcal{P} is a product of spheres of dimension card I_k and radius r_k , where r_k is defined by $r_k^2 = 2^{-4k} \operatorname{card} I_k / S'^2$. It will be very useful to reduce to the case where the radii of these spheres are quite different from each other. This is the purpose of the next lemma.

Lemma 3.2.1. There is a subset J of \mathbb{Z} with the following two properties:

$$\sum_{k \in J} 2^{-2k} \operatorname{card} I_k \ge S^{\prime 2} / L .$$
 (3.25)

$$k, n \in J, \ k < n \Rightarrow r_n \le 2^{-6} r_k$$
 (3.26)

Proof. Since $S < \infty$ the sequence $a_k = 2^{-2k} \operatorname{card} I_k$ is bounded. We apply Lemma 2.9.5 (or more precisely the version of this lemma where the index set is \mathbb{Z} rather than \mathbb{N}) with $\alpha = 2$ to find a finite subset $I \subset \mathbb{Z}$ such that

$$k, n \in I, k \neq n \Rightarrow a_n < a_k 2^{|n-k|} . \tag{3.27}$$

and

$$\sum_{k\in I} a_k \ge S^{\prime 2}/L \ . \tag{3.28}$$

Consider $k, n \in I$ with k < n. Then $a_n < 2^{n-k}a_k$, and recalling the value of a_n this means that card $I_n \leq 2^{3(n-k)}$ card I_k . Recalling the value of r_k this means that $r_n^2 \leq r_k^2/2$. Let us now enumerate I as a finite sequence $(n(\ell))_{\ell \leq N}$ in increasing order, so that $r_{n(\ell+1)} \leq r_{n(\ell)}/\sqrt{2}$ and $r_{n(\ell+12)} \leq 2^{-6}r_{n(\ell)}$. For $0 \leq p \leq 11$ consider the set $J_p = \{n(p+12q); q \geq 0, p+12q \leq N\}$, so that $I = \bigcup_{0 \leq p \leq 11} J_p$. Consequently $\sum_{k \in I} a_k = \sum_{0 \leq p \leq 11} \sum_{k \in J_p} a_k$. Thus using (3.28) there exists some $p, 0 \leq p \leq 11$ such that $\sum_{k \in J_p} a_k \geq S'^2/L$ and the set J_p satisfies the desired requirements.

Consider a set J as constructed in Lemma 3.2.1. We then replace \mathcal{P} by the subset \mathcal{P}' consisting of the points $t \in \mathcal{P}$ such that $t_i = 0$ when $i \in I_k$ and $k \notin J$. We note that $\sum_{k \in J} r_k \sqrt{\operatorname{card} I_k} = \sum_{k \in J} 2^{-2k} \operatorname{card} I_k / S' \geq S' / L$, where the last inequality follows from (3.25).

We have now finished our preliminary reductions. To construct inside any ellipsoid \mathcal{E} an organized tree \mathcal{T} such that its depth $\tau(\mathcal{T})$ witnesses the size of \mathcal{E} it suffices to perform the same task for a set of the type

$$\mathcal{P}' = \left\{ t \in \ell^2(I^*) ; \forall k \le N , \sum_{i \in I_k} t_i^2 \le r_k^2 \right\},\$$

where N is a given integer, where $(I_k)_{k\leq N}$ are disjoint subsets of \mathbb{N}^* of union I^* , and where $r_{k+1} \leq 2^{-6}r_k$. Just as in (3.24), the size of \mathcal{P}' is about $\sum_{k\leq N} r_k \sqrt{\operatorname{card} I_k}$.

For $k \leq N$ let us consider the sphere

$$\mathbb{S}_k = \left\{ t \in \ell^2(I^*) ; \sum_{i \in I_k} t_i^2 \le r_k^2 , i \notin I_k \Rightarrow t_i = 0 \right\}.$$

It follows from the volume argument (2.45) (used for A = B and $\epsilon = 1/2$) that there is a subset U_k of \mathbb{S}_k with

$$\operatorname{card} U_k \ge 2^{\operatorname{card} I_k} , \qquad (3.29)$$

such that any two distinct points of U_k are at distance $\geq r_k/2$. Given $1 \leq m \leq N$ and for $k \leq m$ given $y_k \in U_k$, $y_k = (y_{k,i})_{i \in I^*}$ consider the set $A = A(y_1, \ldots, y_m) \subset \mathcal{P}'$ defined as

$$A(y_1, \dots, y_m) = \left\{ t \in \ell^2(I^*) ; \ \forall k \le m \ , \ \forall i \in I_k \ , \ t_i = y_{k,i} \right\} .$$

We will show that these sets, together with the set \mathcal{P}' form an organized tree \mathcal{T} (as defined in the previous section). When m < N we have c(A) =card U_{m+1} : the children of A are the sets $A(y_1, \ldots, y_m, y)$ where $y \in U_{m+1}$. When m = N we have c(A) = 0 and $A(y_1, \ldots, y_N)$ consists of single point. We now check the condition (3.10). Consider m < N. Define j(A) as the smallest integer j such that $4^{-j-1} \leq r_{m+1}/2$, so that $r_{m+1} \leq 2 \cdot 4^{-j}$. For $y \in U_{m+1}$ consider the unique point $t(y) \in A(y_1, \ldots, y_m, y)$ such that $t(y)_i = 0$ if $i \in I_k$ for k > m + 1. Then for $y, y' \in U_{m+1}, y \neq y'$ the distance d(t(y), t(y'))is the same as the distance between y and y', which are different points of $U_{m+1} \subset \mathbb{S}_{m+1}$. Thus

$$4^{-j(A)-1} \le r_{m+1}/2 \le d(t(y), t(y')) \le 2r_{m+1} \le 4^{-j(A)+1}$$

Furthermore, recalling that $r_{k+1} \leq 2^{-6}r_k$ (so that in particular $\sum_{k\geq m} r_k \leq 2r_m$) if $x \in A(y_1, \ldots, y_m, y)$ then

$$||x - t(y)|| \le \sum_{k \ge m+2} r_k \le 2r_{m+2} \le 2 \cdot 2^{-6} r_{m+1} \le 2^{-6} 4^{-j(A)+1} = 4^{-j(A)-2}$$

so that $A(y_1, \ldots, y_m, y) \subset B(t(y), 4^{-j(A)-2})$ as required. Let us now study $\tau(\mathcal{T})$. A branch in \mathcal{T} is defined by a point $t \in S_{\mathcal{T}}$, which is the unique point of a set of the type $A(y_1, \ldots, y_N)$. Let us set $A_0 = \mathcal{P}'$ and $A_m := A(y_1, \ldots, y_m)$ for $1 \leq m < N$. Then $t \in A_m$ for $0 \leq m < N$. Also $c(A_m) = \operatorname{card} U_{m+1}$ and $4^{-j(A_m)} \geq r_{m+1}/2$. Thus, using (3.29) in the last equality,

$$\sum_{t \in A \in \mathcal{T}} 4^{-j(A)} \sqrt{\log c(A)} \ge \sum_{0 < m < N} 4^{-j(A_m)} \sqrt{\log c(A_m)}$$
$$\ge \sum_{0 < m < N} r_{m+1} \sqrt{\log \operatorname{card} U_{m+1}} / L \ge \sum_{1 \le k \le N} r_k \sqrt{\operatorname{card} I_k} / L ,$$

and, as we have seen, this last quantity is the size of \mathcal{P}' .

3.3 Fernique's Functional

3.3.1 Fernique's Functional

In Section 3.1 we presented four equivalent methods to measure the size of a metric space. (Besides our usual $\gamma_2(T, d)$, these were the maximum depth of a separated or organized tree contained in T and majorizing measures.) Recalling (3.17), it will turn out that a fifth measure of size, the quantity

$$\operatorname{Fer}(T,d) := \sup_{\mu} \int_{T} I_{\mu}(t) \mathrm{d}\mu(t)$$
(3.30)

will play an important role. Here the supremum is over all probability measures on T which are supported by a countable set.

Why is the functional (3.30) important? This is far from obvious. In the context of Gaussian processes, this functional has no special importance, and the related notion of majorizing measures is not particularly useful. We will first understand the usefulness of Fernique's functional in Chapter 5 while studying a class of processes which are conditionally Gaussian. Furthermore in following chapters we will be able to use similar ideas in far more general situations, while the proper generalization of the other functionals has not been found yet. In this section we will prove that Fernique's functional is equivalent to the functional $\gamma_2(T, d)$,

$$\frac{1}{L}\gamma_2(T,d) \le \operatorname{Fer}(T,d) \le L\gamma_2(T,d) .$$
(3.31)

We will not give the simplest possible proof of this fact. Rather we will prepare for future work by giving arguments which contain in germ the ideas which will prove fruitful. The ideas of this section will not be critically used before Chapter 11.

A further understanding of Fernique's functional will be reached is Section 3.5 where we will basically show the remarkable fact that the supremum in the right-hand side of (3.30) is obtained when μ is the "law of the supremum", i.e. the law of a r.v. such that $X_{\tau} = \sup_{t \in T} X_t$, see Theorem 3.5.1 below.

The right-hand side inequality in (3.31) is the easiest, and is a consequence of the following.

Proposition 3.3.1. Consider a probability measure μ on a metric space (T, d). Then

$$\int_{T} I_{\mu}(t) \mathrm{d}\mu(t) \le L\gamma_2(T, d) .$$
(3.32)

Proof. For each $t \in T$ we define $\epsilon_0(t) = \Delta(T)$ and for $n \ge 1$ we define

$$\epsilon_n(t) = \inf\{\epsilon > 0 \; ; \; \mu(B(t,\epsilon)) \ge N_{n+1}^{-1}\} \; . \tag{3.33}$$

Thus $\sqrt{\log(1/\mu(B(t,\epsilon)))} \le L2^{n/2}$ for $\epsilon \ge \epsilon_n(t)$ and then

$$I_{\mu}(t) = \int_{0}^{\Delta(T)} \sqrt{\frac{1}{\log(\mu(B(t,\epsilon)))}} d\epsilon = \sum_{n \ge 0} \int_{\epsilon_{n+1}(t)}^{\epsilon_{n}(t)} \sqrt{\frac{1}{\log(\mu(B(t,\epsilon)))}} d\epsilon$$
$$\leq L \sum_{n \ge 0} 2^{n/2} \epsilon_{n}(t) . \quad (3.34)$$

Consider an admissible sequence (\mathcal{A}_n) of partitions with

$$\forall t \in T , \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \le 2\gamma_2(T, d) .$$
 (3.35)

Let us fix $n \ge 0$ and set

$$T_n = \left\{ t \in T \; ; \; \mu(A_n(t)) \ge N_{n+1}^{-1} \right\}; T'_n = T \setminus T_n = \left\{ t \in T \; ; \; \mu(A_n(t)) < N_{n+1}^{-1} \right\}.$$

Thus T'_n is the union of the sets $A \in \mathcal{A}_n$ which are of measure $\leq N_{n+1}^{-1}$. Since card $\mathcal{A}_n \leq N_n$ we have $\mu(T'_n) \leq N_n N_{n+1}^{-1} = N_n^{-1}$. Also, by definition of $\epsilon_n(t)$ we have $\epsilon_n(t) \leq \Delta(A_n(t))$ if $t \in T_n$ and $\epsilon_n(t) \leq \Delta(T)$ if $t \in T'_n$. Consequently,

$$\int_{T} 2^{n/2} \epsilon_{n}(t) d\mu(t) = \int_{T_{n}} 2^{n/2} \epsilon_{n}(t) d\mu(t) + \int_{T_{n}'} 2^{n/2} \epsilon_{n}(t) d\mu(t)$$
$$\leq \int_{T_{n}} 2^{n/2} \Delta(A_{n}(t)) d\mu(t) + L\Delta(T) 2^{n/2} N_{n}^{-1} .$$

Combining with (3.34) we obtain

$$\int_{T} I_{\mu}(t) \mathrm{d}\mu(t) \leq L \Delta(T) + L \int_{T} \sum_{n \geq 1} 2^{n/2} \Delta(A_n(t)) \mathrm{d}\mu(t) .$$
(3.36)

Integrating (3.35) with respect to μ proves that the last term of (3.36) is $\leq K\gamma_2(T,d)$ and concludes the proof since $\Delta(T) \leq L\gamma_2(T,d)$.

3.3.2 Fernique's Convexity Argument

Our approach to the left-hand side of (3.31) is based on the following elementary fact, which is a consequence of the Hahn-Banach theorem.

Lemma 3.3.2. Consider a number a > 0. Consider a set S of functions on a finite set T. Assume that for each probability measure ν on T there exists $f \in S$ such that $\int f d\nu \leq a$. Then for each $\epsilon > 0$ there is a function f in the convex hull of S such that $f \leq a + \varepsilon$.

96 3. Trees and Other Measures of Size

Proof. Denote S^+ the set of functions g such that there exists $f \in S$ with $f \leq g$. Denote by C the closed convex hull of S^+ . We prove that the constant function \mathbf{a} equal to a everywhere belongs to C. We proceed by contradiction. If this is not the case, by the Hahn-Banach theorem we may separate C and \mathbf{a} . That is, there exists a linear functional φ on the space of functions on T such that $\varphi(f) > \varphi(\mathbf{a})$ for $f \in C$. Consider then a function g on T with $g \geq 0$. For each $\lambda > 0$ and each $f \in S$ we have $f + \lambda g \geq f$ so that $f + \lambda g \in C$ and hence $\varphi(f) + \lambda \varphi(g) > \varphi(\mathbf{a})$. This proves that $\varphi(g) \geq 0$, i.e. φ is positive. Since T is finite, φ is of the form $\varphi(g) = \sum_{t \in T} \alpha_t g(t)$ for numbers $\alpha_t \geq 0$. Setting $\beta = \sum_{t \in T} \alpha(t)$ consider the probability measure ν on T given by $\nu(\{t\}) = \alpha_t/\beta$ for $t \in T$. Then $\varphi(g) = \beta \int g d\nu$ for each function g on T. Taking $g = \mathbf{a}$ shows that $\beta a = \varphi(\mathbf{a})$. By hypothesis there exists $f \in C$ with $\int f d\nu \leq a$. Then $\varphi(f) = \beta \int f d\nu \leq \beta a = \varphi(\mathbf{a})$, a contradiction.

So we have proved that $\mathbf{a} \in \mathcal{C}$, the *closure* of the convex hull of \mathcal{S}^+ . Consequently there is one point of this convex hull which is $\leq a + \epsilon$ everywhere. The result follows.

Of course, the hypothesis that T is finite is inessential, it is just to avoid secondary complications.

Let us give a version of the basic lemma sufficiently general to cover all our needs.

Lemma 3.3.3. Consider a finite metric space (T, d). Consider a convex function $\Phi : [0, 1] \to \mathbb{R}^+$. Assume that for each probability measure μ on T and a certain number D one has

$$\int_{T} \mathrm{d}\mu(t) \int_{0}^{\Delta(T)} \Phi(\mu(B(t,\epsilon))) \mathrm{d}\epsilon \le D .$$
(3.37)

Then there exists a probability measure μ on T for which

$$\sup_{t \in T} \int_0^{\Delta(T)} \varPhi(\mu(B(t,\epsilon))) \mathrm{d}\epsilon \le 2D .$$
(3.38)

Proof. Let us denote by $\mathcal{M}(T)$ the set of probability measures on T. The class \mathcal{C} of functions f on T that satisfy

$$\exists \mu \in \mathcal{M}(T) \; ; \; \forall t \in T \; , \; f_{\mu}(t) := \int_{0}^{\Delta(T)} \varPhi(\mu(B(t,\epsilon))) \mathrm{d}\epsilon \leq f(t)$$

is convex. This is immediate to check using the convexity of Φ . For each probability measure ν on T, there exists f in \mathcal{C} with $\int f d\nu \leq B$: this is true for $f = f_{\nu}$ by (3.37). Consequently by Lemma 3.3.2, there exists $f \in \mathcal{C}$ such that $f \leq 2B$, which is the content of the lemma.

Corollary 3.3.4. Consider a finite metric space (T, d). Assume that for a certain number C and for each probability measure μ on T we have

3.3 Fernique's Functional 97

$$\int_{T} I_{\mu}(t) \mathrm{d}\mu(t) \le C .$$
(3.39)

Then there is probability measure μ on T such that

$$\forall t \in T , \ I_{\mu}(t) \le 2C + 2\Delta(T) . \tag{3.40}$$

Proof. Calculus shows that the function $\Phi(x) = \sqrt{\log(e/x)}$ is convex for $x \in [0,1]$, and $\sqrt{\log(1/x)} \leq \Phi(x) \leq 1 + \sqrt{\log(1/x)}$. Thus

$$I_{\mu}(t) \leq \int_{0}^{\Delta(T)} \Phi(\mu(B(t,\epsilon))) \mathrm{d}\epsilon \leq \Delta(T) + I_{\mu}(t) \;,$$

so that (3.39) implies that (3.37) holds for $D = C + \Delta(T)$ and (3.40) then follows from (3.38).

Lemma 3.3.5. We have $\Delta(T) \leq LFer(T, d)$.

Proof. Consider $s, u \in T$ with $d(s, u) \geq \Delta(T)/2$ and the probability μ on T such that $\mu(\{s\}) = \mu(\{u\}) = 1/2$. For $\epsilon < \Delta(T)/2 \leq d(s, u)$ we have $\mu(B(s, \epsilon)) \leq 1/2$ and this implies that $I_{\mu}(s) \geq \Delta(T)\sqrt{\log 2}/2$. Similarly we have $I_{\mu}(u) \geq \Delta(T)/L$ so that $\int_{T} I_{\mu}(t) d\mu(t) \geq \Delta(T)/L$.

Proof of (3.31) when T is finite. Combining Corollary 3.3.4 and Lemma 3.3.5 we obtain that there exists a probability μ on T such that $\sup_{t\in T} I_{\mu}(t) \leq L \operatorname{Fer}(T, d)$. On the other hand we have proved in Section 3.1 that $\gamma_2(T, d) \leq L \sup_{t\in T} I_{\mu}(t)$.

3.3.3 From Majorizing Measures to Sequences of Partitions

In Section 3.1 we have proved that given a probability measure μ on a metric space T we have

$$\gamma_2(T,d) \le L \sup_{t \in T} I_\mu(t) . \tag{3.41}$$

We do not know how to generalize the arguments of the proof to the more general settings we will consider later. We give now a direct proof, following a scheme which we know how to generalize. The contents of this section will not be relevant until Chapter 11. First, we prove that

$$\Delta(T) \le L \sup_{t \in T} I_{\mu}(t) .$$
(3.42)

For this we consider $s, t \in T$ with $d(s,t) > \Delta(T)/2$ so that since the balls $B(t, \Delta(T)/2)$ and $B(s, \Delta(T)/2)$ are disjoint, one of them, say the first one,

⁵ The argument by which we have proved this inequality will not generalize, but fortunately there is another route, which is described in the next section.
98 3. Trees and Other Measures of Size

has a measure $\leq 1/2$. Then $\mu(B(t,\epsilon)) \leq 1/2$ for $\epsilon \leq \Delta(T)/2$ and thus $I_{\mu}(t) \geq \sqrt{\log 2\Delta(T)/2}$.

We start the main argument. We will construct an admissible sequence (\mathcal{A}_n) of partitions of T which witnesses (3.41). For $A \in \mathcal{A}_n$ we also construct an integer $j_n(A)$ as follows. First, we set $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$ and $j_0(T) = j_1(T) = j_0$, the largest integer with $\Delta(T) \leq 2^{-j_0(T)}$. Next for $n \geq 1$ we require the conditions

$$A \in \mathcal{A}_n \Rightarrow \Delta(A) \le 2^{-j_n(A)+2} , \qquad (3.43)$$

$$t \in A \in \mathcal{A}_n \Rightarrow \mu(B(t, 2^{-j_n(A)})) \ge N_{n-1}^{-1} .$$
(3.44)

The construction proceeds as follows. Having constructed \mathcal{A}_n we split each element A of \mathcal{A}_n into at most N_n pieces, ensuring that card $\mathcal{A}_{n+1} \leq N_n^2 = N_{n+1}$. For this we set

$$A_0 = \{t \in A; \mu(B(t, 2^{-j_n(A)-1})) \ge 1/N_n\}.$$
(3.45)

We claim first that we may cover A_0 by $< N_n$ sets, each of diameter $\leq 2^{-j_n(A)+1}$. For this we consider a subset W of A_0 , maximal with respect to the property that any two points of W are at distance $> 2^{-j_n(A)}$. The balls of radius $2^{-j_n(A)-1}$ centered at the points of W are disjoint, and each of them is of measure $> N_n^{-1}$ by (3.45), so that there are $< N_n$ of them. Since W is maximum, the balls of radius $2^{-j_n(A)+1}$ centered at the points of W cover A_0 , and each of them has diameter $\le 2^{-j_n(A)+1}$. Thus there exists a partition of A_0 in $< N_n$ sets of diameter $\le 2^{-j_n(A)+1}$. The required partition of A consists of these sets B and of $A_1 = A \setminus A_0$. For each set B we set $j_{n+1}(B) = j_n(A)+1$, and we set $j_{n+1}(A_1) = j_n(A)$, so that conditions (3.43) and 3.44 hold.

This completes the construction. The important point is that

$$B \in \mathcal{A}_{n+1}, B \subset A \in A_n, j_{n+1}(B) = j_n(A) \Rightarrow \mu(B(t, 2^{-j_{n+1}(B)}))$$
$$= \mu(B(t, 2^{-j_n(A)})) \le N_n^{-1}. \quad (3.46)$$

This property holds because if $t \in A$ and $\mu(B(t, 2^{-j_n(A)})) > N_n^{-1}$ then $t \in A_0$ and the element B of \mathcal{A}_{n+1} which contains t has been assigned a value $j_{n+1}(B) = j_n(A) + 1$.

To prove (3.41) we will prove that

$$\forall t \in T \ , \ \sum_{n \geq 0} 2^{n/2} \varDelta(A_n(t)) \leq L I_\mu(t) \ .$$

We fix $t \in T$. We set $j(n) = j_n(A_n(t))$, and $a(n) = 2^{n/2}2^{-j(n)}$. Using (3.43) it suffices to prove that

$$\sum_{n \ge 0} a(n) \le LI_{\mu}(t) .$$
 (3.47)

We consider $\alpha = \sqrt{2}$ and the corresponding set *I* as in (2.84). (Leaving to the reader to prove that the sequence (a(n)) is bounded, which will become clear soon.) Thus as in (2.91) we have

$$n \in I, n \ge 1 \Rightarrow j(n+1) = j(n) + 1, \ j(n-1) = j(n).$$
 (3.48)

We enumerate $I \setminus \{0\}$ as a sequence $(n_k)_{k\geq 0}$ (leaving again to the reader the easier case where I is finite), so that $j(n_{k+1}) \geq j(n_k + 1) = j(n_k) + 1$ and

$$\sum_{n \ge 0} a(n) \le La(0) + L \sum_{k \ge 1} a(n_k) .$$
(3.49)

From (3.48) we have $j(n_k-1) = j(n_k)$. Using (3.46) for $n = n_k - 1$ we obtain

$$\mu(B(t, 2^{-j(n_k)})) < N_{n_k-1}^{-1}$$
,

so that $\sqrt{\log(1/\mu(B(t,\epsilon)))} \ge 2^{n_k/2}/L$ for $\epsilon < 2^{-j(n_k)}$. Since $j(n_{k+1}) > j(n_k)$ this implies

$$\begin{aligned} a(n_k) &= 2^{n_k/2 - j(n_k)} \le 2 \cdot 2^{n_k/2} (2^{-j(n_k)} - 2^{j(n_{k+1})}) \\ &\le L \int_{2^{-j(n_k)}}^{2^{-j(n_k)}} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}} \mathrm{d}\mu(\epsilon) \end{aligned}$$

Summation of these inequalities and use of (3.42) and (3.49) proves (3.47).

3.4 Witnessing Measures

Proposition 3.4.1. ([91]) For a metric space (T, d), define

$$\delta_2(T,d) = \sup_{\mu} \inf_{t \in T} I_{\mu}(t) , \qquad (3.50)$$

where the supremum is taken over all probability measures μ on T.⁶ Then

$$\frac{1}{L}\gamma_2(T,d) \le \delta_2(T,d) \le L\gamma_2(T,d) .$$
(3.51)

It is obvious that $\inf_{t\in T} I_{\mu}(t) \leq \int_{T} I_{\mu}(t) d\mu(t)$ so that $\delta_{2}(T, d) \leq \text{Fer}(T, d)$. Thus the right-hand side of (3.51) follows from the right-hand side of (3.31) while the left-hand side of (3.31) follows from the left-hand side of (3.51).

The most important consequence of (3.31) is that there exists a probability measure μ on T for which $\inf_{t \in T} I_{\mu}(t) \geq \gamma_2(T, d)/L$. Such a probability

 $^{^{6}}$ Please observe that the order of the infimum and the supremum is not as in (3.20).

100 3. Trees and Other Measures of Size

measure "witnesses that the value of $\gamma_2(T, d)$ is large" because of the righthand side of (3.51). In this spirit, we will call μ a witnessing measure, and we define its "size" as the quantity $\inf_{t \in T} I_{\mu}(t)$.⁷ Witnessing measures can be magically convenient. One of the first advances the author made beyond the results of [171] was the realization that witnessing measures yield a proof of Theorem 5.2.1 below an order of magnitude easier than the original proof. This approach is now replaced by the use of Fernique's functional because, unfortunately we do not know how to extend the idea of witnessing measure to settings where multiple distances will be considered. Finding proper generalizations of Proposition 3.4.1 to more general settings is an attractive research problem, see in particular Problem 10.15.4.

Proof of Proposition 3.4.1. The right-hand side inequality follows from Proposition 3.3.1 and the trivial fact that $\inf_{t\in T} I_{\mu}(t) \leq \int I_{\mu}(t) d\mu(t)$. The reader should review the material of Section 3.1 to follow the proof of the converse. Recalling (3.5), given an organized tree \mathcal{T} we define a measure μ on T by $\mu(A) = 0$ if $A \cap S_{\mathcal{T}} = \emptyset$ and by

$$t \in S_{\mathcal{T}} \Rightarrow \mu(\{t\}) = \frac{1}{\prod_{t \in A \in \mathcal{T}} c(A)}$$

The intuition is that the mass carried by $A \in \mathcal{T}$ is equally divided between the children of A. Then $I_{\mu}(t) = \infty$ if $t \notin S_{\mathcal{T}}$. Consider $t \in A \in \mathcal{T}$ and j = j(A). Then, since \mathcal{T} is an organized tree, $B(t, 4^{-j-2})$ meets only one child of A, so that $\mu(B(t, 4^{-j-2})) \leq 1/c(A)$. Copying the argument of (3.22) readily implies that $LI_{\mu}(t) \geq \sum_{t \in A \in \mathcal{T}} 4^{-j(A)} \sqrt{\log c(A)}$ from which the result follows by (3.15).

Exercise 3.4.2. For a metric space (T, d) define

$$\chi_2(T,d) = \sup_{\mu} \inf \int \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \mathrm{d}\mu(t)$$

where the infimum is taken over all admissible sequences and the supremum over all probability measures. It is obvious that $\chi_2(T,d) \leq \gamma_2(T,d)$. Prove that $\gamma_2(T,d) \leq L\chi_2(T,d)$. Hint: Prove that the functional $\chi_2(T,d)$ satisfies the appropriate growth condition. Warning: the argument takes about half a page and is fairly non-trivial.

3.5 An inequality of Fernique

We end up this chapter with a beautiful inequality of Fernique. It will not be used anywhere else in this work, but is presented to emphasize Fernique's lasting contributions to the theory of Gaussian processes.

⁷ Thus a probability measure μ on T is both a majorizing and a witnessing measure. It bounds $\gamma_2(T, d)$ from above by $L \sup_{t \in T} I_{\mu}(t)$ and from below by $\inf_{t \in T} I_{\mu}(t)/L$. Furthermore one may find μ such that these two bounds are of the same order.

Theorem 3.5.1. Consider a Gaussian process $(X_t)_{t \in T}$. Provide T with the canonical distance associated to this process and consider a probability measure μ on T. Consider a r.v. τ of law μ . Then for any probability measure ν on T we have

$$\mathsf{E}X_{\tau} \le L \int I_{\nu}(t) \mathrm{d}\mu(t) + L\Delta(T, d) . \qquad (3.52)$$

Here of course X_{τ} is the r.v. $X_{\tau(\omega)}(\omega)$. We leave some technical details aside and prove the result only when T is finite. The basic principle is as follows.

Lemma 3.5.2. Consider a standard Gaussian r.v. g and a set A. Then $\mathsf{E1}_A|g| \leq L\mathsf{P}(A)\sqrt{\log(2/\mathsf{P}(A))}$.

Proof. We write

$$\mathsf{E1}_A|g| = \int_0^\infty \mathsf{P}(A \cap \{|g| \ge t\}) \mathrm{d}t \le \int_0^\infty \min(\mathsf{P}(A), 2\exp(-t^2/2)) \mathrm{d}t \ .$$

Letting $\alpha = \sqrt{2 \log(2/\mathsf{P}(A))}$ we split the integral in the regions $t \leq \alpha$ and $t > \alpha$. We bound the first part by $\alpha \mathsf{P}(A)$ and the second by

$$\int_{\alpha}^{\infty} 2\exp(-t^2/2) \mathrm{d}t \le \frac{1}{\alpha} \int_{\alpha}^{\infty} 2t \exp(-t^2/2) \mathrm{d}t = \mathsf{P}(A)/\alpha \le L\mathsf{P}(A)\alpha \ . \quad \Box$$

Corollary 3.5.3. Consider Gaussian r.v.s $(g_i)_{i \leq N}$ with $\mathsf{E}g_i^2 \leq a^2$. Consider a r.v. τ valued in $\{1, \ldots, N\}$, and let $\alpha_i = \mathsf{P}(\tau = i)$. Then $\mathsf{E}|g_{\tau}| \leq La \sum_{i \leq N} \alpha_i \sqrt{\log 2/\alpha_i}$.

Proof. Since $|g_{\tau}| = \sum_{i \leq N} \mathbf{1}_{\{\tau=i\}} |g_i|$, and using Lemma 3.5.2 to obtain $\mathsf{E1}_{\{\tau=i\}} |g_i| \leq La\alpha_i \sqrt{\log 2/\alpha_i}$.

We also need the following elementary convexity inequality.

Lemma 3.5.4. Consider numbers $\alpha_i > 0$ with $\sum_{i \le N} \alpha_i \le \alpha \le 1$. Then

$$\sum_{i \le N} \alpha_i \sqrt{\log(2/\alpha_i)} \le \alpha \sqrt{\log(2N/\alpha)} .$$
(3.53)

Proof. Calculus shows that the function $\varphi(x) = x\sqrt{\log(2/x)}$ is concave increasing for $x \leq 1$, so that if $\alpha' = \sum_{i \leq N} \alpha_i$ then $N^{-1} \sum_{i \leq N} \varphi(\alpha_i) \leq \varphi(\alpha'/N) \leq \varphi(\alpha/N)$. \Box

The slightly technical part of the proof of Theorem 3.5.1 is contained in the following.

Lemma 3.5.5. Consider probability measures μ and ν on a metric space (T, d). Then there exist $n_0 \in \mathbb{Z}$ and a sequence of partitions $(\mathcal{B}_n)_{n \geq n_0}$ (which need not be increasing) of T with the following properties. First, \mathcal{B}_{n_0} contains only one set. Next, the sets of \mathcal{B}_n are of diameter $\leq 2^{-n+2}$. Finally,

102 3. Trees and Other Measures of Size

$$\sum_{n \ge n_0} 2^{-n} \sum_{A \in \mathcal{B}_n, B \in \mathcal{B}_{n+1}} \mu(A \cap B) \sqrt{\log(2/\mu(A \cap B))}$$
$$\leq L \int I_{\nu}(t) \mathrm{d}\mu(t) + L\Delta(T, d) . \quad (3.54)$$

Proof. We consider the largest integer n_0 with $2^{-n_0} \ge \Delta(T, d)$. We set $\mathcal{B}_{n_0} = \{T\}$. For $n > n_0$ we proceed as follows. For $k \ge 0$ we set

$$T_{n,k} = \{t \in T ; 1/N_{k+1} < \nu(B(t, 2^{-n})) \le 1/N_k\}$$

The sets $(T_{n,k})_{k\geq 0}$ form a partition of T. Consider a subset V of $T_{n,k}$ such that the points of V are at mutual distance $> 2^{-n+1}$. The balls of radius 2^{-n} centered at the points of V are disjoint and by definition of $T_{n,k}$ they have a ν -measure $> 1/N_{k+1}$. Thus card $V \leq N_{k+1}$ and according to Lemma 2.9.3, $e_{k+1}(T_{n,k}) \leq 2^{-n+1}$ and thus $T_{n,k}$ can be partitioned into at most N_{k+1} sets of diameter $\leq 2^{-n+2}$. We construct such a partition $\mathcal{B}_{n,k}$ of $T_{n,k}$ for each k and we consider the corresponding partition \mathcal{B}_n of T.

We now turn to the proof of (3.54). First we note that $\operatorname{card} \mathcal{B}_{n,k} \leq N_{k+1}$ and $\operatorname{card} \mathcal{B}_{n+1,\ell} \leq N_{\ell+1}$. Also $\sum_{A \in \mathcal{B}_{n,k}, B \in \mathcal{B}_{n+1,\ell}} \mu(A \cap B) \leq \mu(T_{n,k} \cap T_{n+1,\ell})$. We then use (3.53) to obtain

$$S_{n,k,\ell} := \sum_{A \in \mathcal{B}_{n,k}, B \in \mathcal{B}_{n+1,\ell}} \mu(A \cap B) \sqrt{\log(2/\mu(A \cap B))}$$

$$\leq \mu(T_{n,k} \cap T_{n+1,\ell}) \sqrt{\log(2N_{k+1}N_{\ell+1}/\mu(T_{n,k} \cap T_{n+1,\ell}))} .$$

The left-hand side of (3.54) is $\sum_{n \ge n_0} 2^{-n} \sum_{k,\ell} S_{n,k,\ell}$. We will use the decomposition

$$\sum_{k,\ell} S_{n,k,\ell} = \sum_{(k,\ell)\in I(n)} S_{n,k,\ell} + \sum_{(k,\ell)\in J(n)} S_{n,k,\ell}$$

where $I(n) = \{(k, \ell); \mu(T_{n,k} \cap T_{n+1,\ell}) \ge 1/(N_{k+1}N_{\ell+1})\}$ and $J(n) = \{(k, \ell); \mu(T_{n,k} \cap T_{n+1,\ell}) < 1/(N_{k+1}N_{\ell+1})\}$. Then

$$\sum_{(k,\ell)\in I(n)} S_{n,k,\ell} \leq \sum_{k,\ell} \mu(T_{n,k} \cap T_{n+1,\ell}) \sqrt{\log(2N_{k+1}^2 N_{\ell+1}^2)}$$
$$\leq L \sum_{k,\ell} \mu(T_{n,k} \cap T_{n+1,\ell}) (2^{k/2} + 2^{\ell/2})$$
$$\leq L \sum_k \mu(T_{n,k}) 2^{k/2} + L \sum_\ell \mu(T_{n+1,\ell}) 2^{\ell/2} . \quad (3.55)$$

Now the definition of $T_{n,k}$ shows that

$$\sum_{k \ge 1} \mu(T_{n,k}) 2^{k/2} \le \int \sqrt{\log(1/\nu(B(t,2^{-n})))} d\mu(t) , \qquad (3.56)$$

and thus

$$\sum_{n \ge n_0} 2^{-n} \sum_{k \ge 1} 2^{k/2} \mu(T_{n,k}) \le L \int I_{\nu}(t) \mathrm{d}\mu(t) \; .$$

Next, since the function $\varphi(x) = x\sqrt{\log 2N_{k+1}N_{\ell+1}/x}$ increases for $x \leq N_{k+1}N_{\ell+1}$, for $(k,\ell) \in J(n)$ we have $\varphi(\mu(T_{n,k} \cap T_{n+1,\ell})) \leq \varphi(1/(N_{k+1}N_{\ell+1}))$ so that

$$\sum_{(k,\ell)\in J(n)} S_{n,k,\ell} \leq \sum_{(k,\ell)\in J(n)} \varphi(\mu(T_{n,k}\cap T_{n+1,\ell})) \leq \sum_{(k,\ell)\in J(n)} \varphi(1/N_{k+1}N_{\ell+1})$$
$$\leq L \sum_{k,\ell} \frac{2^{k/2} + 2^{\ell/2}}{N_{k+1}N_{\ell+1}} \leq L .$$
(3.57)

Combining these estimates yields the desired inequality:

$$\sum_{n \ge n_0} 2^{-n} \sum_{k,\ell \ge 0} S_{n,k,\ell} \le L \int I_{\nu}(t) \mathrm{d}\mu(t) + L 2^{-n_0} .$$

Proof of Theorem 3.5.1. For $n \ge n_0$ and $A \in \mathcal{B}_n$ we fix an arbitrary point $t_{n,A} \in A$. We lighten notation by writing $t_0 = t_{n_0,T}$. We define $\pi_n(t) = t_{n,A}$ for $t \in A \in \mathcal{B}_n$. We write

$$X_{\tau} - X_{t_0} = \sum_{n \ge n_0} X_{\pi_{n+1}(\tau)} - X_{\pi_n(\tau)} , \qquad (3.58)$$

so that defining $Y_{n,\tau} := X_{\pi_{n+1}(\tau)} - X_{\pi_n(\tau)}$ we have

$$\mathsf{E}|X_{\tau} - X_{t_0}| \le \sum_{n \ge n_0} \mathsf{E}|Y_{n,\tau}| .$$
(3.59)

Given $A \in \mathcal{B}_n$ and $B \in \mathcal{B}_{n+1}$ let us define the variable $Y_{A,B} := X_{t_{n+1,B}} - X_{t_{n,A}}$. The sets $A \cap B$ for $A \in \mathcal{B}_n$ and $B \in \mathcal{B}_{n+1}$ form a partition of T. When $\tau(\omega) \in A \cap B$ we have $\pi_n(\tau) = t_{n,A}$ and $\pi_{n+1}(\tau) = t_{n+1,B}$ so that $Y_{n,\tau} = Y_{A,B}$. The event $\tau(\omega) \in A \cap B$ has probability $\mu(A \cap B)$ since τ has law μ . When $A \cap B \neq \emptyset$ we have $d(t_{n+1,B}, t_{n,A}) \leq \Delta(A) + \Delta(B) \leq L2^{-n}$, so that $\mathbb{E}Y_{A,B}^2 = d(t_{n+1,B}, t_{n,A})^2 \leq L2^{-2n}$. It then follows from Corollary 3.5.3 that

$$\mathsf{E}|Y_{n,\tau}| \le L2^{-n} \sum_{A \in \mathcal{B}_n, B \in \mathcal{B}_{n+1}} \mu(A \cap B) \sqrt{\log(2/\mu(A \cap B))} ,$$

and summation over n and use of (3.54) finishes the proof.

It is actually possible to give a complete geometric description of the quantity $\sup_{\tau} \mathsf{E} X_{\tau}$ where the supremum is taken over all the τ of given law μ , see [137].

Key ideas to remember.

104 3. Trees and Other Measures of Size

- Trees in a metric space (T, d) are well-separated structures which are easy to visualize, and provide a convenient way to measure the size of this metric space, by the size of the largest tree it contains. For suitable classes of trees, this measure of size is equivalent to $\gamma_2(T, d)$.
- One may also measure the size of a metric space by the existence of certain probability measures on this space. Fernique's majorizing measures were used early to control from above the size of a metric space in a way very similar to the functional $\gamma_2(T, d)$, which is however far more technically convenient than majorizing measures.
- An offshoot of the idea of majorizing measures, Fernique's functional, is an equivalent way to measure the size of a metric space and will be of fundamental importance in the sequel.
- The size of a metric space (T, d) can also be bounded from below by the existence of well-scattered probability measures on T.

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

4.1 The Ellipsoid Theorem

As pointed out after Proposition 2.13.2, an ellipsoid \mathcal{E} is in some sense quite smaller than what one would predict by looking only at the numbers $e_n(\mathcal{E})$. We will trace the roots of this phenomenon to a simple geometric property, namely that an ellipsoid is "sufficiently convex", and we will formulate a general version of this principle for sufficiently convex bodies. The case of ellipsoids already suffices to provide tight upper bounds on certain matchings, which is the main goal of the present chapter. The general case is at the root of certain very deep facts of Banach space theory, such as Bourgain's celebrated solution of the Λ_p -problem in Sections 19.3.1 and 19.3.2.

Recall the ellipsoid \mathcal{E} of (2.154), which is defined as the set

$$\mathcal{E} = \left\{ t \in \ell^2 \; ; \; \sum_{i \ge 1} \frac{t_i^2}{a_i^2} \le 1 \right\}$$
(2.154)

and is the unit ball of the norm

$$\|x\|_{\mathcal{E}} := \left(\sum_{i\geq 1} \frac{x_i^2}{a_i^2}\right)^{1/2}.$$
(4.1)

Lemma 4.1.1. We have

$$\|x\|_{\mathcal{E}}, \|y\|_{\mathcal{E}} \le 1 \Rightarrow \left\|\frac{x+y}{2}\right\|_{\mathcal{E}} \le 1 - \frac{\|x-y\|_{\mathcal{E}}^2}{8}.$$
 (4.2)

Proof. The parallelogram identity implies

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

so that

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

and

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

Since (4.2) is the only property of ellipsoids we will use, it clarifies matters to state the following definition.

Definition 4.1.2. Consider a number $p \ge 2$. A norm $\|\cdot\|$ in a Banach space is called p-convex if for a certain number $\eta > 0$ we have

$$||x||, ||y|| \le 1 \Rightarrow \left|\left|\frac{x+y}{2}\right|\right| \le 1 - \eta ||x-y||^p$$
. (4.3)

Saying just that the unit ball of the Banach space is convex implies that for $||x||, ||y|| \le 1$ we have $||(x+y)/2|| \le 1$. Here (4.3) quantitatively improves on this inequality. Geometrically it means that the unit ball of the Banach space is "round enough".

Thus (4.2) implies that the Banach space ℓ^2 provided with the norm $\|\cdot\|_{\mathcal{E}}$ is 2-convex. For $1 < q < \infty$ the classical Banach space L^q is *p*-convex where $p = \max(2, q)$. The reader is referred to [80] for this result and any other classical facts about Banach spaces. Let us observe that, taking y = -x in (4.3) we must have

$$2^p \eta \le 1 . \tag{4.4}$$

In this section we shall study the metric space (T, d) where T is the unit ball of a *p*-convex Banach space B, and where d is the distance induced on B by another norm $\|\cdot\|_{\sim}$. This concerns in particular the case where T is the ellipsoid (2.154) and $\|\cdot\|_{\sim}$ is the ℓ^2 norm.

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

$$\gamma_{\alpha,\beta}(T,d) = \inf\left(\sup_{t\in T} \sum_{n\geq 0} \left(2^{n/\alpha} \Delta(A_n(t))\right)^\beta\right)^{1/\beta},\tag{4.5}$$

where α and β are positive numbers, and where the infimum is over all admissible sequences (\mathcal{A}_n) . Thus, with the notation of Definition 2.7.3, we have $\gamma_{\alpha,1}(T,d) = \gamma_{\alpha}(T,d)$. For matchings, the important functionals are $\gamma_{2,2}(T,d)$ and $\gamma_{1,2}(T,d)$ (but it requires no extra effort to consider the general case). The importance of these functionals is that under certain conditions they nicely relate to $\gamma_2(T,d)$ through Hölder's inequality. We explain right now how this is done, even though this spoils the surprise of how the terms $\sqrt{\log N}$ occur in Section 4.5.

Lemma 4.1.3. Consider a finite metric space T, and assume that card $T \leq N_m$. Then

$$\gamma_2(T,d) \le \sqrt{m}\gamma_{2,2}(T,d) . \tag{4.6}$$

Proof. Since T is finite there exists¹ an admissible sequence (\mathcal{A}_n) of T for which

$$\forall t \in T, \sum_{n \ge 0} 2^n \Delta(A_n(t))^2 \le \gamma_{2,2}(T,d)^2.$$
(4.7)

Since card $T \leq N_m$, we may assume that \mathcal{A}_m consists of all the sets $\{t\}$ for $t \in T$. Then $A_m(t) = \{t\}$ for each t, so that in (4.7) the sum is really over $n \leq m-1$. Since for any numbers $(a_n)_{0 \leq n \leq m-1}$ we have $\sum_{0 \leq n \leq m-1} a_n \leq \sqrt{m} (\sum_{0 \leq n \leq m-1} a_n^2)^{1/2}$ by the Cauchy-Schwarz inequality, it follows that

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

How to relate the functionals $\gamma_{1,2}$ and γ_2 by a similar argument is shown in Lemma 4.7.9 below.

We may wonder how it is possible, using something as simple as the Cauchy-Schwarz inequality in Lemma 4.1.3, that we can ever get essentially exact results. At a general level the answer is obvious: it is because we use this inequality in the case of near equality. That this is indeed the case for the ellipsoids of Corollary 4.1.7 below is a non-trivial fact about the geometry of these ellipsoids.

Theorem 4.1.4. If T is the unit ball of a p-convex Banach space, if η is as in (4.3) and if the distance d on T is induced by another norm, then

$$\gamma_{\alpha,p}(T,d) \le K(\alpha,p,\eta) \sup_{n \ge 0} 2^{n/\alpha} e_n(T,d) .$$
(4.8)

Before we prove this result (in Section 4.2) we explore some of its consequences. The following exercise stresses the main point of this theorem.

Exercise 4.1.5. Consider a general metric space (T, d). (a) Prove that

$$\gamma_{\alpha,p}(T,d) \le K(\alpha) \left(\sum_{n \ge 0} \left(2^{n/\alpha} e_n(T,d) \right)^p \right)^{1/p}, \tag{4.9}$$

and that

$$\sup_{n \ge 0} 2^{n/\alpha} e_n(T, d) \le K(\alpha) \gamma_{\alpha, p}(T, d) .$$
(4.10)

(b) Prove that it is essentially impossible in general to improve on (4.9). Hint: you probably want to review Chapter 3 before you try this.

¹ Since there are only finitely many admissible sequences, the infimum over these is achieved.

Thus, knowing only the numbers $e_n(T,d)$, we would expect only the general bound (4.9). The content of Theorem 4.1.4 is that the size of T, as measured by the functional $\gamma_{\alpha,p}$, is actually much smaller than that.

Corollary 4.1.6. (The Ellipsoid Theorem.) Consider the ellipsoid \mathcal{E} of (2.154) and $\alpha \geq 1$. Then²

$$\gamma_{\alpha,2}(\mathcal{E}) \le K(\alpha) \sup_{\epsilon > 0} \epsilon(\operatorname{card}\{i \; ; \; a_i \ge \epsilon\})^{1/\alpha}.$$
(4.11)

Proof. Without loss of generality we may assume that the sequence (a_i) is non-increasing. We apply Theorem 4.1.4 to the case $\|\cdot\| = \|\cdot\|_{\mathcal{E}}$, and where d is the distance of ℓ^2 , and we get

$$\gamma_{\alpha,2}(\mathcal{E}) \leq K(\alpha) \sup_{n \geq 0} 2^{n/\alpha} e_n(\mathcal{E})$$

To bound the right-hand side we write

$$\sup_{n \ge 0} 2^{n/\alpha} e_n(\mathcal{E}) \le 2^{2/\alpha} e_0(\mathcal{E}) + \sup_{n \ge 0} 2^{(n+3)/\alpha} e_{n+3}(\mathcal{E})$$

We now proceed as in the proof of Proposition 2.13.2. Using (2.166), we have

$$\sup_{n \ge 0} 2^{(n+3)/\alpha} e_{n+3}(\mathcal{E}) \le K(\alpha) \sup_{n \ge 0} 2^{n/\alpha} \max_{k \le n} 2^{k-n} a_{2^k}$$
$$= K(\alpha) \sup_{0 \le k \le n} 2^{k-n(1-1/\alpha)} a_{2^k}$$
$$= K(\alpha) \sup_{k \ge 0} 2^{k/\alpha} a_{2^k} , \qquad (4.12)$$

and since $e_0(\mathcal{E}) \leq a_1$ we have proved that $\gamma_{\alpha,2}(\mathcal{E}) \leq K(\alpha) \sup_{n\geq 0} 2^{n/\alpha} a_{2^n}$. Finally, the choice $\epsilon = a_{2^n}$ shows that

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

since $\operatorname{card}\{i; a_i \ge a_{2^n}\} \ge 2^n$ because the sequence (a_i) is non-increasing. \Box

The restriction $\alpha \geq 1$ is inessential and can be removed by a suitable modification of (2.166). The important cases are $\alpha = 1$ and $\alpha = 2$. We will use the following convenient reformulation.

Corollary 4.1.7. Consider a countable set J, numbers $(b_i)_{i \in J}$ and the ellipsoid

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

Then

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

 $^{^2}$ Recalling that a subset of ℓ^2 is always provided with the distance induced by the ℓ^2 norm.

Proof. Without loss of generality we can assume that $J = \mathbb{N}$. We then set $a_i = 1/b_i$, we apply Corollary 4.1.6, and we set $\epsilon = 1/u$.

We give right away a striking application of this result. This application is at the root of the results of Section 4.7. We denote by λ Lebesgue's measure.

Proposition 4.1.8. Consider the set \mathcal{L} of functions $f : [0,1] \to \mathbb{R}$ such that f(0) = f(1) = 0, f is continuous on [0,1], f is differentiable outside a finite set and $\sup |f'| \leq 1.^3$ Then $\gamma_{1,2}(\mathcal{L}, d_2) \leq L$, where $d_2(f,g) = ||f - g||_2 = (\int_{[0,1]} (f - g)^2 d\lambda)^{1/2}$.

Proof. The very beautiful idea (due to Coffman and Shor [36]) is to use the Fourier transform to represent \mathcal{L} as a subset of an ellipsoid. The Fourier coefficients of a function $f \in \mathcal{L}$ are defined for $p \in \mathbb{Z}$ by

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

The key fact is the Plancherel formula,

$$||f||_{2} = \left(\sum_{p \in \mathbb{Z}} |c_{p}(f)|^{2}\right)^{1/2}, \qquad (4.13)$$

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

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

the metric space (\mathcal{L}, d_2) is isometric to a subspace of (\mathcal{D}, d) , where d is the distance induced by $\ell^2_{\mathbb{C}}(\mathbb{Z})$. It is then obvious from the definition that $\gamma_{1,2}(\mathcal{L}, d_2) \leq \gamma_{1,2}(\mathcal{D}, d)$, so that it suffices to prove that $\gamma_{1,2}(\mathcal{D}, d) < \infty$. By integration by parts, and since f(0) = f(1) = 0, $c_p(f') = -2\pi i p c_p(f)$, so that, using (4.13) for f', we get

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

For $f \in \mathcal{L}$ we have f(0) = 0 and $|f'| \leq 1$ so that $|f| \leq 1$ and $|c_0(f)| \leq 1$. Thus for $f \in \mathcal{L}$ we have

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

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

$$\mathcal{E} := \left\{ (c_p) \in \ell^2_{\mathbb{C}}(\mathbb{Z}) \; ; \; \sum_{p \in \mathbb{Z}} \max(1, p^2) |c_p|^2 \le 2 \right\}.$$

³ The same result holds for the set \mathcal{L}' of 1-Lipschitz functions f with f(0) = f(1) = 0, since \mathcal{L} is dense in \mathcal{L}' .

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

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

We then apply Corollary 4.1.7 as follows. The set J consists of two copies of \mathbb{Z} . There is a two-to-one map φ from J to \mathbb{Z} and $b_j = \max(1, |\varphi(j)|)$. Then $\operatorname{card}\{j \in J; |b_j| \leq u\} \leq Lu$ for $u \geq 1$ and = 0 for u < 1. \Box

Exercise 4.1.9. (a) For $k \geq 1$ consider the space $T = \{0,1\}^{2^k}$. Writing $t = (t_i)_{i \leq 2^k}$ a point of T, consider on T the distance $d(t,t') = 2^{-j-1}$, where $j = \min\{i \leq 2^k; t_i \neq t'_i\}$. Consider the set \mathcal{L} of 1-Lipschitz functions on (T,d) which are zero at $t = (0, \ldots, 0)$. Prove that $\gamma_{1,2}(\mathcal{L}, d_{\infty}) \leq L\sqrt{k}$, where d_{∞} denotes the distance induced by the uniform norm. Hint: use Lemma 4.5.18 below to prove that $e_n(\mathcal{L}, d_{\infty}) \leq L2^{-n}$ and conclude using (4.9).

(b) Let μ denote the uniform probability on T and d_2 the distance induced by $L^2(\mu)$. It can be shown that $\gamma_{1,2}(\mathcal{L}, d_2) \geq \sqrt{k}/L$. (This could be challenging even if you master Chapter 3.) Meditate upon the difference with Proposition 4.1.8.

4.2 Partitioning Scheme, II

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

Theorem 4.2.1. Consider a metric space (T, d) and a number $r \ge 4$. Assume that for $j \in \mathbb{Z}$ we are given functions $s_j \ge 0$ on T with the following property:

Whenever we consider a subset A of T and $j \in \mathbb{Z}$ with $\Delta(A) \leq 2r^{-j}$ then for each $n \geq 1$ either $e_n(A) \leq r^{-j-1}$ or else there exists $t \in A$ with $s_j(t) \geq (2^{n/\alpha}r^{-j-1})^p$. (4.14)

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

$$\forall t \in T \; ; \; \sum_{n \ge 0} (2^{n/\alpha} \Delta(A_n(t)))^p \le K(\alpha, p, r) \Big(\Delta(T, d)^p + \sup_{t \in T} \sum_{j \in \mathbb{Z}} s_j(t) \Big) \; . \; (4.15)$$

The proof is identical to that of Theorem 2.9.8 which corresponds to the case $\alpha = 2$ and p = 1.

Proof of Theorem 4.1.4. We recall that by hypothesis T is the unit ball for the norm $\|\cdot\|$ of p-convex Banach space (but we study T for the metric dinduced by a different norm). For $t \in T$ and $j \in \mathbb{Z}$ we set

$$c_j(t) = \inf\{\|v\| \; ; \; v \in B_d(t, r^{-j}) \cap T\} \le 1 \; , \tag{4.16}$$

where the index d emphasizes that the ball is for the distance d rather than for the norm. Since T is the unit ball we have $c_i(t) \leq 1$. Let us set

$$D = \sup_{n \ge 0} 2^{n/\alpha} e_n(T, d) .$$
 (4.17)

The proof relies on Theorem 4.2.1 for the functions

$$s_j(t) = KD^p(c_{j+2}(t) - c_{j-1}(t)) , \qquad (4.18)$$

for a suitable value of K. Since $c_i(t) \leq 1$ it is clear that

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

and (using also that $\Delta(T, d) \leq 2e_0(T, d)$) the issue is to prove that (4.14) holds for a suitable constant K in (4.18). Consider then a set $A \subset T$ with $\Delta(A) \leq 2r^{-j}$, consider $n \geq 1$ and assume that $e_n(A) > a := r^{-j-1}$. The goal is to find $t \in A$ such that $s_j(t) \geq (2^{n/\alpha}r^{-j-1})^p$ i.e.

$$KD^{p}(c_{j+2}(t) - c_{j-1}(t)) \ge (2^{n/\alpha}r^{-j-1})^{p}$$
. (4.19)

For this let $m = N_n$, According to Lemma 2.9.3, (a) there exist points $(t_\ell)_{\ell \leq m}$ in A, such that $d(t_\ell, t_{\ell'}) \geq a$ whenever $\ell \neq \ell'$. We will show that one of the points t_ℓ satisfies (4.19). Consider $H_\ell = T \cap B_d(t_\ell, a/r) = T \cap B_d(t_\ell, r^{-j-2})$. By definition of $c_{j+1}(t_\ell)$ we have $c_{j+2}(t_\ell) = \inf\{\|v\| ; v \in H_\ell\}$. The basic idea is that the points of the different sets H_ℓ cannot be too close to each other for the norm of T because there are N_n such sets. So, since the norm is sufficiently convex, we will find a point in the convex hull of these sets with a norm quite smaller than $\max_{\ell \leq m} c_{j+2}(t_\ell)$. To implement the idea, consider u' such that

$$2 > u' > \max_{\ell \le m} \inf\{ \|v\| \; ; \; v \in H_{\ell} \} = \max_{\ell \le m} c_{j+2}(t_{\ell}) \; . \tag{4.20}$$

For $\ell \leq m$ consider $v_{\ell} \in H_{\ell}$ with $||v_{\ell}|| \leq u'$. It follows from (4.3) that for $\ell, \ell' \leq m$,

$$\left\|\frac{v_{\ell} + v_{\ell'}}{2u'}\right\| \le 1 - \eta \left\|\frac{v_{\ell} - v_{\ell'}}{u'}\right\|^p.$$
(4.21)

Set

$$u = \inf \left\{ \|v\| \; ; \; v \in \operatorname{conv} \bigcup_{\ell \le m} H_{\ell} \right\}.$$
(4.22)

Since $(v_{\ell} + v_{\ell'})/2 \in \operatorname{conv} \bigcup_{\ell \leq m} H_{\ell}$, by definition of u we have $u \leq ||v_{\ell} + v'_{\ell}||/2$, and (4.21) implies

$$\frac{u}{u'} \le 1 - \eta \left\| \frac{v_\ell - v_{\ell'}}{u'} \right\|^p,$$

so that, using that u' < 2 in the second inequality below,

$$||v_{\ell} - v_{\ell'}|| \le u' \left(\frac{u' - u}{\eta u'}\right)^{1/p} < R := 2 \left(\frac{u' - u}{\eta}\right)^{1/p},$$

and hence the points $w_{\ell} := R^{-1}(v_{\ell} - v_1)$ belong to the unit ball T. Now, since $H_{\ell} \subset B_d(t_{\ell}, a/r)$ we have $v_{\ell} \in B_d(t_{\ell}, a/r)$. Since $r \ge 4$, we have $d(v_{\ell}, v_{\ell'}) \ge a/2$ for $\ell \ne \ell'$, and, since the distance d arises from a norm, by homogeneity we have $d(w_{\ell}, w_{\ell'}) \ge R^{-1}a/2$ for $\ell \ne \ell'$. Then Lemma 2.9.3, (c) implies that $e_{n-1}(T, d) \ge R^{-1}a/4$, so that from (4.17) it holds that $2^{(n-1)/\alpha}R^{-1}a/4 \le D$, and recalling that $R = 2((u'-u)/\eta)^{1/p}$ we obtain

$$(2^{n/\alpha}r^{j-1})^p \le KD^p(u'-u) ,$$

where K depends on α only. Since this holds for any u' as in (4.20), there exists ℓ such that

$$(2^{n/\alpha}r^{j-1})^p \le KD^p(c_{j+2}(t_\ell) - u) .$$
(4.23)

Now, by construction, for $\ell' \leq m$ we have

$$H_{\ell'} \subset B_d(t_{\ell'}, a/r) = B_d(t_{\ell'}, r^{-j-2}) \subset B_d(t_{\ell}, r^{-j+1})$$

since $d(t_{\ell}, t_{\ell'}) \leq 2r^{-j}$ as $t_{\ell}, t_{\ell'} \in A$ and $\Delta(A) \leq 2r^{-j}$. Thus conv $\bigcup_{\ell' \leq m} H_{\ell'} \subset B_d(t_{\ell}, r^{-j+1}) \cap T$, and from (4.16) and (4.22) we have $u \geq c_{j-1}(t_{\ell})$ and we have proved (4.19).

Exercise 4.2.2. Write the previous proof using a certain functional with an appropriate growth condition.

The following generalization of Theorem 4.1.4 yields very precise results when applied to ellipsoids. It will not be used in the sequel, so we refer to [171] for a proof.

Theorem 4.2.3. Consider β , β' , p > 0 with

$$\frac{1}{\beta} = \frac{1}{\beta'} + \frac{1}{p} \,. \tag{4.24}$$

Then, under the conditions of Theorem 4.1.4 we have

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

Exercise 4.2.4. Use Theorem 4.2.3 to obtain a geometrical proof of (2.159). Hint: Choose $\alpha = 2, \beta = 1, \beta' = p = 2$ and use (2.166).

4.3 Matchings

The rest of this chapter is devoted to the following problem. Consider N r.v.s X_1, \ldots, X_N independently and uniformly distributed in the unit cube $[0, 1]^d$, where $d \geq 1$. Consider a typical realization of these points. How evenly distributed in $[0, 1]^d$ are the points X_1, \ldots, X_N ? To measure this, we will match the points $(X_i)_{i\leq N}$ with non-random "evenly distributed" points $(Y_i)_{i\leq N}$, that is, we will find a permutation π of $\{1, \ldots, N\}$ such that the points X_i and $Y_{\pi(i)}$ are "close". There are different ways to measure "closeness". For example one may wish that the sum of the distances $d(X_i, Y_{\pi(i)})$ be as small as possible (Section 4.5), that the maximum distance $d(X_i, Y_{\pi(i)})$ be as small as possible (Section 4.7), or one can use more complicated measures of "closeness" (Section 17.1).

The case d = 1 is by far the simplest. Assuming that the X_i are labeled in a way that $X_1 \leq X_2 \leq \ldots$ and similarly for the Y_i one has $\mathsf{E}\sup_{i\leq N} |X_i - Y_i| \leq L\sqrt{N}$. This is a consequence of the classical inequality (which we will later prove as an exercise)

$$\mathsf{E}\sup_{0 \le t \le 1} |\operatorname{card}\{i \le N \; ; \; X_i \le t\} - Nt| \le L\sqrt{N} \; . \tag{4.25}$$

The case where d = 2 is very special, and is the object of the present chapter. The case $d \geq 3$ will be studied in Chapter 18. The reader having never thought of the matter might think that the points X_1, \ldots, X_N are very evenly distributed. 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 $\sqrt{\log N}/\sqrt{N}$. This specific irregularity is mentioned just as an easy illustration, and plays no part in the considerations of the present chapter. What matters here⁴ is that in some sense there are irregularities at all scales 2^{-k} for $1 \le k \le L^{-1} \log N$, and that these are all of the same order. To see this, let us think that we actually move the points X_i to the points $Y_{\pi(i)}$ in straight lines. In a given small square of side 2^{-k} there is often an excess of points X_i of order $\sqrt{N2^{-2k}} = 2^{-k}\sqrt{N}$. When matched these points will *leave* the square and will cross its boundary. The number of points crossing this boundary per unit of length is independent of the scale 2^{-k} . It will also often happen that there is a deficit of points X_i in this square of side 2^{-k} , and in this case some points X_i will have to cross the boundary to *enter* it. The flows at really different scales should be roughly independent, and there are about $\log N$ such scales, so when we combine what happens at different scales we should get an extra factor $\sqrt{\log N}$ (and not $\log N$). Crossing our fingers, we should believe that about $\sqrt{N \log N}$ points X_i per unit of length cross a typical interval contained in the square, so that the total length of the segments joining the points X_i to the points $Y_{\pi(i)}$

⁴ This is much harder to visualize and is specific to the case d = 2.

should be of that order.⁵ This fact that all scales have the same weight is typical of dimension 2. In dimension 1, it is the large scales that matter most, while in dimension ≥ 3 it is the small ones.

Exercise 4.3.1. Perform this calculation.

One can summarize the situation by saying that

obstacles to matchings at different scales may combine in dimension 2 but not in dimension ≥ 3 . (4.26)

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

What does it mean to say that the non-random points $(Y_i)_{i \leq N}$ are evenly distributed? When N is a square, $N = n^2$, everybody will agree that the N points $(k/n, \ell/n), 1 \leq k, \ell \leq n$ are evenly distributed, and unless you love details you are welcomed to stick to this case. More generally we will say that the non-random points $(Y_i)_{i \leq N}$ are evenly spread if one can cover $[0, 1]^2$ with N rectangles with disjoint interiors, such that each rectangle R has an area 1/N, contains exactly one point Y_i , and is such that $R \subset B(Y_i, 10/\sqrt{N})$. To construct such points one may proceed as follows. Consider the largest integer k with $k^2 \leq N$, and observe that $k(k+3) \geq (k+1)^2 \geq N$, so that there exists integers $(n_i)_{i\leq k}$ with $k \leq n_i \leq k+3$ and $\sum_{i\leq k} n_i = N$. Cut the unit square into k vertical strips, in a way that the *i*-th strip has width n_i/N and to this *i*-th strip attribute n_i points placed at even intervals $1/n_i$.⁷

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

Proposition 4.3.2. Consider a matrix $C = (c_{ij})_{i,j \le N}$. Let

 $^{^{5}}$ As we will see later, we have guessed the correct result.

⁶ There is nothing magic about the number 10. Thinks of it as a universal constant. The last thing I want is to figure out the best possible value. That 10 works should be obvious from the following construction.

⁷ A more elegant approach dispenses from this slightly awkward construction. It is the concept of "transportation cost". One attributes mass 1/N to each point X_i , and one measures the "cost of transporting" the resulting probability measure to the uniform probability on $[0, 1]^2$. In the presentation one thus replaces the evenly spread points Y_i by a more canonical object, the uniform probability on $[0, 1]^2$. This approach does not make the proofs any easier, so we shall not use it despite its aesthetic appeal.

4.3 Matchings 115

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

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

$$M(C) = \sup \sum_{i \le N} (w_i + w'_i) , \qquad (4.27)$$

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

$$\forall i, j \le N, w_i + w'_j \le c_{ij} . \tag{4.28}$$

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" (4.27).

A well-known application of Proposition 4.3.2 is another "duality formula".

Proposition 4.3.3. Consider points $(X_i)_{i \leq N}$ and $(Y_i)_{i \leq N}$ in a metric space (T, d). Then

$$\inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) = \sup_{f \in \mathcal{C}} \sum_{i \le N} (f(X_i) - f(Y_i)) ,$$
(4.29)

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

Proof. Given any permutation π and any 1-Lipschitz function f we have

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

This proves the inequality \geq in (4.29). To prove the converse, we use (4.27) with $c_{ij} = d(X_i, Y_j)$, so that

$$\inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) = \sup \sum_{i \le N} (w_i + w'_i) , \qquad (4.30)$$

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

$$\forall i, j \leq N, w_i + w'_j \leq d(X_i, Y_j)$$
. (4.31)

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

$$f(x) = \min_{j \le N} (-w'_j + d(x, Y_j)) .$$
(4.32)

It is 1-Lipschitz, since it is the minimum of functions which are themselves 1-Lipschitz. By definition we have $f(Y_j) \leq -w'_j$ and by (4.31) for $i \leq N$ we have $w_i \leq f(X_i)$, so that

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

Exercise 4.3.4. Consider a function f which achieves the supremum in the right-hand side of (4.29). Prove that for an optimal matching we have $f(X_i) - f(Y_{\pi(i)}) = d(X_i, Y_{\pi(i)})$. If you know f, this basically tells you how to find the matching. To find $Y_{\pi(i)}$, move from X_i in the direction of steepest descent of f until you find a points Y_j .

The following is a well-known, and rather useful result of combinatorics. We deduce it from Proposition 4.3.2 in Section B.1, but other proofs exist, based on different ideas, see e.g. [10] § 2.

Corollary 4.3.5 (Hall's Marriage Lemma). Assume that to each $i \leq N$ we associate a subset A(i) of $\{1, \ldots, N\}$ and that, for each subset I of $\{1, \ldots, N\}$ we have

$$\operatorname{card}\left(\bigcup_{i\in I} A(i)\right) \ge \operatorname{card} I.$$
 (4.33)

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

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

4.4 Discrepancy Bounds

Generally speaking, the study of expressions of the type

$$\sup_{f \in \mathcal{F}} \left| \sum_{i \le N} (f(X_i) - \int f d\mu) \right|$$
(4.34)

for a class of functions \mathcal{F} will be important in the present book, particularly in Chapter 14. A bound on such a quantity is called a *discrepancy bound* because since

$$\left|\sum_{i\leq N} (f(X_i) - \int f d\mu)\right| = N \left|\frac{1}{N}\sum_{i\leq N} f(X_i) - \int f d\mu\right|$$

it bounds uniformly on \mathcal{F} the "discrepancy" between the true measure $\int f d\mu$ and the "empirical measure" $N^{-1} \sum_{i \leq N} f(X_i)$. Finding such a bound simply requires finding a bound for the supremum of the process $(|Z_f|)_{f \in \mathcal{F}}$, where the (centered) r.v.s Z_f are given by⁸

$$Z_f = \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\mu) , \qquad (4.35)$$

a topic at the very center of our attention.

A relation between discrepancy bounds and matching theorems can be guessed from Proposition 4.3.3 and will be made explicit in the next section. In this book *every* matching theorem will be proved through a discrepancy bound.

⁸ Please remember this notation which is used throughout this chapter.

4.5 The Ajtai-Komlós-Tusnády Matching Theorem

Theorem 4.5.1 ([5]). If the points $(Y_i)_{i \leq N}$ are evenly spread and the points $(X_i)_{i < N}$ are *i.i.d.* uniform on $[0, 1]^2$, then (for $N \geq 2$)

$$\mathsf{E} \inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) \le L\sqrt{N \log N} , \qquad (4.36)$$

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

The term \sqrt{N} is just a scaling effect. There are N terms $d(X_i, Y_{\pi(i)})$ each of which should be about $1/\sqrt{N}$. The non-trivial part of the theorem is the factor $\sqrt{\log N}$. In Section 4.6 we shall show that (4.36) can be reversed, i.e.

$$\mathsf{E} \ \inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) \ge \frac{1}{L} \sqrt{N \log N} \ . \tag{4.37}$$

In order to understand that the bound (4.36) is not trivial, you can study the following greedy matching algorithm which was shown to me by Yash Kanoria.

Exercise 4.5.2. For each $n \geq 0$ consider the partition \mathcal{H}_n of $[0,1]^2$ into 2^{2n} equal squares. Consider the largest integer n_0 with $2^{2n_0} \leq N$ and proceed as follows. For each small square in \mathcal{H}_{n_0} match as many as possible of the points X_i with points Y_i in the same square. Remove the points X_i and the points Y_i that you have matched this way. For the remaining points proceed as follows. In each small square of \mathcal{H}_{n_0-1} match as many of the remaining points X_i to remaining points Y_i inside the same square. Remove all the points X_i and the points X_i and the points Y_i that you have removed at this stage and continue in this manner. Prove that the expected cost of the matching thus constructed is $\leq L\sqrt{N} \log N.^9$

Let us state the "discrepancy bound" at the root of Theorem 4.5.1. Consider the class C of 1-Lipschitz functions on $[0,1]^2$, i.e. of functions f that satisfy

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

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

Theorem 4.5.3. We have

$$\mathsf{E}\sup_{f\in\mathcal{C}}\left|\sum_{i\leq N} (f(X_i) - \int f \mathrm{d}\lambda)\right| \leq L\sqrt{N\log N} .$$
(4.38)

⁹ It can be shown that this bound can be reversed.

Research problem 4.5.4. Prove that the following limit

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

exists.

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

Theorem 4.5.3 is obviously interesting in its own right, and proving it is the goal of this section. Before we discuss it, let us put matchings behind us. *Proof of Theorem 4.5.1.* We recall (4.29), i.e.

$$\inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) = \sup_{f \in \mathcal{C}} \sum_{i \le N} (f(X_i) - f(Y_i)) ,$$
(4.39)

and we simply write

$$\sum_{i\leq N} (f(X_i) - f(Y_i)) \leq \left| \sum_{i\leq N} (f(X_i) - \int f d\lambda) \right| + \left| \sum_{i\leq N} (f(Y_i) - \int f d\lambda) \right|.$$
(4.40)

Next, we claim that

$$\left|\sum_{i\leq N} (f(Y_i) - \int f \mathrm{d}\lambda)\right| \leq L\sqrt{N} .$$
(4.41)

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

$$\begin{split} \left|\sum_{i\leq N} (f(Y_i) - \int f d\lambda)\right| &= \left|\sum_{i\leq N} f(Y_i) - N \int f d\lambda\right| \\ &\leq \sum_{i\leq N} \left| (f(Y_i) - N \int_{R_i} f d\lambda) \right| \\ &\leq \sum_{i\leq N} N \left| \int_{R_i} (f(Y_i) - f(x)) d\lambda(x) \right| . \quad (4.42) \end{split}$$

Since f is Lipschitz and R_i is of diameter $\leq L/\sqrt{N}$ we have $|f(Y_i) - f(x)| \leq L/\sqrt{N}$ when $x \in R_i$. This proves the claim.

Now, using (4.39) and taking expectation

$$\mathsf{E} \inf_{\pi} \sum_{i \leq N} d(X_i, Y_{\pi(i)}) \leq L\sqrt{N} + \mathsf{E} \sup_{f \in \mathcal{C}} \left| \sum_{i \leq N} (f(X_i) - \int f d\lambda) \right|$$
$$\leq L\sqrt{N \log N}$$

by (4.38).

4.5.1 The Long and Instructive Way

S. Bobkov and M. Ledoux recently found [27] a magically simple proof of Theorem 4.5.3. We will present it in Subsection 4.5.2. This proof relies on very specific features, and it is unclear as to whether it will apply to other matching theorems. In the present section we write a far more pedestrian (but far more instructive) proof with the general result Theorem 6.8.3 in mind.

To prove Theorem 4.5.3 the overall strategy is clear. We think of the left-hand side as $\operatorname{Esup}_{f \in \mathcal{C}} |Z_f|$, where Z_f is the random variable of (4.35). We then find nice tail properties for these r.v.s, and we use the methods of Chapter 2. In the end (and because we are dealing with a deep fact) we shall have to prove some delicate "smallness" property of the class \mathcal{C} . This smallness property will ultimately be derived from the ellipsoid theorem. The (very beautiful) strategy for the hard part of the estimates relies on a kind of two-dimensional version of Proposition 4.1.8 and is outlined on page 121.

The class C of 1-Lipschitz function on the unit square is not small in any sense for the simple reason that it contains all the constant functions. However the expression $\sum_{i \leq N} (f(X_i) - \int f d\lambda)$ does not change if we replace f by f + a where a is a constant. In particular

$$\sup_{f \in \mathcal{C}} \left| \sum_{i \le N} (f(X_i) - \int f d\lambda) \right| = \sup_{f \in \widehat{\mathcal{C}}} \left| \sum_{i \le N} (f(X_i) - \int f d\lambda) \right|$$

where we define $\widehat{\mathcal{C}}$ as the set of 1-Lipschitz functions on the unit square for which f(1/2, 1/2) = 0.¹⁰ The gain is that we now may hope that $\widehat{\mathcal{C}}$ is small in the appropriate sense. To prove Theorem 4.5.3, we will prove the following.

Theorem 4.5.5. We have

$$\mathsf{E}\sup_{f\in\mathcal{C}}\left|\sum_{i\leq N} (f(X_i) - \int f \mathrm{d}\lambda)\right| \leq L\sqrt{N\log N} .$$
(4.43)

The following fundamental classical result will allow us to control the tails of the r.v. Z_f of (4.35). It will be used many times.

¹⁰ There is no real reason other than my own fancy to impose that the functions are zero right in the middle of the square.

Lemma 4.5.6 (Bernstein's inequality). Let $(W_i)_{i\geq 1}$ be independent r.v.s with $\mathsf{E}W_i = 0$ and consider a number a with $|W_i| \leq a$ for each i. Then, for v > 0,

$$\mathsf{P}\left(\left|\sum_{i\geq 1} W_i\right|\geq v\right)\leq 2\exp\left(-\min\left(\frac{v^2}{4\sum_{i\geq 1}\mathsf{E}W_i^2},\frac{v}{2a}\right)\right).$$
(4.44)

Proof. For $|x| \leq 1$, we have

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

and thus, since $\mathsf{E}W_i = 0$, for $a|\lambda| \leq 1$, we have

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

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

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

Now, for $0 \leq \lambda \leq 1/a$ we have

$$\begin{split} \mathsf{P}\Bigl(\sum_{i\geq 1} W_i \geq v\Bigr) &\leq \exp(-\lambda v) \mathsf{E} \exp \lambda \sum_{i\geq 1} W_i \\ &\leq \exp\Bigl(\lambda^2 \sum_{i\geq 1} \mathsf{E} W_i^2 - \lambda v\Bigr) \;. \end{split}$$

If $av \leq 2\sum_{i\geq 1} \mathsf{E}W_i^2$, we take $\lambda = v/(2\sum_{i\geq 1} \mathsf{E}W_i^2)$, obtaining a bound $\exp(-v^2/(4\sum_{i\geq 1} \mathsf{E}W_i^2))$. If $av > 2\sum_{i\geq 1} \mathsf{E}W_i^2$, we take $\lambda = 1/a$, and we note that

$$\frac{1}{a^2}\sum_{i\geq 1}\mathsf{E}W_i^2 - \frac{v}{a} \leq \frac{av}{2a^2} - \frac{v}{a} = -\frac{v}{2a}$$

so that $\mathsf{P}(\sum_{i\geq 1} W_i \geq v) \leq \exp(-\min(v^2/4\sum_{i\geq 1}\mathsf{E}W_i^2, v/2a))$. Changing W_i into $-W_i$ we obtain the same bound for $\mathsf{P}(\sum_{i\geq 1} W_i \leq -v)$. \Box

Corollary 4.5.7. For each v > 0 we have

$$\mathsf{P}(|Z_f| \ge v) \le 2\exp\left(-\min\left(\frac{v^2}{4N\|f\|_2^2}, \frac{v}{4\|f\|_{\infty}}\right)\right), \tag{4.45}$$

,

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

Proof. We use Bernstein's inequality with $W_i = f(X_i) - \int f d\lambda$ if $i \leq N$ and $W_i = 0$ if i > N. We then observe that $\mathsf{E}W_i^2 \leq \mathsf{E}f^2 = \|f\|_2^2$ and $|W_i| \leq 2 \sup |f| = 2\|f\|_{\infty}$. Let us then pretend for a while that in (4.45) the bound was instead $2 \exp(-v^2/(4N||f||_2^2))$. Thus we would be back to the problem we considered first, bounding the supremum of a stochastic process under the increment condition (2.4), where the distance on C is given by $d(f_1, f_2) = \sqrt{2N}||f_1 - f_2||_2$. The first thing to point out is that Theorem 4.5.3 is a prime example of a natural situation where using covering numbers does not yield the correct result, where we recall that for a metric space (T, d), the covering number $N(T, d, \epsilon)$ denotes the smallest number of balls of radius ϵ that are needed to cover T. This is closely related to the fact that, as explained in Section 2.13, covering numbers do not describe well the size of ellipsoids. It is hard to formulate a theorem to the effect that covering numbers do not suffice, but the root of the problem is described in the next exercise, and a more precise version can be found later in Exercise 4.5.20.

Exercise 4.5.8. Prove that for each $0 < \epsilon \le 1$

$$\log N(\widehat{\mathcal{C}}, d_2, \epsilon) \ge \frac{1}{L\epsilon^2} , \qquad (4.46)$$

where d_2 denotes the distance in $L^2([0, 1]^2)$. Hint: Consider an integer $n \ge 0$, and divide $[0, 1]^2$ into 2^{2n} equal squares of area 2^{-2n} . For every such square Cconsider a number $\epsilon_C = \pm 1$. Consider then the function $f \in C$ such that for $x \in C$ one has $f(x) = \epsilon_C d(x, B)$, where B denotes the boundary of C. There are $2^{2^{2n}}$ such functions. Prove that by appropriate choices of the signs ϵ_C one may find at least $\exp(2^{2n}/L)$ functions of this type which are at mutual distance $\ge 2^{-n}/L$.

Since covering numbers do not suffice, we will appeal to the generic chaining, Theorem 2.7.2. As we will show later, in Exercise 4.5.21 we have $\gamma_2(\widehat{\mathcal{C}}, d_2) = \infty$. To overcome this issue we will replace $\widehat{\mathcal{C}}$ by a sufficiently large finite subset $\mathcal{F} \subset \widehat{\mathcal{C}}$, for which we shall need the crucial estimate $\gamma_2(\mathcal{F}, d_2) \leq L\sqrt{\log N}$. This will be done by proving that $\gamma_{2,2}(\widehat{\mathcal{C}}, d_2) < \infty$ where $\gamma_{2,2}$ is the functional of (4.5), so that $\gamma_{2,2}(\mathcal{F}, d_2) < \infty$, and appealing to Lemma 4.1.3.

The main ingredient towards the control of $\gamma_{2,2}(\widehat{\mathcal{C}}, d_2)$ is the following 2-dimensional version of Proposition 4.1.8.

Lemma 4.5.9. Consider the space C^* of 1-Lipschitz functions on $[0,1]^2$ which are zero on the boundary of $[0,1]^2$. Then $\gamma_{2,2}(C^*, d_2) < \infty$.

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

$$c_{p,q}(f) = \iint_{[0,1]^2} f(x_1, x_2) \exp(2i\pi(px_1 + qx_2)) dx_1 dx_2 .$$
(4.47)

The Plancherel formula

$$||f||_2 = \left(\sum_{p,q \in \mathbb{Z}} |c_{p,q}(f)|^2\right)^{1/2}$$
(4.48)

asserts that Fourier transform is an isometry, so that if

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

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 (4.47) and integration by parts we get

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

Using (4.48) for $\partial f/\partial x$, and since $\|\partial f/\partial x\|_2 \leq 1$ we get $\sum_{p,q\in\mathbb{Z}} p^2 |c_{p,q}(f)|^2 \leq 1/4\pi^2$. Proceeding similarly for $\partial f/\partial y$, we get

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

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

$$\mathcal{E} = \left\{ \left((x_{p,q}), (y_{p,q}) \right) \in \ell^2(\mathbb{Z} \times \mathbb{Z}) \times \ell^2(\mathbb{Z} \times \mathbb{Z}) ; \\ x_{0,0}^2 + y_{0,0}^2 \leq 1 , \sum_{p,q \in \mathbb{Z}} (p^2 + q^2)(x_{p,q}^2 + y_{p,q}^2) \leq 1 \right\}.$$
(4.49)

For $u \geq 1$, we have

$$\operatorname{card}\left\{(p,q) \in \mathbb{Z} \times \mathbb{Z} ; \ p^2 + q^2 \le u^2\right\} \le (2u+1)^2 \le Lu^2$$

We then deduce from Corollary 4.1.7 that $\gamma_{2,2}(\mathcal{E}, d) < \infty$.

Proposition 4.5.10. We have $\gamma_{2,2}(\widehat{\mathcal{C}}, d_2) < \infty$.

I am grateful to R. van Handel who showed me the following simple arguments, which replaces pages of gritty work in [171]. The basic idea is to deduce this from Lemma 4.5.9, essentially by showing that $\hat{\mathcal{C}}$ is a Lipschitz image of a subset of \mathcal{C}^* , or more exactly of the clone considered in the next lemma.

Lemma 4.5.11. The set C^{\sharp} of 1-Lipschitz functions on $[-1,2]^2$ which are zero on the boundary of this set satisfies $\gamma_{2,2}(C^{\sharp}, d^{\sharp}) < \infty$ where d^{\sharp} is the distance induced by $L^2([-1,2]^2, d\lambda)$.

Proof. This should be obvious form Lemma 4.5.9, we just perform the same construction on two squares of different sizes, $[0,1]^2$ and $[-1,2]^2$.

Lemma 4.5.12. Each Lipschitz function $f \in \widehat{\mathcal{C}}$ is the restriction to $[0,1]^2$ of a function f^{\sharp} of \mathcal{C}^{\sharp} .

Proof. A function $f \in \widehat{\mathcal{C}}$ may be extended to a 1-Lipschitz function \tilde{f} on \mathbb{R}^2 by the formula $\tilde{f}(y) = \inf_{x \in [0,1]^2} f(x) + d(x,y)$. Since f(1/2, 1/2) = 0 by definition of $\widehat{\mathcal{C}}$ and since f is 1-Lipschitz, then $|f(x)| \leq 1/\sqrt{2} \leq 1$ for $x \in [0,1]^2$. The function $f^{\sharp}(y) = \min(\tilde{f}(y), d(y, \mathbb{R}^2 \setminus [-1,2]^2))$ is 1-Lipschitz. Since each point of $[0,1]^2$ is at distance ≥ 1 of $\mathbb{R}^2 \setminus [-1,2]^2$, f^{\sharp} coincides with f on $[0,1]^2$, and it is zero on the boundary of $[-1,2]^2$.

Proof of Proposition 4.5.10. To each function f of \mathcal{C}^{\sharp} we associate its restriction $\varphi(f)$ to $[0,1]^2$. Since the map φ is a contraction, by Lemma 4.5.11 we have $\gamma_{2,2}(\varphi(\mathcal{C}^{\sharp})) < \infty$, and by Lemma 4.5.12 we have $\widehat{\mathcal{C}} \subset \varphi(\mathcal{C}^{\sharp})$.

Let us now come back to Earth and deal with the actual bound (4.45). For this we develop an appropriate version of Theorem 2.7.2. It will be used many times. The ease with which one deals with two distances is remarkable. The proof of the theorem contain a principle which will be used many times: if we have two admissible sequences of partitions such that for each of them, the sets of the partition as small in a certain sense, then we can construct an admissible sequence of partitions whose sets are small in both senses.

Theorem 4.5.13. Consider a set T provided with two distances d_1 and d_2 . Consider a centered process $(X_t)_{t \in T}$ which satisfies

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

$$\mathsf{P}(|X_s - X_t| \ge u) \le 2 \exp\left(-\min\left(\frac{u^2}{d_2(s,t)^2}, \frac{u}{d_1(s,t)}\right)\right). \tag{4.50}$$

Then

$$\mathsf{E}\sup_{s,t\in T} |X_s - X_t| \le L(\gamma_1(T,d_1) + \gamma_2(T,d_2)) . \tag{4.51}$$

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

Proof. We denote by $\Delta_j(A)$ the diameter of the set A for d_j . We consider an admissible sequence $(\mathcal{B}_n)_{n>0}$ such that¹¹

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

$$(4.52)$$

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

$$\forall t \in T , \sum_{n \ge 0} 2^{n/2} \Delta_2(C_n(t)) \le 2\gamma_2(T, d_2) .$$
(4.53)

 $^{^{11}}$ The factor 2 in the right-hand side below is just in case the infimum over all partitions is not attained.

Here $B_n(t)$ is the unique element of \mathcal{B}_n that contains t (etc.). We define partitions \mathcal{A}_n of T as follows. We set $\mathcal{A}_0 = \{T\}$, and, for $n \ge 1$, we define \mathcal{A}_n as the partition generated by \mathcal{B}_{n-1} and \mathcal{C}_{n-1} , i.e. the partition that consists of the sets $B \cap C$ for $B \in \mathcal{B}_{n-1}$ and $C \in \mathcal{C}_{n-1}$. Thus card $\mathcal{A}_n \le N_{n-1}^2 \le N_n$, and the sequence (\mathcal{A}_n) is admissible. We then choose for each $n \ge 0$ a set T_n such that card $T_n \le N_n$ which meets all the sets in \mathcal{A}_n . It is convenient to reformulate (4.50) as follows: when $u \ge 1$ we have

$$\forall s, t \in T$$
, $\mathsf{P}(|X_s - X_t| \ge u^2 d_1(s, t) + u d_2(s, t)) \le 2 \exp(-u^2)$.

We then copy the proof of (2.34), replacing (2.31) by

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

Exercise 4.5.14. The purpose of this exercise is to deduce Theorem 4.5.13 from Theorem 2.7.14.

(a) Prove that if for some numbers A, B > 0 a r.v. $Y \ge 0$ satisfies

$$\mathsf{P}(Y \ge u) \le 2 \exp\left(-\min\left(\frac{u^2}{A^2}, \frac{u}{B}\right)\right),$$

then for $p \ge 1$ we have $||Y||_p \le L(A\sqrt{p} + Bp)$.

(b) We denote by $D_n(A)$ the diameter of a subset A of T for the distance $\delta_n(s,t) = ||X_s - X_t||_{2^n}$. Prove that under the conditions of Theorem 4.5.13 there exists an admissible sequence of partitions (\mathcal{A}_n) such that

$$\sup_{t \in T} \sum_{n \ge 0} D_n(A_n(t)) \le L(\gamma_1(T, d_1) + \gamma_2(T, d_2)) .$$
(4.54)

Exercise 4.5.15. Consider a space T equipped with two different distances d_1 and d_2 . Prove that

$$\gamma_2(T, d_1 + d_2) \le L(\gamma_2(T, d_1) + \gamma_2(T, d_2)).$$
(4.55)

We can now state a general bound, from which we will deduce Theorem 4.5.3.

Theorem 4.5.16. Consider a class \mathcal{F} of functions on $[0,1]^2$ and assume that $0 \in \mathcal{F}$. Then

$$\mathsf{E}\sup_{f\in\mathcal{F}}\left|\sum_{i\leq N} (f(X_i) - \int f \mathrm{d}\lambda)\right| \leq L\left(\sqrt{N}\gamma_2(\mathcal{F}, d_2) + \gamma_1(\mathcal{F}, d_\infty)\right), \quad (4.56)$$

where d_2 and d_{∞} are the distances induced on \mathcal{F} by the norms of L^2 and L^{∞} respectively.

Proof. Combining Corollary 4.5.7 with Theorem 4.5.13 we get, since $0 \in \mathcal{F}$ and $Z_0 = 0$,

$$\mathsf{E}\sup_{f\in\mathcal{F}}|Z_f| \le \mathsf{E}\sup_{f,f'\in\mathcal{F}}|Z_f - Z_{f'}| \le L\big(\gamma_2(\mathcal{F}, 2\sqrt{N}d_2) + \gamma_1(\mathcal{F}, 4d_\infty)\big).$$
(4.57)

Finally,
$$\gamma_2(\mathcal{F}, 2\sqrt{N}d_2) = 2\sqrt{N}\gamma_2(\mathcal{F}, d_2)$$
 and $\gamma_1(\mathcal{F}, 4d_\infty) = 4\gamma_1(\mathcal{F}, d_\infty)$. \Box

Exercise 4.5.17. Try to prove (4.25) now. Hint: Consider $\mathcal{F} = \{\mathbf{1}_{[0,k/N]}; k \leq N\}$. Use Exercise 2.7.5 and entropy numbers.

In the situation which interests us, there will plenty of room to control the term $\gamma_1(\mathcal{F}, d_\infty)$, and this term is a lower-order term, which can be considered as a simple nuisance. For this term, entropy numbers suffice. To control these, we first state a general principle, which was already known to Kolmogorov.

Lemma 4.5.18. Consider a metric space (U, d) and assume that for certain numbers B and $\alpha \geq 1$ and each $0 < \epsilon < B$ we have

$$N(U, d, \epsilon) \le \left(\frac{B}{\epsilon}\right)^{\alpha}.$$
(4.58)

Consider the set \mathcal{B} of 1-Lipschitz functions f on U with $||f||_{\infty} \leq B$. Then for each $\epsilon > 0$ we have

$$\log N(\mathcal{B}, d_{\infty}, \epsilon) \le K(\alpha) \left(\frac{B}{\epsilon}\right)^{\alpha}, \qquad (4.59)$$

where $K(\alpha)$ depends only on α . In particular,

$$e_n(\mathcal{B}, d_\infty) \le K(\alpha) B 2^{-n/\alpha} . \tag{4.60}$$

Proof. By homogeneity we may and do assume that B = 1. Using (4.58) for $\epsilon = 2^{-n}$, for each $n \ge 0$ consider a set $V_n \subset U$ with card $V_n \le 2^{n\alpha}$ such that any point of U is within distance 2^{-n} of a point of V_n . We define on \mathcal{B} the distance d_n by $d_n(f,g) = \max_{x \in V_n} |f(x) - g(x)|$. We prove first that

$$d_{\infty}(f,g) \le 2^{-n+1} + d_n(f,g) . \tag{4.61}$$

Indeed, for any $x \in U$ we can find $y \in V_n$ with $d(x,y) \le 2^{-n}$ and then $|f(x) - g(x)| \le 2^{-n+1} + |f(y) - g(y)| \le 2^{-n+1} + d_n(f,g).$

Denote by $W_n(f,r)$ the ball for d_n of center f and radius r. We claim that

$$W_{n-1}(f, 2^{-n+1}) \subset W_n(f, 2^{-n+3})$$
 (4.62)

Indeed, using (4.61) for n-1 rather than n we see that $d_n(f,g) \le d_{\infty}(f,g) \le 2^{-n+2} + 2^{-n+1} \le 2^{-n+3}$ for $g \in W_{n-1}(f, 2^{-n+1})$.

Next, we claim that

$$N(W_n(f, 2^{-n+3}), d_n, 2^{-n}) \le L^{\operatorname{card} V_n} .$$
(4.63)

Since $d_n(f,g) = \|\varphi_n(f) - \varphi_n(g)\|_{\infty}$ where $\varphi_n(f) = (f(x))_{x \in V_n}$ we are actually working here in $\mathbb{R}^{\operatorname{card} V_n}$, and (4.63) is a consequence of (2.47): in $\mathbb{R}^{\operatorname{card} V_n}$ we are covering a ball of radius 2^{-n+3} by balls of radius 2^{-n} .

Covering \mathcal{B} by $N(\mathcal{B}, d_{n-1}, 2^{-n+1})$ balls $W_{n-1}(f, 2^{-n+1})$ and hence by $N(\mathcal{B}, d_{n-1}, 2^{-n+1})$ balls $W_n(f, 2^{-n+3})$ and then covering each of these by $N(W_n(f, 2^{-n+3}), d_n, 2^{-n}) \leq L^{\operatorname{card} V_n}$ balls for d_n of radius 2^{-n} we obtain

$$N(\mathcal{B}, d_n, 2^{-n}) \le L^{\operatorname{card} V_n} N(\mathcal{B}, d_{n-1}, 2^{-n+1}) .$$
(4.64)

Since card $V_n = 2^{\alpha n}$, iteration of (4.64) proves that $\log N(\mathcal{B}, d_n, 2^{-n}) \leq K2^{\alpha n}$. Finally (4.61) implies that

$$\log N(\mathcal{B}, d_{\infty}, 2^{-n+2}) \le \log N(\mathcal{B}, d_{n-1}, 2^{-n-1}) \le K 2^{\alpha n}$$

and concludes the proof.

We apply the previous lemma to $U = [0, 1]^2$ which obviously satisfies (4.58) for $\alpha = 2$, so that (4.60) implies that for $n \ge 0$,

$$e_n(\hat{\mathcal{C}}, d_\infty) \le L2^{-n/2} . \tag{4.65}$$

Proposition 4.5.19. We have

$$\mathsf{E}\sup_{f\in\widehat{\mathcal{C}}}\left|\sum_{i\leq N}f(X_i) - \int f\mathrm{d}\lambda\right| \leq L\sqrt{N\log N} .$$
(4.66)

An interesting feature of this proof is that it does not work to try to use (4.56) directly. Rather we will use (4.56) for an appropriate subset T of \hat{C} , which can be thought of as the "main part" of \hat{C} , and for the "rest" of \hat{C} we will use other (and much cruder) bounds. This method is not artificial. As we will learn much later, in Theorem 6.8.3, when properly used, it always yields the best possible estimates.

Proof. Consider the largest integer m with $2^{-m} \ge 1/N$. By (4.65) we may find a subset T of $\widehat{\mathcal{C}}$ with card $T \le N_m$ and

$$\forall f \in \widehat{\mathcal{C}}, d_{\infty}(f,T) \leq L2^{-m/2} \leq L/\sqrt{N}$$
.

Thus for each $f \in \widehat{\mathcal{C}}$ consider $\tilde{f} \in T$ with $d_{\infty}(f, \tilde{f}) \leq L/\sqrt{N}$. Then

$$|Z_f| \le |Z_{\tilde{f}}| + |Z_f - Z_{\tilde{f}}| = |Z_{\tilde{f}}| + |Z_{f-\tilde{f}}| \le |Z_{\tilde{f}}| + L\sqrt{N}$$

where we have used the obvious inequality $|Z_{f-\tilde{f}}| \leq 2d_\infty(f,\tilde{f}).$ Since $\tilde{f} \in T$ we obtain

$$\mathsf{E}\sup_{f\in\widehat{\mathcal{C}}}|Z_f| \le \mathsf{E}\sup_{f\in T}|Z_f| + L\sqrt{N} . \tag{4.67}$$

To prove (4.66) it suffices to show that

$$\mathsf{E}\sup_{f\in T} |Z_f| \le L\sqrt{N\log N} . \tag{4.68}$$

Proposition 4.5.10 and Lemma 4.1.3 imply $\gamma_2(T, d_2) \leq L\sqrt{m} \leq L\sqrt{\log N}$. Now, as in (2.56) we have

$$\gamma_1(T, d_\infty) \le L \sum_{n \ge 0} 2^n e_n(T, d_\infty)$$
.

Since $e_n(T, d_{\infty}) = 0$ for $n \ge m$, (4.65) yields $\gamma_1(T, d_{\infty}) \le L2^{m/2} \le L\sqrt{N}$. Thus (4.68) follows from Theorem 4.5.16 and this completes the proof. \Box

Exercise 4.5.20. Use Exercise 4.5.8 to prove that Dudley's bound cannot yield better than the estimate $\gamma_2(T, d_2) \leq L \log N$.

Exercise 4.5.21. Assuming $\gamma_2(\mathcal{C}^*, d_2) < \infty$, show that the previous arguments prove that

$$\mathsf{E}\sup_{f\in\mathcal{C}^*}\left|\sum_{i\leq N}f(X_i)-\int f\mathrm{d}\lambda\right|\leq L\sqrt{N}(1+\gamma_2(\mathcal{C}^*,d_2))\ .$$

Comparing with (4.78) conclude that $\gamma_2(\mathcal{C}^*, d_2) = \infty$. Convince yourself that the separated trees implicitly constructed in the proof of (4.78) also witness this.

Exercise 4.5.22. Suppose now that you are in dimension d = 3. Prove that $\mathsf{E}\sup_{f\in\widehat{\mathcal{C}}}|\sum_{i\leq N} f(X_i) - \int f d\lambda| \leq LN^{2/3}$. Hint: according to Lemma 4.5.18 we have $e_n(\widehat{\mathcal{C}}, d_\infty) \leq L2^{-n/3}$. This is the only estimate you need, using the trivial fact that $e_n(\widehat{\mathcal{C}}, d_2) \leq e_n(\widehat{\mathcal{C}}, d_\infty)$.

Exercise 4.5.23. Consider the space $T = \{0, 1\}^{\mathbb{N}}$ provided with the distance $d(t, t') = 2^{-j/2}$, where $j = \min\{i \ge 1; t_i \ne t'_i\}$ for $t = (t_i)_{i\ge 1}$. This space somewhat resembles the unit square, in the sense that $N(T, d, \epsilon) \le L\epsilon^{-2}$ for $\epsilon \le 1$. Prove that if $(X_i)_{i\le N}$ are i.i.d. uniformly distributed in T and $(Y_i)_{i\le N}$ are uniformly spread (in a manner which is left to the reader to define precisely) then

$$\mathsf{E}\inf_{\pi} \sum_{i \le N} d(X_i, Y_{\pi(i)}) \le L\sqrt{N} \log N , \qquad (4.69)$$

where the infimum is over all the permutations of $\{1, \ldots, N\}$. Hint: You can do this from scratch, and for this covering numbers suffice, e.g. in the form of (4.59). The method of Exercise 4.5.2 also works here. In Exercise 4.6.8 below, you will be asked to prove that this bound *is of the correct order*.

4.5.2 The Short and Magic Way

We now start studying the Bobkov-Ledoux approach [27] which considerably simplifies previous results such as the following one.

Theorem 4.5.24. [[149]] Consider the class C^* of 1-Lipschitz functions on $[0,1]^2$ which are zero on the boundary of $[0,1]^2$. Consider points $(z_i)_{i\leq N}$ in $[0,1]^2$ and standard independent Gaussian r.v.s g_i . Then

$$\mathsf{E}\sup_{f\in\mathcal{C}^*}\left|\sum_{i\leq N}g_if(z_i)\right|^2 \leq LN\log N .$$
(4.70)

It should be obvious from Lemma 4.5.12 that in the previous result one may replace C^* by \widehat{C} of Theorem 4.5.5. The following improves on Theorem 4.5.3.

Corollary 4.5.25. Consider an independent sequence $(X_i)_{i \leq N}$ of r.v.s valued in $[0,1]^2$. (It is not assumed that these r.v.s have the same distribution.) Then

$$\mathsf{E}\sup_{f\in\widehat{\mathcal{C}}}\left|\sum_{i\leq N} (f(X_i) - \mathsf{E}f(X_i))\right| \leq L\sqrt{N\log N} .$$
(4.71)

Proof. Consider i.i.d standard Gaussian r.v.s g_i . Taking first expectation in the g_i given the X_i , it follows from Theorem 4.5.24 (or more accurately from the version of this theorem for the class \widehat{C}) that $\operatorname{E}\sup_{f\in\widehat{C}}|\sum_{i\leq N}g_if(X_i)|^2 \leq LN\log N$. The Cauchy-Schwarz inequality yields $\operatorname{E}\sup_{f\in\widehat{C}}|\sum_{i\leq N}g_if(X_i)| \leq L\sqrt{N\log N}$. We will learn later the simple tools which allow to deduce (4.71) from this inequality, in particular the Giné-Zinn inequalities and specifically (11.35) (which has to be combined with (6.6)).

Let us consider an integer $n \ge \sqrt{N}$ and the set

$$G = \{ (k/n, \ell/n) ; 0 \le k, \ell \le n-1 \} .$$

Using the fact that the functions $f \in C^*$ are 1-Lipschitz, and replacing each point z_i by the closest point in G (which is at distance $\leq \sqrt{2}/n \leq L/\sqrt{N}$ of z), the following is obvious.

Lemma 4.5.26. To prove Theorem 4.5.24 we may assume that each $z_i \in G$.

Let us define an ellipsoid \mathcal{E} in \mathbb{R}^N as a set $\mathcal{E} = \{\sum_{k\geq 1} \alpha_k u_k\}$ where $(u_k)_{k\geq 1}$ is a given sequence in \mathbb{R}^N and where $(\alpha_k)_{k\geq 1}$ varies over all the possible sequences with $\sum_{k\geq 1} \alpha_k^2 \leq 1$.¹² For $t \in \mathbb{R}^N$ we write $X_t = \sum_{i\leq N} t_i g_i$ as usual.

 $^{^{12}}$ The name is justified, a bit a of algebra allows one to show that such a set is an ellipsoid in the usual sense, but we do not need that.

Lemma 4.5.27. We have

$$\mathsf{E}\sup_{t\in\mathcal{E}}|X_t|^2 \leq \sum_{k\geq 1} \|u_k\|^2 \; .$$

Proof. This is exactly the same argument as to prove (2.157). For $t = \sum_{k\geq 1} \alpha_k u_k \in \mathcal{E}$ we have $X_t = \sum_{k\geq 1} \alpha_k X_{u_k}$ so that by the Cauchy-Schwarz inequality we have $|X_t|^2 \leq \sum_{k\geq 1} |X_{u_k}|^2$ and the result follows from taking the supremum in t and expectation since $\mathsf{E}|X_{u_k}|^2 = ||u_k||^2$. \Box

To prove Theorem 4.5.24 we will show that the set $\{(f(z_i))_{i \leq N}; f \in \mathcal{C}^*\}$ is a subset of an appropriate ellipsoid.

For this we identify G with the group $\mathbb{Z}_n \times \mathbb{Z}_n$ where $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$, with the idea to use Fourier analysis in G, keeping in mind that a function on $[0,1]^2$ which is zero on the boundary of this set will give rise to a function on $\mathbb{Z}_n \times \mathbb{Z}_n$. Consider the elements $\tau_1 = (1,0)$ and $\tau_2 = (0,1)$ of G. For a function $f: G \to \mathbb{R}$ we define the functions $f_1(\tau) = f(\tau + \tau_1) - f(\tau)$ and $f_2(\tau) = f(\tau + \tau_2) - f(\tau)$, and the class $\tilde{\mathcal{C}}$ of functions $G \to \mathbb{R}$ which satisfy

$$\forall \tau \in G, |f(\tau)| \le 1 \; ; \; \forall \tau \in G \; ; \; |f_1(\tau)| \le 1/n \; ; \; |f_2(\tau)| \le 1/n \; .$$
 (4.72)

Thus, seeing the functions on C^* as functions on G, they belong to \tilde{C} . Let us denote by \hat{G} the set of characters χ on G.¹³ The Fourier transform \hat{f} of a function f on G is the function \hat{f} on \hat{G} given by $\hat{f}(\chi) = (\operatorname{card} G)^{-1} \sum_{\tau \in G} f(\tau) \bar{\chi}(\tau)$ where we recall that $|\chi(\tau)| = 1$. One then has the Fourier expansion

$$f = \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\chi , \qquad (4.73)$$

and the Plancherel formula

$$\frac{1}{\operatorname{card} G} \sum_{\tau \in G} |f(\tau)|^2 = \sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^2 .$$
 (4.74)

The key to the argument is the following.

Proposition 4.5.28. There exist positive numbers $(c(\chi))_{\chi \in \widehat{G}}$ such that

$$\sum_{\chi \in \widehat{G}} \frac{1}{c(\chi)} \le L \log n \tag{4.75}$$

and

$$\forall f \in \tilde{\mathcal{C}} , \sum_{\chi \in \widehat{G}} c(\chi) |\widehat{f}(\chi)|^2 \le 1 .$$
(4.76)

 13 A character of a group G is a group homomorphism from G to the unit circle.

We start the preparations for the proof of Proposition 4.5.28. The following lemma performs integration by parts.

Lemma 4.5.29. For each function f on G and every $\chi \in \widehat{G}$ we have $\widehat{f}_1(\chi) = (\chi(-\tau_1) - 1)\widehat{f}(\chi)$.

Proof. Since $\sum_{\tau \in G} f(\tau + \tau_1)\chi(\tau) = \sum_{\tau \in G} f(\tau)\chi(\tau - \tau_1)$ by change of variable, we have

$$(\operatorname{card} G)\hat{f}_{1}(\chi) = \sum_{\tau \in G} (f(\tau + \tau_{1}) - f(\tau))\chi(\tau) = \sum_{\tau \in G} f(\tau)(\chi(\tau - \tau_{1}) - \chi(\tau))$$
$$= (\chi(\tau_{1}) - 1)\sum_{\tau \in G} f(\tau)\chi(\tau) = (\chi(\tau_{1}) - 1)(\operatorname{card} G)\hat{f}(\chi) ,$$

where we have used in the third equality that $\chi(\tau - \tau_1) = \chi(\tau)\chi(-\tau_1)$. \Box

Corollary 4.5.30. For $f \in \tilde{\mathcal{C}}$ we have

$$\sum_{\chi \in \widehat{G}} (|\chi(-\tau_1) - 1|^2 + |\chi(-\tau_2) - 1|^2) |\widehat{f}(\chi)|^2 \le \frac{2}{n^2} .$$
(4.77)

Proof. Using Lemma 4.5.29 and then the Plancherel formula (4.74) and (4.72) we obtain

$$\sum_{\chi \in \widehat{G}} |\chi(-\tau_1) - 1|^2 |\widehat{f}(\chi)|^2 = \sum_{\chi \in \widehat{G}} |\widehat{f}_1(\chi)|^2 = \frac{1}{\operatorname{card} G} \sum_{\tau \in G} |f_1(\tau)|^2 \le \frac{1}{n^2} ,$$

and we proceed similarly for τ_2 .

Proof of Proposition 4.5.28. For $\chi \in \widehat{G}$ let us set $c(\chi) = 1/2$ if χ is constant character χ_0 equal to 1, and otherwise

$$c(\chi) = \frac{n^2}{4} \left(|\chi(-\tau_1) - 1|^2 + |\chi(-\tau_2) - 1|^2 \right) \,.$$

Then, since $|\hat{f}(\chi_0)| \leq 1$ because $|f(\tau)| \leq 1$ for each τ , and using (4.77) in the second inequality,

$$\sum_{\chi \in \hat{G}} c(\chi) |\hat{f}(\chi)|^2 = \frac{1}{2} |\hat{f}(\chi_0)|^2 + \sum_{\chi \in \hat{G}, \chi \neq \chi_0} c(\chi) |\hat{f}(\chi)|^2 \le \frac{1}{2} + \frac{n^2}{4} \frac{2}{n^2} \le 1 ,$$

and this proves (4.76). To prove (4.75) we use that \widehat{G} is exactly the set of characters of the type $\chi_{p,q}(a,b) = \exp(2i\pi(ap+bq)/n)$ where $0 \le p,q \le n-1$. Thus $\chi_{p,q}(-\tau_1) = \exp(-2i\pi p/n)$ and $\chi_{p,q}(-\tau_2) = \exp(-2i\pi q/n)$. Now, for $0 \le x \le 1$ we have $|1 - \exp(-2i\pi x)| \ge \min(x, 1-x)$ so that 4.6 Lower Bound for the Ajtai-Komlós-Tusnády Theorem 131

$$|\chi_{p,q}(-\tau_1) - 1| \ge \frac{1}{Ln} \min(p, n-p) ; |\chi_{p,q}(-\tau_2) - 1| \ge \frac{1}{Ln} \min(q, n-q)$$

Thus

$$\sum_{\chi \in \widehat{G}} \frac{1}{c(\chi)} \le \frac{1}{c(\chi_0)} + L \sum \frac{1}{\min(p, n-p)^2 + \min(q, n-q)^2}$$

where the sum is over $0 \le p, q \le n-1$ and $(p,q) \ne (0,0)$. Distinguishing whether $p \le n/2$ or $p \ge n/2$ (and similarly for q) we obtain

$$\sum \frac{1}{\min(p, n-p)^2 + \min(q, n-q)^2} \le 4 \sum \frac{1}{p^2 + q^2} ,$$

where the sum is over the same set and this sum is $\leq L \log n$.

Proof of Theorem 4.5.24. We write (4.73) as $f = \sum_{\chi \in \widehat{G}} \alpha_{\chi} \chi / \sqrt{c(\chi)}$ where $\alpha_{\chi} = \sqrt{c(\chi)} \widehat{f}(\chi)$ so that $\sum_{\chi \in \widehat{G}} |\alpha_{\chi}|^2 \leq 1$ by (4.76). Now we come back to real numbers by taking the real part of the identity $f = \sum_{\chi \in \widehat{G}} \alpha_{\chi} \chi / \sqrt{c(\chi)}$. This gives an equality of the type $f = \sum_{\chi \in \widehat{G}} (\alpha'_{\chi} \chi' + \beta'_{\chi} \chi'') / \sqrt{c(\chi)}$ where $\sum_{\chi \in \widehat{G}} ((\alpha'_{\chi})^2 + (\beta'_{\chi})^2) \leq 1$ and $|\chi'|, |\chi''| \leq 1$. That is, the set $\{(f(z_i))_{i \leq N}; f \in \mathcal{C}^*\}$ is a subset of the ellipsoid $\mathcal{E} = \{\sum_{k \geq 1} \alpha_k u_k; \sum_k \alpha_k^2 \leq 1\}$, where the family (u_k) of points of \mathbb{R}^N consists of the points $(\chi'(z_i)/\sqrt{c(\chi)})_{i \leq N}$ and $(\chi''(z_i)/\sqrt{c(\chi)})_{i \leq N}$ where χ takes all possible values in \widehat{G} . For such a u_k we have $|u_k(z_i)| \leq 1/\sqrt{c(\chi)}$ so that $||u_k||^2 = \sum_{i \leq N} u_k(z_i)^2 \leq N/c(\chi)$, and then by (4.75) we have $\sum_k ||u_k||^2 \leq LN \log n$. Finally we apply Lemma 4.5.27, and we take for n the smallest integer $\geq \sqrt{N}$.

Exercise 4.5.31. Let us denote by ν the uniform measure on G and by d_{ν} the distance in the space $L^2(\nu)$. Prove that $\gamma_2(\mathcal{C}^*, d_{\nu}) \geq \sqrt{\log n}/L$. Warning: this is not so easy, and the solution is not provided. Hint: make sure you understand the previous chapter, an construct an appropriate tree. The ingredients on how to build that tree are contained in the proof given in the next section, and Section 3.2 should also be useful. You may assume that N is a power of 2 to save technical work. Furthermore, you may also look at [171] where trees were explicitly used.

4.6 Lower Bound for the Ajtai-Komlós-Tusnády Theorem

Recall that \mathcal{C}^* denotes the class of 1-Lipschitz functions on the unit square which are zero on the boundary of the square. We shall prove the following, where $(X_i)_{i < N}$ are i.i.d. in $[0, 1]^2$.

Theorem 4.6.1. We have

$$\mathsf{E}\sup_{f\in\mathcal{C}^*} \left|\sum_{i\leq N} \left(f(X_i) - \int f \mathrm{d}\lambda\right)\right| \geq \frac{1}{L}\sqrt{N\log N} .$$
(4.78)

Since

$$\sup_{f \in \mathcal{C}^*} \left| \sum_{i \le N} (f(X_i) - \int f d\lambda) \right| \le \sup_{f \in \mathcal{C}^*} \left| \sum_{i \le N} (f(X_i) - f(Y_i)) \right| \\ + \sup_{f \in \mathcal{C}^*} \left| \sum_{i \le N} (f(Y_i) - \int f d\lambda) \right|,$$

taking expectation and using (4.41) it follows from (4.78) that if the points Y_i are evenly spread then (provided $N \ge L$),

$$\mathsf{E}\sup_{f\in\mathcal{C}^*} \left|\sum_{i\leq N} (f(X_i) - f(Y_i))\right| \geq \frac{1}{L}\sqrt{N\log N} ,$$

and since $\mathcal{C}^* \subset \mathcal{C}$ the duality formula (4.29) implies that the expected cost of matching the points X_i and the points Y_i is at least $\sqrt{N \log N}/L$.

The proof of Theorem 4.6.1 occupies this entire section and starts now. The beautiful argument we present goes back to [5]. We can expect that this proof is non-trivial. To explain why, let us recall the set T used in the proof of Proposition 4.5.19. Analysis of the proof of that proposition leads us to guess that the reason why the bound it provides cannot be improved is that $\gamma_2(T, d_2)$ is actually of order $\sqrt{\log N}$ (and not of a smaller order). So a proof of (4.78) must contain a proof that this is the case. In the previous chapter, we learned a technique to prove such results, the construction of "trees". Not surprisingly, our proof implicitly uses a tree which witnesses just this, somewhat similar to the tree we constructed in Section 3.2.¹⁴

We may assume N large and we consider a number $r \in \mathbb{N}$ which is a small proportion of log N, say $r \simeq (\log N)/100$.¹⁵ The structure of the proof is to recursively construct for $k \leq r$ certain (random) functions f_k such that for any $q \leq r$

$$\sum_{k \le q} f_k \text{ is 1-Lipschitz} \tag{4.79}$$

and for each $k \leq r$,

$$\mathsf{E}\sum_{i\leq N} \left(f_k(X_i) - \int f_k \mathrm{d}\lambda \right) \geq \frac{\sqrt{N}}{L\sqrt{r}} .$$
(4.80)

 $^{^{14}}$ It should also be useful to solve Exercise 4.5.31 above.

¹⁵ We absolutely need for the proof a number r which is a proportion of log N, and taking a small proportion gives us some room. More specifically, each square of side 2^{-r} will have an area larger than, say, $1/\sqrt{N}$, so that it will typically contain many points X_i , as we will use when we perform normal approximation.

The function $f^* = \sum_{k < r} f_k$ is then 1-Lipschitz and satisfies

$$\mathsf{E}\sum_{i\leq N} \left(f^*(X_i) - \int f^* \mathrm{d}\lambda \right) \geq \frac{\sqrt{Nr}}{L}$$

This completes the proof of (4.78). The function f_k looks to what happens at scale 2^{-k} , and (4.80) states that each such scale 2^{-k} contributes about equally to the final result.

Following the details of the construction is not that difficult, despite the fact that ensuring (4.79) requires technical work. What is more difficult is to see why one makes such choices as we do. There is no magic there, making the right choices means that we have understood which aspect of the geometric complexity of the class C^* is relevant here.

The main idea behind the construction of the function f_k is that if we divide $[0,1]^2$ into little squares of side 2^{-k} , in each of these little squares there is some irregularity of the distribution of the X_i . The function f_k is a sum of terms, each corresponding to one of the little squares, see (4.90). It is designed to, in a sense, *add* the irregularities over these different squares.

The functions f_k will be built out of simple functions which we describe now. For $1 \le k \le r$ and $1 \le \ell \le 2^k$ we consider the function $f'_{k,\ell}$ on [0,1]defined as follows:

$$f'_{k,\ell}(x) = \begin{cases} 0 & \text{unless } x \in [(\ell-1)2^{-k}, \ell 2^{-k}[\\ 1 & \text{for } x \in [(\ell-1)2^{-k}, (\ell-1/2)2^{-k}[\\ -1 & \text{for } x \in [(\ell-1/2)2^{-k}, \ell 2^{-k}[. \end{cases}$$
(4.81)

We define

$$f_{k,\ell}(x) = \int_0^x f'_{k,\ell}(y) dy .$$
 (4.82)



Fig. 4.1. The graphs of $f_{k,1}$ and $f_{k,3}$.

We now list a few useful properties of these functions. In these formulas $\|.\|_2$ denotes the norm in $L^2([0,1])$, etc. The proofs of these assertions are completely straightforward and better left to the reader.
Lemma 4.6.2. The following holds true:

The family
$$(f'_{k,\ell})$$
 is orthogonal in $L^2([0,1])$. (4.83)

$$\|f'_{k,\ell}\|_2^2 = 2^{-k} . (4.84)$$

$$\|f_{k,\ell}'\|_1 = 2^{-k} . (4.85)$$

$$f_{k,\ell}$$
 is zero outside $](\ell-1)2^{-k}, \ell 2^{-k}[$. (4.86)

$$|f_{k,\ell}||_1 = 2^{-2k-2} . (4.87)$$

$$\|f'_{k,\ell}\|_{\infty} = 1 \; ; \; \|f_{k,\ell}\|_{\infty} = 2^{-k-1} \; . \tag{4.88}$$

$$\|f_{k,\ell}\|_2^2 = \frac{1}{12} 2^{-3k} .$$
(4.89)

The functions f_k will be of the type

$$f_k = \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell,\ell' \le 2^k} z_{k,\ell,\ell'} f_{k,\ell} \otimes f_{k,\ell'} , \qquad (4.90)$$

where $f_{k,\ell} \otimes f_{k,\ell'}(x,y) = f_{k,\ell}(x)f_{k,\ell'}(y)$ and $z_{k,\ell,\ell'} \in \{0,1,-1\}$. Note that $f_{k,\ell} \otimes f_{k,\ell'}$ is zero outside the little square $[(\ell - 1)2^{-k}, \ell 2^{-k}[\times](\ell' - 1)2^{-k}, \ell' 2^{-k}]$, and that these little squares are disjoint as ℓ and ℓ' vary. The term $z_{k,\ell,\ell'}f_{k,\ell} \otimes f_{k,\ell'}$ is designed to take advantage of the irregularity of the distribution of the X_i in the corresponding little square. The problem of course is to choose the numbers $z_{k,\ell,\ell'}$. There are two different ideas here. First, as a technical device to ensure (4.79), $z_{k,\ell,\ell'}$ may be set to zero. This will happen on a few little squares, those where are getting dangerously close to violate this condition (4.79). The second idea is that we will adjust the signs of $z_{k,\ell,\ell'}$ in a way that the contributions of the different little squares add properly (rather than canceling each other).

Let us now explain the scaling term $2^{k-5}/\sqrt{r}$ in (4.90). The coefficient 2^{-5} is just a small numerical constant ensuring that we have enough room. The idea of the term $2^k/\sqrt{r}$ is that the partial derivatives of f_k will be of order $1/\sqrt{r}$. Taking a sum of a most r such terms and taking cancellations in effect will give us partial derivatives which, at most of the points are ≤ 1 . This is formally expressed in the next lemma. So, such sums are not necessarily 1-Lipschitz, but are pretty close to being so and some minor tweaking will ensure that they are.

Lemma 4.6.3. Consider $q \leq r$. Consider a function of the type $f = \sum_{k \leq q} f_k$, where f_k is given by (4.90) and where $z_{k,\ell,\ell'} \in \{0, 1, -1\}$. Then

$$\left\|\frac{\partial f}{\partial x}\right\|_2 \le 2^{-6} . \tag{4.91}$$

Proof. First we note that

$$\frac{\partial f}{\partial x}(x,y) = \sum_{k \le q} \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell,\ell' \le 2^k} z_{k,\ell,\ell'} f'_{k,\ell}(x) f_{k,\ell'}(y) , \qquad (4.92)$$

which we rewrite as

$$\frac{\partial f}{\partial x}(x,y) = \sum_{k \le q} \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell \le 2^k} a_{k,\ell}(y) f'_{k,\ell}(x) ,$$

where $a_{k,\ell}(y) = \sum_{\ell' \leq 2^k} z_{k,\ell,\ell'} f_{k,\ell'}(y)$. Using (4.83) we obtain

$$\int \left(\frac{\partial f}{\partial x}\right)^2 dx = \sum_{k \le q} \frac{2^{2k-10}}{r} \sum_{\ell \le 2^k} a_{k,\ell}(y)^2 \|f'_{k,\ell}\|_2^2 \,.$$

Since $z_{k,\ell,\ell'}^2 \leq 1$, and since the functions $(f_{k,\ell})_{\ell \leq 2^k}$ have disjoint support, we have $a_{k,\ell}(y)^2 \leq \sum_{\ell' \leq 2^k} f_{k,\ell'}(y)^2$, so that

$$\int \left(\frac{\partial f}{\partial x}\right)^2 \mathrm{d}x \le \sum_{k \le q} \frac{2^{2k-10}}{r} \sum_{\ell,\ell' \le 2^k} \|f'_{k,\ell}\|_2^2 f_{k,\ell'}(y)^2.$$

Integrating in y and using (4.84) and (4.89) yields

$$\left\|\frac{\partial f}{\partial x}\right\|_{2}^{2} \leq \sum_{k \leq q} \frac{2^{2k-10}}{r} \frac{2^{-2k}}{12} = \frac{q}{r} \frac{2^{-10}}{12} \leq 2^{-12} .$$

Naturally, we have the same bound for $\|\partial f/\partial y\|_2$. These bounds do not imply that f is 1-Lipschitz, but they imply that it is 1-Lipschitz "most of the time".

We construct the functions f_k recursively. Having constructed f_1, \ldots, f_q , let $f := \sum_{k \leq q} f_k$, and assume that it is 1-Lipschitz. We will construct f_{q+1} of the type (4.90) by choosing the coefficients $z_{q+1,\ell,\ell'}$. Let us say that a square of the type

$$[(\ell-1)2^{-q}, \ell 2^{-q} [\times [(\ell'-1)2^{-q}, \ell' 2^{-q} [(4.93)$$

for $1 \leq \ell, \ell' \leq 2^q$ is a *q*-square. There are 2^{2q} such *q*-squares.

Definition 4.6.4. We say that a (q+1)-square is dangerous if it contains a point for which either $|\partial f/\partial x| \ge 1/2$ or $|\partial f/\partial y| \ge 1/2$. We say that it is safe if it is not dangerous.

The danger is that on this square (4.93) the function $f + f_{q+1}$ might not be 1-Lipschitz.

Lemma 4.6.5. At most half of the (q+1)-squares are dangerous, so at least half of the (q+1)-squares are safe.

This lemma is a consequence of the fact that "f is 1-Lipschitz most of the time." The proof is a bit technical, so we delay it to the end of the section.

The following is also a bit technical, but is certainly expected. It will also be proved later.

Lemma 4.6.6. If $z_{q+1,\ell,\ell'} = 0$ whenever the corresponding (q+1)-square (4.93) is dangerous, then $f + f_{q+1}$ is 1-Lipschitz.

We now complete the construction of the function f_{q+1} . For a dangerous square we set $z_{q+1,\ell,\ell'} = 0$. Let us define

$$h_{\ell,\ell'}(x) = f_{q+1,\ell} \otimes f_{q+1,\ell'}(x) - \int f_{q+1,\ell} \otimes f_{q+1,\ell'} \mathrm{d}\lambda$$
 .

Using (4.89) and (4.87) we obtain

$$\|h_{\ell,\ell'}\|_2^2 \ge 2^{-6q}/L \ . \tag{4.94}$$

Let us define then

$$D_{\ell,\ell'} = \sum_{i \le N} h_{\ell,\ell'}(X_i) .$$
(4.95)

For a safe square, we choose $z_{q+1,\ell,\ell'} = \pm 1$ such that

$$z_{q+1,\ell,\ell'} D_{\ell,\ell'} = |D_{\ell,\ell'}|.$$

Thus, if

$$f_{q+1} = \frac{2^{q-4}}{\sqrt{r}} \sum_{\ell,\ell' \le 2^{q+1}} z_{q+1,\ell,\ell'} f_{q+1,\ell} \otimes f_{q+1,\ell'}$$

we have

$$\sum_{i \le N} \left(f_{q+1}(X_i) - \int f_{q+1} d\lambda \right) = \frac{2^{q-4}}{\sqrt{r}} \sum_{\text{safe}} |D_{\ell,\ell'}| .$$
 (4.96)

where the sum is over all values of (ℓ, ℓ') such that the corresponding square (4.93) is safe.

We turn to the proof of (4.80) and for this we estimate $\mathsf{E}\sum_{\text{safe}} |D_{\ell,\ell'}|$. An obvious obstacle to perform this estimate is that the r.v.s $D_{\ell,\ell'}$ are not independent of the set of safe squares. But we know that at least half of the squares are safe, so we can bound below $\sum_{\text{safe}} |D_{\ell,\ell'}|$ by the sum of the 2^{2q+1} smallest among the 2^{2q+2} r.v.s $|D_{\ell,\ell'}|$.

Let us estimate $\mathsf{E}D^2_{\ell,\ell'}$. By definition (4.95) $D_{\ell,\ell'}$ is a sum $\sum_{i\leq N} h_{\ell,\ell'}(X_i)$ of independent centered r.v.s so that $\mathsf{E}D^2_{\ell,\ell'} = N \|h_{\ell,\ell'}\|_2^2$, and using (4.94) we obtain an estimate

$$\mathsf{E}D^2_{\ell,\ell'} \ge 2^{-6q} N/L \,. \tag{4.97}$$

Let us then *pretend* for a moment that the r.v.s $D_{\ell,\ell'}$ are Gaussian and independent as ℓ, ℓ' vary. For a Gaussian r.v. g we have $\mathsf{P}(|g| \ge (\mathsf{E}g^2)^{1/2}/100) \ge$

7/8. Then for each ℓ, ℓ' we have $|D_{\ell,\ell'}| \ge 2^{-3q}\sqrt{N}/L$ with probability $\ge 7/8$. In other words the r.v. $Y_{\ell,\ell'} = \mathbf{1}_{|D_{\ell,\ell'}|\ge 2^{-3q}\sqrt{N}/L}$ satisfies $\mathsf{E}Y_{\ell,\ell'} \ge 7/8$. Then Bernstein's inequality shows that with overwhelming probability at least 3/4 of these variables equal 1. For further use let us state the following more general principle. Considering M independent r.v.s $Z_i \in \{0,1\}$ with $\mathsf{P}(Z_i = 1) = a_i = \mathsf{E}Z_i$ then for u > 0 we have

$$\mathsf{P}\Big(\Big|\sum_{i\leq M} (Z_i - a_i)\Big| \ge uM\Big) \le 2\exp(-Mu^2/L) , \qquad (4.98)$$

and in particular $\mathsf{P}(\sum_{i \leq \underline{M}} Z_i \leq \sum_{i \leq \underline{M}} a_i - Mu) \leq 2 \exp(-Mu^2/L).$

Thus $|D_{\ell,\ell'}| \ge 2^{-3q} \sqrt{N}/L$ for at least 3/4 of the squares, so that at least 1/4 of the squares are both safe and satisfy this inequality. Consequently (4.80) holds (for q + 1 rather than q) follows from (4.96) as desired.

It is not exactly true that the r.v.s $D_{\ell,\ell'}$ are independent and Gaussian. Standard techniques exist to take care of this, namely Poissonization and normal approximation. There is all the room in the world because $r \leq (\log N)/100$. As these considerations are not related to the rest of the material of this work they are better omitted.

We now turn to the proofs of Lemmas 4.6.5 and 4.6.6. The next lemma prepares for these proofs.

Lemma 4.6.7. Consider $q \leq r$ and a function of the type $f = \sum_{k \leq q} f_k$, where f_k is given by (4.90) and where $z_{k,\ell,\ell'} \in \{0,1,-1\}$. Then

$$\left|\frac{\partial^2 f}{\partial x \partial y}\right| \le \frac{2^{q-4}}{\sqrt{r}} . \tag{4.99}$$

Proof. We have

$$\left|\frac{\partial^2 f}{\partial x \partial y}\right| = \left|\sum_{k \le q} \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell, \ell' \le 2^k} z_{k,\ell,\ell'} f'_{k,\ell} \otimes f'_{k,\ell'}\right| \le \sum_{k \le q} \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell,\ell' \le 2^k} |f'_{k,\ell} \otimes f'_{k,\ell'}| \le \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell,\ell' \le 2^k} |f'_{k,\ell'} \otimes f'_{k,\ell'}| \le \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell' \le 2^k} |f'_{k,\ell'} \otimes f'_{k,\ell'} \otimes f'_{k,\ell'}| \le \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell' \le 2^k} |f'_{k,\ell'} \otimes f'_{k,\ell'} \otimes f'_{k,\ell'}| \le \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell' \le 2^k} |f'_{k,\ell'} \otimes f'_{k,\ell'} \otimes f'_{k,\ell'}| \le \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell' \le 2^k} |f'_{k,\ell'} \otimes f'_{k,\ell'} \otimes f'_{k,\ell'} \otimes f'_{k,\ell'}| \le \frac{2^{k-5}}{\sqrt{r}} \sum_{\ell' \le 2^k} |f'_{k,\ell'} \otimes f'_{k,\ell'} \otimes f'$$

The functions $f'_{k,\ell} \otimes f'_{k,\ell'}$ have disjoint support, and by the first part of (4.88) $|f'_{k,\ell} \otimes f'_{k,\ell'}| \leq 1$. Also, $\sum_{k \leq q} 2^k \leq 2^{q+1}$.

Proof of Lemma 4.6.5. We will observe from the definition that all functions $f'_{k,\ell}$ for $k \leq q$ are constant on the intervals $[\ell 2^{-q-1}, (\ell+1)2^{-q-1}]$. Thus according to (4.92), on a (q+1)-square, $\partial f/\partial x$ does not depend on x. If (x, y) and (x', y') belong to the same (q+1)-square then

$$\frac{\partial f}{\partial x}(x',y') = \frac{\partial f}{\partial x}(x,y') . \qquad (4.100)$$

Moreover $|y - y'| \le 2^{-q-1}$ so that (4.99) implies

$$\left|\frac{\partial f}{\partial x}(x,y) - \frac{\partial f}{\partial x}(x,y')\right| \le |y-y'|\frac{2^{q-5}}{\sqrt{r}} \le \frac{2^{-6}}{\sqrt{r}}$$

and combining with (4.100) we obtain

$$\left|\frac{\partial f}{\partial x}(x,y) - \frac{\partial f}{\partial x}(x',y')\right| = \left|\frac{\partial f}{\partial x}(x,y) - \frac{\partial f}{\partial x}(x,y')\right| \le \frac{2^{-6}}{\sqrt{r}}$$

In particular if a (q + 1)-square contains a point at which $|\partial f/\partial x| \geq 1/2$, then at each point of this square we have $|\partial f/\partial x| \geq 1/2 - 2^{-6}/\sqrt{r} \geq 1/4$. The proportions α of (q + 1)-squares with this property satisfies $\alpha(1/4)^2 \leq ||\partial f/\partial x||_2^2 \leq 2^{-12}$, where we have used (4.91) in the last inequality. This implies that at most a proportion 2^{-8} of (q + 1)-squares can contain a point with $|\partial f/\partial x| \geq 1/2$. Repeating the same argument for $\partial f/\partial y$ shows that as desired at most half of (q + 1)-squares are dangerous.

Proof of Lemma 4.6.6. To ensure that $g := f + f_{q+1}$ is 1-Lipschitz, it suffices to ensure that it is 1-Lipschitz on each (q + 1)-square. When the square is dangerous, $f_{q+1} = 0$ on this square by construction and g is 1-Lipschitz on it because there g = f and f is 1-Lipschitz.

When the square is safe, everywhere on the square we have $|\partial f/\partial x| \leq 1/2$ and $|\partial f/\partial y| \leq 1/2$. Now the second part of (4.88) implies

$$\left\|\frac{\partial g}{\partial x} - \frac{\partial f}{\partial x}\right\|_{\infty} = \left\|\frac{1}{\sqrt{r}}2^{q-4}\sum_{\ell,\ell' \le 2^{q+1}} z_{q+1,\ell,\ell'}f'_{q+1,\ell} \otimes f_{q+1,\ell'}\right\|_{\infty} \le \frac{1}{2^5\sqrt{r}}$$

and

$$\left\|\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\right\|_{\infty} = \left\|\frac{1}{\sqrt{r}} 2^{q-4} \sum_{\ell,\ell' \le 2^{q+1}} z_{q+1,\ell,\ell'} f_{q+1,\ell} \otimes f'_{q+1,\ell'}\right\|_{\infty} \le \frac{1}{2^5 \sqrt{r}} ,$$

where we have used that the elements of the sum have disjoint supports. So we are certain that at each point of a safe square we have $|\partial g/\partial x| \leq 1/\sqrt{2}$ and $|\partial g/\partial y| \leq 1/\sqrt{2}$, and hence that g is 1-Lipschitz on a safe square. \Box

Exercise 4.6.8. This is a continuation of Exercise 4.5.23. Adapt the method you learned in this section to prove that the bound (4.69) is of the correct order.

4.7 The Leighton-Shor Grid Matching Theorem

Theorem 4.7.1 ([77]). If the points $(Y_i)_{i \leq N}$ are evenly spread and if $(X_i)_{i \leq N}$ are *i.i.d.* uniform over $[0,1]^2$, then (for $N \geq 2$), with probability at least $1 - L \exp(-(\log N)^{3/2}/L)$ we have

4.7 The Leighton-Shor Grid Matching Theorem 139

$$\inf_{\pi} \sup_{i \le N} d(X_i, Y_{\pi(i)}) \le L \frac{(\log N)^{3/4}}{\sqrt{N}} .$$
(4.101)

In particular

$$\mathsf{E}\inf_{\pi} \sup_{i \le N} d(X_i, Y_{\pi(i)}) \le L \frac{(\log N)^{3/4}}{\sqrt{N}} .$$
(4.102)

To deduce (4.102) from (4.101) one simply uses any matching in the (rare) event that (4.101) fails. We shall prove in Section 4.8 that the inequality (4.102) can be reversed. A close cousin of this theorem can be found in Appendix A.

A first simple idea is that to prove Theorem 4.7.1 we do not care about what happens at a scale smaller than $(\log N)^{3/4}/\sqrt{N}$. Therefore consider the largest integer ℓ_1 with $2^{-\ell_1} \ge (\log N)^{3/4}/\sqrt{N}$ (so that in particular $2^{\ell_1} \le \sqrt{N}$). We divide $[0,1]^2$ into little squares of side $2^{-\ell_1}$. For each such square, we are interested in how many points (X_i) it contains, but we do not care where these points are located in the square. We shall deduce Theorem 4.7.1 from a discrepancy theorem for a certain class of functions.¹⁶ What we really have in mind is the class of functions which are indicators of a union A of little squares with sides of length $2^{-\ell_1}$, and such that the boundary of A has a given length. It turns out that we shall have to parametrize the boundaries of these sets by curves, so it is convenient to turn things around and to consider the class of sets A that are the interiors of curves of given length.

To make things precise, let us define the grid G of $[0,1]^2$ of mesh width $2^{-\ell_1}$ by

$$G = \left\{ (x_1, x_2) \in [0, 1]^2 ; 2^{\ell_1} x_1 \in \mathbb{N} \text{ or } 2^{\ell_1} x_2 \in \mathbb{N} \right\}$$

A vertex of the grid is a point $(x_1, x_2) \in [0, 1]^2$ with $2^{\ell_1} x_1 \in \mathbb{N}$ and $2^{\ell_1} x_2 \in \mathbb{N}$. There are $(2^{\ell_1} + 1)^2$ vertices. An *edge* of the grid is the segment between two vertices that are at distance $2^{-\ell_1}$ of each other. A square of the grid is a square of side $2^{-\ell_1}$ whose edges are edges of the grid. Thus, an edge of the grid is a subset of the grid, but a square of the grid is not a subset of the grid, see Figure 4.2.

A curve is the image of a continuous map $\varphi : [0,1] \to \mathbb{R}^2$. We say that the curve is a simple curve if it is one-to-one on [0,1[. We say that the curve is traced on G if $\varphi([0,1]) \subset G$, and that it is closed if $\varphi(0) = \varphi(1)$. If C is a closed simple curve in \mathbb{R}^2 , the set $\mathbb{R}^2 \setminus C$ has two connected components. One of these is bounded. It is called the interior of C and is denoted by $\overset{o}{C}$.

The proof of Theorem 4.7.1 has a probabilistic part (the hard one) and a deterministic part. The probabilistic part states that with high probability the number of points inside a closed curve differs from its expected value by at most the length of the curve times $L\sqrt{N}(\log N)^{3/4}$. The deterministic part will be given at the end of the section and will show how to deduce Theorem 4.7.1 from Theorem 4.7.2.

¹⁶ This is the case for every matching theorem we prove.



Fig. 4.2. A square A, and edge e, a vertex V and a simple curve C traced on G.

Theorem 4.7.2. With probability at least $1 - L \exp(-(\log N)^{3/2}/L)$, the following occurs. Given any closed simple curve C traced on G, we have

$$\left|\sum_{i\leq N} \left(\mathbf{1}_{\overset{o}{C}}(X_i) - \lambda(\overset{o}{C})\right)\right| \leq L\ell(C)\sqrt{N}(\log N)^{3/4}, \qquad (4.103)$$

where $\lambda(\overset{\circ}{C})$ is the area of $\overset{\circ}{C}$ and $\ell(C)$ is the length of C.

We will reduce the proof of this theorem to the following result, which concerns curves of a given length going through a given vertex.

Proposition 4.7.3. Consider a vertex τ of G and $k \in \mathbb{Z}$. Define $\mathcal{C}(\tau, k)$ as the set of closed simple curves traced on G that pass through τ^{17} and have length $\leq 2^k$. Then, if $-\ell_1 \leq k \leq \ell_1 + 2$, with probability at least $1 - L \exp(-(\log N)^{3/2}/L)$, for each $C \in \mathcal{C}(\tau, k)$ we have

$$\left|\sum_{i \le N} \left(\mathbf{1}_{\stackrel{o}{C}}(X_i) - \lambda(\stackrel{o}{C})\right)\right| \le L2^k \sqrt{N} (\log N)^{3/4} .$$
(4.104)

It would be easy to control the left-hand side if one considered only curves with a simple pattern, such as boundaries of rectangles. The point however is that the curves we consider can be very complicated, and the longer we allow them to be, the more so. Before we discuss Proposition 4.7.3 further, we show that it implies Theorem 4.7.2.

¹⁷ that is, τ is an end vertex of an edge which belongs to the curve.

Proof of Theorem 4.7.2. Since there are at most $(2^{\ell_1} + 1)^2 \leq LN$ choices for the vertex τ , we can assume with probability at least

$$1 - L(2^{\ell_1} + 1)^2 (2\ell_1 + 4) \exp(-(\log N)^{3/2}/L) \ge 1 - L' \exp(-(\log N)^{3/2}/L')$$
(4.105)

that (4.104) occurs for all choices of $C \in \mathcal{C}(\tau, k)$, for any τ and any k with $-\ell_1 \leq k \leq \ell_1 + 2$.

Consider a simple curve C traced on G. Bounding the length of C by the total length of the edges of G, we have $2^{-\ell_1} \leq \ell(C) \leq 2(2^{\ell_1} + 1) \leq 2^{\ell_1+2}$. Then the smallest integer k for which $\ell(C) \leq 2^k$ satisfies $-\ell_1 \leq k \leq \ell_1 + 2$. Since $2^k \leq 2\ell(C)$ the proof is finished by (4.104).

Exercise 4.7.4. Prove the second inequality in (4.105) in complete detail.

The main step to prove Proposition 4.7.3 is the following.

Proposition 4.7.5. Consider a vertex τ of G and $k \in \mathbb{Z}$. Define $C(\tau, k)$ as in Proposition 4.7.3. Then, if $-\ell_1 \leq k \leq \ell_1 + 2$ we have

$$\mathsf{E}\sup_{C \in C(\tau,k)} \left| \sum_{i \le N} \left(\mathbf{1}_{\mathring{C}}(X_i) - \lambda(\mathring{C}) \right) \right| \le L2^k \sqrt{N} (\log N)^{3/4} .$$
(4.106)

Proof of Proposition 4.7.3. To prove Proposition 4.7.3 we have to go from the control in expectation provided by (4.106) to the control in probability of (4.104). There is powerful tool to do this: concentration of measure. The function

$$f(x_1,\ldots,x_N) = \sup_{C \in \mathcal{C}(\tau,k)} \left| \sum_{i \le N} \left(\mathbf{1}_{\overset{o}{C}}(x_i) - \lambda(\overset{o}{C}) \right) \right|$$

of points $x_1, \ldots, x_N \in [0, 1]^2$ has the property that changing the value of a given variable x_i can change the value of f by at most one. One of the earliest "concentration of measure" results (for which we refer to [74]) asserts that for such a function the r.v. $W = f(X_1, \ldots, X_N)$ satisfies a deviation inequality of the form

$$\mathsf{P}(|W - \mathsf{E}W| \ge u) \le 2\exp\left(-\frac{u^2}{2N}\right). \tag{4.107}$$

Using (4.106) to control EW and taking $u = L2^k \sqrt{N} (\log N)^{3/4}$ proves Proposition 4.7.3 in the case $k \ge 0$. A little bit more work is needed when k < 0. In that case a curve of length 2^k is entirely contained in the square V of center τ and side 2^{k+1} and $\mathbf{1}_{\hat{C}}(X_i) = 0$ unless $X_i \in V$. To take advantage of this, we work conditionally on $I = \{i \le N; X_i \in V\}$ and we can then use (4.107) with card I instead of N. This provides the desired inequality when card $I \le L2^{2k}N$. On the other hand, by (4.98) and since $\lambda(V) = 2^{2k+2}$ we have $\mathsf{P}(\operatorname{card} I \ge L2^{2k}N) \le \exp(-N2^{2k}) \le L\exp(-(\log N)^{3/2}/L)$ because $k \ge -\ell_1$ and the choice of ℓ_1 .

We start the proof of Proposition 4.7.5. We denote by \mathcal{F}_k the class of functions of the type $\mathbf{1}_{\mathcal{C}}^{\circ}$, where $C \in \mathcal{C}(\tau, k)$ so we can rewrite (4.106) as

$$\mathsf{E}\sup_{f\in\mathcal{F}_{k}}\left|\sum_{i\leq N} (f(X_{i}) - \int f \mathrm{d}\lambda)\right| \leq L2^{k}\sqrt{N}(\log N)^{3/4}.$$
(4.108)

The key point again is the control on the size of \mathcal{F}_k with respect to the distance of $L^2(\lambda)$. The difficult part of this control is the following.

Proposition 4.7.6. We have

$$\gamma_2(\mathcal{F}_k, d_2) \le L2^k (\log N)^{3/4}$$
. (4.109)

Another much easier fact is the following.

Proposition 4.7.7. We have

$$\gamma_1(\mathcal{F}_k, d_\infty) \le L2^k \sqrt{N} . \tag{4.110}$$

Proof of (4.108) and of Proposition 4.7.5. Combine Proposition 4.7.6, Proposition 4.7.7 and Theorem 4.5.16. $\hfill \Box$

Let us first prove the easy Proposition 4.7.7.

Lemma 4.7.8. We have card $C(\tau, k) \leq 2^{2^{k+\ell_1+1}} = N_{k+\ell_1+1}$.

Proof. A curve $C \in \mathcal{C}(\tau, k)$ consists of at most $2^{k+\ell_1}$ edges of G. If we move through C, at each vertex of G we have at most 4 choices for the next edge, so $\operatorname{card} \mathcal{C}(\tau, k) \leq 4^{2^{k+\ell_1}} = N_{k+\ell_1+1}$.

Proof of Proposition 4.7.7. Generally speaking, a set T of cardinality $\leq N_m$ and diameter Δ satisfies $\gamma_1(T,d) \leq L\Delta 2^m$, as is shown by taking $\mathcal{A}_n = \{T\}$ for n < m and $A_m(t) = \{t\}$. We use this for $T = \mathcal{F}_k$, so that card T =card $\mathcal{C}(\tau,k) \leq N_{k+\ell_1+1}$ by Lemma 4.7.8, and $2^{k+\ell_1+1} \leq L2^k \sqrt{N}$.

We now attack the difficult part, the proof of Proposition 4.7.6. The exponent 3/4 occurs through the following general principle, where we recall that if d is a distance, so is \sqrt{d} .

Lemma 4.7.9. Consider a finite metric space (T,d) with card $T \leq N_m$. Then

$$\gamma_2(T, \sqrt{d}) \le m^{3/4} \gamma_{1,2}(T, d)^{1/2}$$
. (4.111)

Proof. Since T is finite there exists an admissible sequence (\mathcal{A}_n) of T such that

$$\forall t \in T, \sum_{n \ge 0} (2^n \Delta(A_n(t), d))^2 \le \gamma_{1,2}(T, d)^2.$$
(4.112)

Without loss of generality we can assume that $A_m(t) = \{t\}$ for each t, so that in (4.112) the sum is over $n \leq m-1$. Now

$$\Delta(A,\sqrt{d}) = \Delta(A,d)^{1/2}$$

so that, using Hölder's inequality,

$$\sum_{0 \le n \le m-1} 2^{n/2} \Delta(A_n(t), \sqrt{d}) = \sum_{0 \le n \le m-1} (2^n \Delta(A_n(t), d))^{1/2}$$
$$\le m^{3/4} \Big(\sum_{n \ge 0} (2^n \Delta(A_n(t), d))^2 \Big)^{1/4}$$
$$\le m^{3/4} \gamma_{1,2} (T, d)^{1/2} ,$$

which concludes the proof.

Let us denote by $A \triangle B$ the symmetric difference $(A \setminus B) \cup (B \setminus A)$ between two sets A and B. On the set of closed simple curves traced on G, we define the distance d_1 by $d_1(C, C') = \lambda(\overset{\circ}{C} \triangle \overset{\circ}{C}')$ and the distance

$$\delta(C_1, C_2) := \left\| \mathbf{1}_{\mathring{C}_1} - \mathbf{1}_{\mathring{C}_2} \right\|_2 = (\lambda(\mathring{C}_1 \triangle \mathring{C}_2))^{1/2} = \left(d_1(C_1, C_2) \right)^{1/2}, \quad (4.113)$$

so that

 $\gamma_2(\mathcal{F}_k, d_2) = \gamma_2(\mathcal{C}(\tau, k), \delta) = \gamma_2(\mathcal{C}(\tau, k), \sqrt{d_1}) ,$

and using Lemma 4.7.8 and (4.111) for $m := k + \ell_1 + 1$ we obtain

$$\gamma_2(\mathcal{F}_k, d_2) \le L(\log N)^{3/4} \gamma_{1,2}(\mathcal{C}(\tau, k), d_1)^{1/2}$$

because $m \leq L \log N$ for $k \leq \ell_1 + 2$.

Therefore it remains only to prove the following.

Proposition 4.7.10. We have

$$\gamma_{1,2}(\mathcal{C}(\tau,k), d_1) \le L2^{2k}$$
 (4.114)

The reason why this is true is that the metric space (\mathcal{L}, d_2) of Proposition 4.1.8 satisfies $\gamma_{1,2}(\mathcal{L}, d_2) < \infty$, while $(\mathcal{C}(\tau, k), d_1)$ is a Lipschitz image of a subset of this metric space (\mathcal{L}, d_2) . The elementary proof of the following may be found in Section B.2.

Lemma 4.7.11. There exists a map W from a subset T of \mathcal{L} onto $\mathcal{C}(\tau, k)$ which for any $f_0, f_1 \in T$ satisfies

$$d_1(W(f_0), W(f_1)) \le L 2^{2k} \|f_0 - f_1\|_2 .$$
(4.115)

To conclude the proof of Proposition 4.7.10 we check that the functionals $\gamma_{\alpha,\beta}$ behave as expected under Lipschitz maps.

Lemma 4.7.12. Consider two metric spaces (T,d) and (U,d') and a map $f:(T,d) \to (U,d')$ which is onto and satisfies

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

for a certain constant A. Then

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

Proof. This is really obvious when f is one-to-one. We reduce to that case by considering a map $\varphi : U \to T$ with $f(\varphi(x)) = x$ and replacing T by $\varphi(U)$. \Box



Fig. 4.3. A union A of little squares, and the boundary of A'.

It remains to deduce Theorem 4.7.1 from Theorem 4.7.2. The argument is purely deterministic and unrelated to any other material in the present book. The basic idea is very simple, and to keep it simple we describe it in slightly imprecise terms. Consider a union A of little squares of side length $2^{-\ell_1}$ and the union A' of all the little squares that touch A, see Figure 4.3.

We want to prove that A' contains as many points Y_i as A contains points X_i , so that by Hall's Marriage Lemma each point X_i can be matched to a point Y_i in the same little square, or in a neighbor of it. Since the points Y_i are evenly spread the number of such points in A' is very nearly $N\lambda(A')$. There may be more than $N\lambda(A)$ points X_i in A, but (4.103) tells us that the excess number of points cannot be more than a proportion of the length ℓ of the boundary of A. The marvelous fact is that we may also expect that $\lambda(A') - \lambda(A)$ is also proportional to ℓ , so that we may hope that the excess number of points X_i in A should not exceed $N(\lambda(A') - \lambda(A))$, proving the result. The proportionality constant is not quite right to make the argument work, but this difficulty is bypassed simply by applying the same argument to a slightly coarser grid.

When one tries to describe precisely what is meant by the previous argument, one has to check a number of details. This elementary task which requires patience is performed in Appendix B.3.

4.8 Lower Bound for the Leighton-Shor Theorem

Theorem 4.8.1. If the points $(X_i)_{i \leq N}$ are *i.i.d.* uniform over $[0,1]^2$ and the points $(Y_i)_{i \leq N}$ are evenly spread, then

$$\mathsf{E}\inf_{\pi} \max_{i \le N} d(X_i, Y_{\pi(i)}) \ge \frac{(\log N)^{3/4}}{L\sqrt{N}} .$$
(4.116)

We consider the class of functions

$$\mathcal{C} = \left\{ f: [0,1] \to [0,1] \; ; \; f(0) = f(1) = 0 \; ; \; \int_0^1 f'^2(x) \mathrm{d}x \le 1 \right\} \; . \tag{4.117}$$

For $f \in \mathcal{C}$ we consider its subgraph

$$S(f) := \{ (x, y) \in [0, 1]^2 ; y \le f(x) \} .$$
(4.118)

To prove (4.116) the key step will be to show that with high probability we may find $f \in \mathcal{C}$ with

card{
$$i \le N$$
; $X_i \in S(f)$ } $\ge N\lambda(S(f)) + \frac{1}{L}\sqrt{N}(\log N)^{3/4}$. (4.119)

With a little more work, we could actually prove that we can find such a function f which moreover satisfies $|f'| \leq 1$. This extra work is not needed. The key property of f here is that its graph has a bounded length and this is already implied by the condition $||f'||_2 \leq 1$, since the length of this graph is $\int_0^1 \sqrt{1 + f'^2(x)} dx \le 2$.

Lemma 4.8.2. The set of points within distance $\epsilon > 0$ of the graph of f has an area $\leq L\epsilon$. The set of points within distance $\epsilon > 0$ of S(f) has an area $\leq \lambda(S(f)) + L\epsilon.$

Proof. The graph of $f \in \mathcal{C}$ has length ≤ 2 . One can find a subset of the graph of f of cardinality $\leq L/\epsilon$ such that each point of the graph is within distance ϵ of this set¹⁸ A point within distance ϵ of the graph then belongs to one of L/ϵ balls of radius 2ϵ . This proves the first assertion. The second assertion follows from the fact that a point which is within distance ϵ of S(f) either belongs to S(f) or is within distance ϵ of the graph of f. \square

Proof of Theorem 4.8.1. We prove that when there exists a function f satisfying (4.119) then $\inf_{\pi} \max_{i < N} d(X_i, Y_{\pi(i)}) \ge (\log N)^{3/4} / L \sqrt{N}$. Let us denote by $S(f)_{\epsilon}$ the ϵ -neighborhood¹⁹ of S(f) in $[0,1]^2$. We first observe that for any $f \in \mathcal{C}$ we have

 $^{^{18}}$ This is true for any curve of length 2. If one consider a parameterization $\varphi(t)$ $0 \le t \le 2$ of the curve by arc length, the points $\varphi(k\epsilon)$ for $k \le 2/\epsilon$ have this property. ¹⁹ That is, the set of points within distance $\leq \epsilon$ of a point of S(f).

$$\operatorname{card}\{i \le N \; ; \; Y_i \in S(f)_{\epsilon}\} \le N\lambda(S(f)) + L\epsilon N + L\sqrt{N} \; . \tag{4.120}$$

This is because, by definition of an evenly spread family, each point Y_i belongs to a small rectangle R_i of area 1/N and of diameter $\leq 10/\sqrt{N}$, and a pessimistic upper bound for the left-hand side of (4.120) is the number of such rectangles that intersect $S(f)_{\epsilon}$. These rectangles are entirely contained in the set of points within distance L/\sqrt{N} of $S(f)_{\epsilon}$, i.e. in the set of points within distance $\leq \epsilon + L/\sqrt{N}$ of S(f) and by Lemma 4.8.2 this set has area $\leq \lambda(S(f)) + L\epsilon + L/\sqrt{N}$, hence the bound (4.120).

Consequently (and since we may assume that N is large enough) (4.119) implies that for $\epsilon = (\log N)^{3/4}/(L\sqrt{N})$ it holds that

$$\operatorname{card}\{i \leq N ; Y_i \in S(f)_{\epsilon}\} < \operatorname{card}\{i \leq N ; X_i \in S(f)\}$$

and therefore any matching must pair at least one point $X_i \in S(f)$ with a point $Y_j \notin S(f)_{\epsilon}$, so that $\max_{i \leq N} d(X_i, Y_{\pi(i)}) \geq \epsilon$.

Recalling the functions $f_{k,\ell}$ of (4.82), we consider now an integer $c \geq 2$ which will be determined later. The purpose of c is to give us room. Thus, by (4.87),

$$\int_0^1 f_{ck,\ell}(x) \mathrm{d}x = 2^{-2ck-2} . \tag{4.121}$$

Let us set

$$ilde{f}_{k,\ell} = rac{1}{\sqrt{r}} f_{ck,\ell} \; .$$

Consider the functions of the type

$$f = \sum_{k \le r} f_k$$
 with $f_k = \sum_{1 \le \ell \le 2^{ck}} x_{k,\ell} \tilde{f}_{k,\ell}$, (4.122)

where $x_{k,\ell} \in \{0,1\}$. Then f(0) = f(1) = 0.

Lemma 4.8.3. A function f of the type (4.122) satisfies

$$\int_0^1 f'(x)^2 \mathrm{d}x \le 1 \;. \tag{4.123}$$

Proof. Using (4.83) and (4.84) we obtain

$$\int_0^1 f'(x)^2 \mathrm{d}x = \sum_{k \le r} \sum_{\ell \le 2^{ck}} \frac{x_{k,\ell}^2}{r} \|f_{ck,\ell}\|_2^2 = \sum_{k \le r} \sum_{\ell \le 2^{ck}} \frac{x_{k,\ell}^2}{r} 2^{-ck} \le 1 \,. \qquad \Box$$

Consequently each function of the type (4.122) belongs the class C of (4.117). *Proof of (4.119).* Given N large we choose r as the largest integer for which $2^{cr} \leq N^{1/100}$, so that $r \leq \log N/Lc$. The construction of the functions f_k is inductive. Assume that we have already constructed f_1, \ldots, f_q , and let $g = \sum_{k \leq q} f_k$. For $\ell \leq 2^{c(q+1)}$ let us consider the region

$$R_{\ell} := S(g + \tilde{f}_{q+1,\ell}) \setminus S(g) ,$$

so that by (4.121)

$$\lambda(R_{\ell}) = \frac{2^{-2c(q+1)}}{4\sqrt{r}} . \tag{4.124}$$

These regions are disjoint because the functions $f_{q+1,\ell}$ have disjoint support. Furthermore if we choose $f_{q+1} = \sum_{\ell \leq 2^{c(q+1)}} x_{q+1,\ell} \tilde{f}_{q+1,\ell}$ where $x_{q+1,\ell} \in \{0,1\}$ then we have

$$S(g+f_{q+1})\setminus S(g) = \bigcup_{\ell\in J} R_\ell$$
,

where

$$J = \{\ell \le 2^{c(q+1)} ; x_{q+1,\ell} = 1\},\$$

and thus

$$\lambda(S(g+f_{q+1}) \setminus S(g)) = \sum_{J} \lambda(R_{\ell}) . \qquad (4.125)$$

Since our goal is to construct functions such that there is an excess of points X_i in their subgraph, we do the obvious thing, we take $x_{q+1,\ell} = 1$ if there is an excess of points X_i in R_ℓ , that if

$$\delta_{\ell} := \operatorname{card}\{i \le N; X_i \in R_{\ell}\} - N\lambda(R_{\ell}) \ge 0 , \qquad (4.126)$$

and otherwise we set $x_{k+1,\ell} = 0$. We have, recalling (4.125),

$$\operatorname{card}\{i \leq N ; X_i \in S(g+f_{q+1}) \setminus S(g)\} = \sum_J \operatorname{card}\{i \leq N ; X_i \in R_\ell\}$$
$$= \sum_J \delta_\ell + N\lambda(S(g+f_{q+1}) \setminus S(g)) . \quad (4.127)$$

We will show that with high probability we have $\sum_J \delta_\ell \geq \sqrt{N}/(Lr^{1/4})$. Recalling that $g = \sum_{k \leq q} f_k$ and $g + f_{q+1} = \sum_{k \leq q+1} f_k$, summation of the inequalities (4.127) over q < r then proves (4.119), where f is the function $\sum_{k \leq r} f_k$.

Let us say that the region R_{ℓ} is *favorable* if

$$\delta_{\ell} \ge \sqrt{N\lambda(R_{\ell})}/L^* = 2^{-c(q+1)}\sqrt{N}/(Lr^{1/4})$$
,

where the universal constant L^* will be determined later. The idea underlying this definition is that given a subset A of the square, with $1/N \ll \lambda(A) \leq 1/2$, the number of points X_i which belong to A has typical fluctuations of order

 $\sqrt{N\lambda(A)}$. Since $\delta_{\ell} \geq 0$ for $\ell \in J$, and since by construction $\ell \in J$ when R_{ℓ} is favorable, we have

$$\sum_{J} \delta_{\ell} \ge \operatorname{card}\{\ell; R_{\ell} | \text{favorable}\} \times 2^{-c(q+1)} \sqrt{N} / (Lr^{1/4}) .$$

To conclude the proof it then suffices to show that with overwhelming probability at least a fixed proportion of the regions R_{ℓ} for $\ell \leq 2^{c(q+1)}$ are favorable. One has to be cautious that the r.v.s X_i are *not* independent of the function g and of the regions R_{ℓ} because in particular the construction of g uses the values of the X_i .

One simple way around that difficulty is to proceed as follows. There are at most $\prod_{k \leq q} 2^{2^{ck}} \leq 2^{2^{cq+1}}$ possibilities for g. To each of these possibilities corresponds a family of $2^{2^{c(q+1)}}$ regions R_{ℓ} . If we can ensure that with overwhelming probability for each of these families a least a fixed proportion of the R_{ℓ} are favorable we are done. Since there are at most $2^{2^{cq+1}}$ families, it suffices to prove that for a given family this fails with probability $\leq 2^{-2^{cq+2}}$. To achieve this we proceed as follows. By normal approximation of the tails of the binomial law, there exists a constant L^* and a number $N_0 > 0$ such that given any set $A \subset [0, 1]^2$ with $1/2 \geq \lambda(A)$ and $N\lambda(A) \geq N_0$ we have

$$\mathsf{P}\big(\operatorname{card}\{i \le N; X_i \in A\} - N\lambda(A) \ge \sqrt{N\lambda(A)}/L^*\big) \ge 1/4 .$$
(4.128)

Since c is a universal constant and $2^{rc} \leq N^{1/100}$, (4.124) shows that $N\lambda(R_{\ell})$ becomes large with N. In particular (4.128) shows that the probability that a given region R_{ℓ} is favorable is $\geq 1/4$. Now, using Poissonization, we can pretend that these probabilities are independent as ℓ varies. As noted in (4.98), given M independent r.v.s $Z_i \in \{0,1\}$ with $\mathsf{P}(Z_i = 1) \geq 1/4$, then $\mathsf{P}(\sum_{i \leq M} Z_i \leq M/8) \leq \exp(-\beta M)$ for some universal constant β . Since here we have $M = 2^{c(q+1)}$ then $\exp(-\beta M) = \exp(-\beta 2^{c-2}2^{cq+2})$. This is $\leq 2^{-2^{cq+2}}$ as required provided we have chosen c large enough that $\beta 2^{c-2} \geq 1$.

4.9 For the Expert Only

Having proved both the Ajtai-Komlós-Tusnády and the Leighton-Shor matching theorems, we should not fall under the illusion that we understand everything about matchings. The most important problem left is arguably the ultimate matching conjecture, stated later as Problem 17.1.2. A first step in that direction would be to answer the following question.²⁰

Question 4.9.1 Can we find a matching which achieves simultaneously both (4.36) and (4.101)?

 $^{^{20}\,}$ The difference between a Problem and a Question is that a Question is permitted to sound less central.

The existence of such a matching does not seem to be of any particular importance, but the challenge is that the Ajtai-Komlós-Tusnády (AKT) theorem and the Leighton-Shor matching theorems are proved by rather different routes and it is far from obvious to find a common proof.

In the rest of the section, we discuss a special matching result. Consider the space $T = \{0,1\}^{\mathbb{N}}$ provided with the distance $d(t,t') = 2^{-j}$, where $j = \min\{i \geq 1; t_i \neq t'_i\}$ for $t = (t_i)_{i\geq 1}$. This space somewhat resembles the unit interval, in the sense that $N(T, d, \epsilon) \leq L\epsilon^{-1}$ for $\epsilon \leq 1$. The space of Exercise 4.5.23 is essentially the space $T \times T$. The AKT theorem tells us what happens for matchings in $[0, 1]^2$ and Exercise 4.5.23 tells us what happens for matchings in T^2 . But what happens in the space $U := [0, 1] \times T$? It does not really matter which specific sensible distance we use on U, let us say that we define d((x, t), (x', t')) = |x - x'| + d(t, t').

Theorem 4.9.2. The expected cost of the optimal matching of N random *i.i.d* uniformly distributed²¹ points in U with N evenly spread points is exactly of order $\sqrt{N}(\log N)^{3/4}$.

The appealing part of this special result is of course the fractional power of log. This result is as pretty as almost anything found in this book, but its special nature makes it appropriate to guide the (expert) reader to the proof through exercises.

Let us start with a finite approximation of T. We consider the space $T_m = \{0, 1\}^m$ provided with the distance defined for $t \neq t'$ by $d(t, t') = 2^{-j}$, where $j = \min\{i \ge 1; t_i \neq t'_i\}$ for $t = (t_i)_{i \le m}$. We set $U_m = [0, 1] \times T_m$ and we denote by θ_m the uniform measure on U_m . Surely the reader who has reached this stage knows how to deduce²² the upper bound of Theorem 4.9.2 from the following.

Theorem 4.9.3. The set \mathcal{L} of 1-Lipschitz functions f on U_m which satisfies $|f| \leq 1$ satisfy $\gamma_2(\mathcal{L}, d_2) \leq Lm^{3/4}$.

Here of course \mathcal{L} is seen as a subset of $L^2(U_m, \theta_m)$. The proof of Theorem 4.9.3 will use expansion of the elements of \mathcal{L} on a suitable basis. Using the same method as in Lemma 4.5.12 one can assume furthermore that the functions of \mathcal{L} are zero on $\{0\} \times T_m$ and $\{1\} \times T_m$. For $0 \leq n \leq m$ we consider the natural partition \mathcal{C}_n of T_m into 2^n sets obtained by fixing the first n coordinates of $t \in T_m$. Denoting by μ_m the uniform measure on T_m , for $C \in \mathcal{C}_n$ we have $\mu_m(C) = 2^{-n}$. A set $C \in \mathcal{C}_n$ with n < m is the union of two sets C_1 and C_2 in \mathcal{C}_{n+1} . We denote by h_C a function on T_n which equals $2^{n/2}$ on one of these sets and $-2^{n/2}$ on the other. Consider also the function h_{\emptyset} on T_m , constant equal to 1. In this manner we obtain an orthogonal basis (h_C) of $L^2(T_m, \mu_m)$. For $f \in \mathcal{L}$ we consider the coefficients of f on this basis.

²¹ It should be obvious what is meant by "uniform probability on U".

 $^{^{22}}$ By following the scheme of proof of (4.43).

$$a_{p,C}(f) := \int_{U_m} \exp(2ip\pi x) h_C(t) f(x,t) \mathrm{d}x \mathrm{d}\mu_m(t) \; .$$

There and always, $p \in \mathbb{Z}$, $n \ge 0, C \in \mathcal{C}_n$ or $p \in \mathbb{Z}$ and $C = \emptyset$. We will lighten notation by writing simply $\sum_{p,C}$ sums over all possible values of (p, C) as above.

Exercise 4.9.4. (a) Prove that

$$\sum_{p,C} p^2 |a_{p,C}|^2 \le L .$$
(4.129)

Hint: just use that $|\partial f/\partial x| \leq 1$.

(b) Prove that for each n and each $C \in \mathcal{C}_n$ we have

$$\sum_{p \in \mathbb{Z}} |a_{p,C}|^2 \le L 2^{-3n} . \tag{4.130}$$

Hint: prove that $\left|\int h_C(t)f(x,t)d\mu_m(t)\right| \leq L2^{-3n/2}$.

We have just shown that \mathcal{L} is isometric to a subset of the set \mathcal{A} of sequences $(a_{p,C})$ which satisfy (4.129) and (4.130).

Exercise 4.9.5. We will now show that $\gamma_2(\mathcal{A}) \leq Lm^{3/4}$. (a) Prove that \mathcal{A} is contained in an ellipsoid of the type

$$\mathcal{E} = \left\{ (a_{c,C}) \; ; \; \sum_{p,C} \alpha_{p,C}^2 |a_{p,C}|^2 \le 1 \right\}$$

where $\alpha_{p,C}^2 = (p^2 + 2^{2n}/m)/L$ if $C \in C_n, n \ge 0$ and $\alpha_{p,\emptyset}^2 = p^2/L$. (b) Conclude using (2.155). (The reader must be careful for the unfortunate clash of notation.)

The goal of the next exercise is to prove the lower bound in Theorem 4.9.2. This lower bound is obtained by a non-trivial twist on the proof of the lower bound for the AKT theorem, so you must fully master that argument to have a chance.

Exercise 4.9.6. Let us recall the functions $f_{q,\ell}$ of (4.82) where we take $r \simeq (\log N)/100$. For $n \ge 0$ we still consider the natural partition \mathcal{C}_n of T into 2^n sets obtained by fixing the first n coordinates of $t \in T$. We consider an integer p with $2^{-p} \simeq 1/\sqrt{r}$. For each q, each $\ell \le 2^q$ and $C \in \mathcal{C}_{q+p}$ we consider the function $f_{q,\ell,C}$ on U given by $f_{q,\ell,C}(x,t) = 2^{-p-20}f_{q,\ell}(x)\mathbf{1}_C(t)$. We consider functions of the type $f_q = \sum_{\ell \le 2^q, C \in \mathcal{C}_{q+p}} z_{q,\ell,C}f_{q,\ell,C}$ where $z_{q,\ell,C} \in \{0, 1, -1\}$. Copy the proof of the lower bound of the AKT theorem to prove that with high probability one can construct these functions such that $\sum_{k \le q} f_k$ is 1-Lipschitz and for each $q, \sum_{i \le N} (f_q(X_i) - \int f_q d\theta) \ge \sqrt{N}/(Lr^{1/4})$, where X_i are i.i.d uniform on U and θ is the uniform probability measure on U. Summation over $q \le r$ yields the desired result.

While making futile attempts in the direction of Theorem 4.9.2, arose further questions which we cannot answer. We describe one of these now. We recall the functionals $\gamma_{\alpha,\beta}$ of (4.5) and the uniform measure μ_m on T_m .

Question 4.9.7 Is it true that for any metric space (T, d) the space U of 1-Lipschitz maps f from T_m to T, provided with the distance D given by $D(f, f')^2 = \int_{T_m} d(f(s), f'(s))^2 d\mu_m(s)$ satisfies $\gamma_2(U, D) \leq Lm^{3/4}\gamma_{1,2}(T)$?

The motivation for this result is that if T is the set of Lipschitz functions on [0, 1] then $\gamma_{1,2}(T) \leq L$ (using Fourier transform to compare with an ellipsoids) and with minimal effort this would provide an alternate and more conceptual proof for the upper bound of Theorem 4.9.2.

Exercise 4.9.8. In the setting of Question 4.9.7 assume that $e_n(T,d) \leq 2^{-n}$. Prove that $e_{2n}(U,D) \leq L2^{-n}$ (and better if n > m). Conclude that $\sum_{n>0} 2^{n/2} e_n(U,D) \leq Lm$. Prove that $\gamma_2(U,D) \leq Lm\gamma_{1,2}(T)$.

Key ideas to remember

- Ellipsoids in a Hilbert space are in a sense smaller than their entropy numbers indicate. This is true more generally for sufficiently convex sets in a Banach space. This phenomenon explains the fractional powers of logarithms occurring in the most famous matching theorems.
- The size of ellipsoids in sometimes accurately described by using proper generalizations $\gamma_{\alpha,\beta}(T,d)$ of the basic functional $\gamma_2(T,d)$.
- Matching theorems are typically proved through a discrepancy bound, which evaluates the supremum of the empirical process over a class \mathcal{F} of functions.
- Bernstein's inequality is a convenient tool to prove discrepancy bounds. It involves the control of \mathcal{F} both for the L^2 and the supremum distance.
- Using two different distances reveals the power of approaching chaining through sequences of partitions.

4.10 Notes and Comments

The original proof of the Leighton-Shor theorem amounts basically to perform by hand a kind of generic chaining in this highly non-trivial case, an incredible tour de force.²³ A first attempt was made in [124] to relate (an important consequence of) the Leighton-Shor theorem to general methods for bounding stochastic processes, but runs into technical complications. Coffman and Shor [36] then introduced the use of Fourier transforms and brought to light the role of ellipsoids, after which it became clear that the structure

²³ There is a simple explanation as to why this was possible: as you can check through Wikipedia, both authors are geniuses.

of these ellipsoids plays a central part in these matching results, a point of view systematically expounded in [153].

Chapter 17 is a continuation of the present chapter. The more difficult material it contains is presented later for fear of scaring readers at this early stage. A notable feature of the result presented there is that ellipsoids do not suffice, a considerable source of complication. The material of Appendix A is closely related to the Leighton-Shor theorem.

The original results of [5] are proved using an interesting technique called the *transportation method*. A version of this method, which avoids many of the technical difficulties of the original approach. is presented in [173]. With the notation of Theorem 4.5.1, it is proved in [173] (a stronger version of the fact) that with probability $\geq 9/10$ one has

$$\inf_{\pi} \frac{1}{N} \sum_{i \le N} \exp\left(\frac{Nd(X_i, Y_{\pi(i)})^2}{L \log N}\right) \le 2.$$
(4.131)

Since exp $x \ge x$, (4.131) implies that $\sum_{i \le N} d(X_i, Y_{\pi(i)})^2 \le L \log N$ and hence using the Cauchy-Schwarz inequality $\sum_{i \le N} d(X_i, Y_{\pi(i)}) \le L \sqrt{N \log N}$. Moreover (4.131) also implies $\max_{i \le N} d(X_i, Y_{\pi(i)}) \le L \log N / \sqrt{N}$. This unfortunately fails to bring a positive answer to Question 4.9.1.

For results about matching for unbounded distributions, see the work of J. Yukich [186] as well as the non-standard results of [172].