Fifty years ago, a theorem by Xavier Fernique and some more old memories

(08-12-2024)

Contents

Links in the text are *red*, the color *blue* is just for emphasis

Introduction

Our main —but not only— objective in these Notes is to present a remarkable result of Xavier Fernique about Gaussian processes, from the year 1974 [Fer_3]. We shall go on with some of Talagrand's further developments on the same theme, and in the last section, we shall also recall a few facts from the '70s, more or less related to Gaussian processes, in many cases less, rather than more. I will try my best to keep the exposition as self-contained as possible, and free from any recent discovery. Also, I won't be able to resist giving quite often much more details than necessary for a decent reader to fully understand what's going on.

Let us begin with a short presentation of the background for Fernique's result. A Gaussian process $(X_t)_{t\in T}$ consists of a collection of Gaussian random variables X_t, indexed by a non-empty set T and belonging to a *Gaussian space*, that is to say, a linear subspace of $L^2(\Omega, \mathrm{P})$ all of whose elements are Gaussian random variables, where (Ω, P) is some probability space. We shall restrict ourselves to *centered* processes, namely, the case when

$$\mathrm{E}\, X_t = 0 \quad \text{for all} \quad t \in T.$$

The index set T will be equipped with the L^2-metric given by the distance in $L^2(\Omega, P)$ of the corresponding random variables,

$$d(s,t)^2 = \mathrm{E}(X_s - X_t)^2, \qquad s, t \in T,$$

and we shall consider *open* balls in T of radius $r > 0$ for that metric,

$$B(s,r) = \{t \in T : d(t,s) < r\}, \qquad s \in T.$$

It is well known that some *entropy conditions* for that metric on T allow one to control the supremum $\sup_{t\in T} X_t$ of the process (1): given $\varepsilon > 0$, let $\mathcal{N}(T, \varepsilon)$ denote the minimal number of open balls of radius ε needed to cover T. We assume that T is bounded for the metric d and (2) we let Δ be the diameter of T. The so-called *Dudley's integral* is defined by

$$I_D(T) = \int_0^\Delta \sqrt{\ln \mathcal{N}(T, \varepsilon)}\, \mathrm{d}\varepsilon.$$

Notice that $\ln \mathcal{N}(T, \varepsilon) = 0$ when $\varepsilon > \Delta$, because $\mathcal{N}(T, \varepsilon) = 1$ is that case (one ball of such a radius $\varepsilon > \Delta$ is enough). If the Dudley integral satisfies

$$I_D(T) < \infty, \quad \text{it follows that} \quad \mathrm{E}\Bigl(\sup_{t\in T} |X_t|\Bigr) < \infty.$$

The domination by the Dudley integral of the expectation (the integral) of the supremum of a Gaussian process was often attributed to Richard Dudley [Dud_1], who himself, shortly after Vladimir Sudakov died in 2016, pointed out [Dud_2] that Sudakov was actually the one to credit for that result (a result that, in essence, is far from being the hardest point in what will be recalled here from Sudakov's and others' works).

The result of Fernique is that, under some "group invariance" of the process, one can prove the reverse implication. Perhaps surprisingly for whom is far from my own domain of interests, Fernique's theorem was one crucial piece for a theorem of Gilles Pisier [Pis_2] on lacunary trigonometrical series, namely, the characterization of Sidon sets $\Lambda \subset \mathbb{Z}$ by the rate of growth as p tends to ∞ of the L^p-norm of the functions with spectrum in Λ. This rate of growth of order $\sqrt{p}$ for Sidon sets was established by Walter

Rudin [Rud]. Conversely, Pisier proves that: *if there exists a constant C such that for every trigonometric polynomial P with spectrum in $\Lambda \subset \mathbb{N}$, we have*

$$\|P\|_{L^p(\mathbb{T},m)} \leqslant C\sqrt{p} \quad \textit{for every} \quad p \geqslant 2,$$

then Λ is a Sidon set, which means that for some $c > 0$, one has

$$\|P\|_{C(\mathbb{T})} \geqslant c \sum_{n\in\Lambda} |\widehat{P}(n)|$$

for all trigonometric polynomials $P(t) = \sum_{n\in\Lambda} \widehat{P}(n)\,\mathrm{e}^{\mathrm{i}nt}$ with spectrum in the set Λ. Here, m is the invariant measure on the torus $\mathbb{T}$ and $C(\mathbb{T})$ is the space of (complex) continuous functions on $\mathbb{T}$, equipped with the maximum norm. In the proof of that result, the Fernique theorem is applied to the complex valued Gaussian process indexed by $t \in \mathbb{T}$ and defined by

$$X_t(\omega) = \sum_{n\in\Lambda} \widehat{P}(n)\, g_n(\omega)\,\mathrm{e}^{\mathrm{i}nt},$$

where (g_n) is a sequence of independent $\mathrm{N}(0,1)$ Gaussian random variables.

The first section gives a few basic facts about Gaussian variables, especially about the maximum of a finite number of Gaussian variables, and culminates at the comparison result of Slepian–Sudakov, stated here without proof. The second section presents a proof of the Fernique theorem. It is a bit embarrassing to admit that apart from the Sudakov–Slepian comparison result, the most advanced mathematical tools present in our two first sections are the logarithmical function $\ln x$, the generalized Riemann integral $\int_0^\infty \mathrm{d}x$, comparing an integral to a series, and integrating by parts... The third section concerns Gaussian processes that are no longer assumed to be stationary; the results are due to Talagrand, who first proved the existence of a *majorizing measure* (3) and introduced later the notion of generic chaining, see Theorem 2.

The section 4, the last one, is loosely connected to the preceding section 3. Being in that small and nonincreasing number of the Banach space enthusiasts from the '70s still in a position to write one more paper, I felt that I had to tell the younger about some of the names, facts and results of that remote epoch: Laurent Schwartz and the radonifying maps, Albrecht Pietsch and the p-summing operators, the 1-summing version of the Grothendieck inequality due to Joram Lindenstrauss and Aleksander Pełczyński, among many more people and results mentioned in Section 4.4.

1. Preliminaries

We start slowly, with elementary observations on the discretization of the Dudley integral, then with the definition of Gaussian random variables, and we proceed calmly toward the comparisons of Gaussian processes due to David Slepian and Sudakov.

1.1. *Discrete versions of the Dudley integral*

We may discretize the Dudley integral, by choosing first a radius $r_0 > \Delta$, then letting for example $r_i = 3^{-i} r_0$ for every integer $i \geqslant 0$ and introducing the series

$$\Sigma_1(T) := \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(T, r_{i+1})}. \tag{1}$$

For every $i \geqslant 0$, we see that

$$\int_{r_{i+1}}^{r_i} \sqrt{\ln \mathcal{N}(T,\varepsilon)}\,\mathrm{d}\varepsilon \leqslant r_i \sqrt{\ln \mathcal{N}(T, r_{i+1})},$$

and thus

$$I_D(T) = \sum_{i=0}^{\infty} \int_{r_{i+1}}^{r_i} \sqrt{\ln \mathcal{N}(T,\varepsilon)}\, \mathrm{d}\varepsilon \leqslant \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(T, r_{i+1})} = \Sigma_1(T).$$

The latter series $\Sigma_1(T)$ is of the sort that will appear several times later. Let us now write simply $\mathcal{N}(r_{i+1})$ instad of $\mathcal{N}(T, r_{i+1})$. With the present choice $r_{i+1} = r_i/3$, we also have conversely that

$$\int_{r_{i+1}}^{r_i} \sqrt{\ln \mathcal{N}(\varepsilon)}\, \mathrm{d}\varepsilon \geqslant \frac{2}{3}\, r_i \sqrt{\ln \mathcal{N}(r_i)}, \ \text{ so } \ I_D(T) \geqslant \frac{2}{9} \sum_{i=1}^{\infty} r_{i-1} \sqrt{\ln \mathcal{N}(r_i)} = \frac{2}{9}\, \Sigma_1(T).$$

A similar reverse inequality holds true for any choice where $r_0 > \Delta$ and $r_i = a^i r_0$, with $0 < a < 1$, or simply a choice where $r_0 > \Delta$ and $r_i - r_{i+1} \geqslant c r_{i-1}$, with $0 < c < 1$.

We may obtain yet another equivalent form for the Dudley integral as a series, by a kind of "change of variable": instead of fixing the radius r and looking for the number of balls of that radius necessary to cover T, we fix the number N of balls and look for a radius such that we can cover T by N balls with that radius. This second series (2) will not be mentioned again until much later in the text, the reader can at first directly jump to the next section. In this change of variable $i \leftrightarrow k$, we think of k and i to be linked by

$$b^k \sim \ln \mathcal{N}(r_i),$$

for some fixed real number $b > 1$.

To be more specific, choose b such that $b^{-1} < \ln 2$, suppose that $\Delta < r_0 < 3\Delta/2$ and let again $r_{i+1} = r_i/3$ for $i \geqslant 0$. With this choice, we have $\mathcal{N}(r_0) = 1$ and we see that $2r_1 = 2r_0/3 < \Delta$, so that $B(t, r_1) \neq T$ for any $t \in T$ and thus $\mathcal{N}(r_1) \geqslant 2$. For simplicity, assume that $\mathcal{N}(\varepsilon)$ is unbounded as $\varepsilon \to 0$. For every integer $k \geqslant 0$, let $i(k)$ be the smallest $i \geqslant 0$ for which we have $b^{k-1} < \ln \mathcal{N}(r_{i+1})$; when $k = 0$ for example, we obtain that $i(0) = 0$ because we know that $\ln \mathcal{N}(r_0) = 0 < b^{-1} < \ln 2 \leqslant \ln \mathcal{N}(r_1)$. Consider

$$\Sigma_2(T) = \sum_{k=0}^{\infty} r_{i(k)}\, b^{k/2}. \tag{2}$$

We will see that, up to a multiplicative constant depending only upon b, the value $\Sigma_2(T)$ is equivalent to $\Sigma_1(T)$, hence also equivalent to the Dudley integral. Let $I \subset \mathbb{N}$ be the set of integers $i(k)$, $k \geqslant 0$. For every $i \in I$, let $k(i)$ be the largest k such that $i(k) = i$. Adding geometric progressions, we get

$$\Sigma_2(T) = \sum_{i \in I} \Bigl(\sum_{i(k)=i} r_i\, b^{k/2} \Bigr) \leqslant \sum_{i \in I} r_i\, \frac{\bigl(\sqrt{b}\bigr)^{k(i)+1} - 1}{\sqrt{b} - 1} < \frac{b^{1/2}}{b^{1/2} - 1} \sum_{i \in I} r_i\, b^{k(i)/2}.$$

When $i(k) = i$ we have $b^{k-1} < \ln \mathcal{N}(r_{i+1})$, thus $b^{k(i)} < b \ln \mathcal{N}(r_{i+1})$ and we go on with

$$\Sigma_2(T) < \frac{b}{b^{1/2} - 1} \sum_{i \in I} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} \leqslant \frac{b}{b^{1/2} - 1} \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} = \frac{b}{b^{1/2} - 1}\, \Sigma_1(T).$$

In the other direction, consider for each $k \geqslant 0$ the (perhaps empty) interval of integers

$$I_k = \bigl\{ i \in \mathbb{N} : b^{k-1} < \ln \mathcal{N}(r_{i+1}) \leqslant b^k \bigr\}.$$

These intervals cover $\mathbb{N}$, because $b^{-1} < \ln \mathcal{N}(r_1)$, so that $i = 0$ belongs to some I_k, and then every $i \geqslant 0$ does as well. Let $K \subset \mathbb{N}$ denote the set of k such that I_k is not empty.

When $k \in K$, we see that $\min I_k = i(k)$, and we observe that

$$\sum_{i \in I_k} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} \leqslant \Bigl(\sum_{i \in I_k} r_i\Bigr) b^{k/2} < \frac{3}{2} r_{i(k)} \, b^{k/2}.$$

Then

$$\Sigma_1(T) = \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} = \sum_{k \in K} \sum_{i \in I_k} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} < \frac{3}{2} \sum_{k \in K} r_{i(k)} \, b^{k/2} \leqslant \frac{3}{2} \Sigma_2(T).$$

By the definition of $i(k)$, we know that $\ln \mathcal{N}(r_{i(k)}) \leqslant b^{k-1} < \ln \mathcal{N}(r_{i(k)+1})$. Hence, for every integer $k \geqslant 0$, there exists a finite set $T_k \subset T$ such that $\ln |T_k| \leqslant b^{k-1}$ and such that the balls of radius $\rho_k = r_{i(k)}$ centered at the points of T_k cover T; for $k = 0$, we have $|T_0| \leqslant \exp(b^{-1}) < 2$ thus $|T_0| = 1$, and $\rho_0 = r_{i(0)} = r_0$. If $\Sigma_2(T)$ is finite, we may summarize the situation as follows:

$$(3) \qquad |T_0| = 1; \ \ \ln |T_k| < b^k \ \text{and} \ T = \bigcup_{s \in T_k} B(s, \rho_k) \ \text{for all} \ k \geqslant 0; \quad \sum_{k=0}^{\infty} \rho_k \, b^{k/2} < \infty.$$

1.2. *Gaussian random variables*

This section is completely elementary and contains some basic definitions, together with the standard Lemmas 1 and 2, given here with explicit constants resulting sometimes from tedious calculations. A random variable X defined on a probability space (Ω, P) is said to be a $\mathrm{N}(0,1)$ *Gaussian random variable* when for every $x \in \mathbb{R}$, we have

$$\mathrm{P}(X > x) = \int_x^{\infty} \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}}.$$

We then get for the *expectation* $\mathrm{E}\, X$ and the *variance* $\mathrm{Var}\, X$ of X the values

$$\mathrm{E}\, X = \int_{\mathbb{R}} u \, \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} = 0,$$

and

$$\mathrm{Var}\, X := \mathrm{E}(X - \mathrm{E}\, X)^2 = \mathrm{E}\, X^2 = \int_{\mathbb{R}} u^2 \, \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} = 1.$$

So, the "0" and the "1" in $\mathrm{N}(0,1)$ refer to the expectation and variance of X.

Let $x > 0$ and observe that

$$(4) \qquad \mathrm{P}(X > x) = \int_x^{\infty} \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} < \int_x^{\infty} \frac{u}{x} \, \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} = \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}}.$$

This estimate is essentially correct, for example because

$$\int_x^{\infty} \mathrm{e}^{-u^2/2} \, \mathrm{d}u \geqslant \int_x^{x+1/x} \mathrm{e}^{-u^2/2} \, \mathrm{d}u \geqslant \frac{1}{x} \, \mathrm{e}^{-(x+x^{-1})^2/2} = \frac{\mathrm{e}^{-(x^2/2)-1-(x^{-2}/2)}}{x},$$

so that we can infer that

$$(5) \qquad x \geqslant 1 \quad \Rightarrow \quad \mathrm{P}(X > x) \geqslant \mathrm{e}^{-3/2} \, \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}}.$$

When x goes to $+\infty$ we can do better, integrating by parts and writing for $x > 0$ the equalities

$$\int_x^{\infty} \mathrm{e}^{-u^2/2} \, \mathrm{d}u = \int_x^{\infty} \frac{1}{u} \, (u \, \mathrm{e}^{-u^2/2}) \, \mathrm{d}u = \frac{\mathrm{e}^{-x^2/2}}{x} - \int_x^{\infty} \frac{1}{u^2} \, \mathrm{e}^{-u^2/2} \, \mathrm{d}u$$

and, repeating the trick,

$$\int_x^\infty \frac{1}{u^2}\,\mathrm{e}^{-u^2/2}\,\mathrm{d}u = \int_x^\infty \frac{1}{u^3}\,(u\,\mathrm{e}^{-u^2/2})\,\mathrm{d}u = \frac{\mathrm{e}^{-x^2/2}}{x^3} - \int_x^\infty \frac{3}{u^4}\,\mathrm{e}^{-u^2/2}\,\mathrm{d}u.$$

We obtain that

$$x > 0 \quad\Rightarrow\quad \mathrm{P}(X > x) = \int_x^\infty \mathrm{e}^{-u^2/2}\,\frac{\mathrm{d}u}{\sqrt{2\pi}} > \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\Bigl(\frac{1}{x} - \frac{1}{x^3}\Bigr) = \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}}\Bigl(1 - \frac{1}{x^2}\Bigr), \tag{6}$$

a certainly uninteresting assertion when $0 < x \leqslant 1$. It is easy to guess how to go on and produce an asymptotic expansion of $\mathrm{P}(X > x)$ in terms of the variable $x > 1$.

When g is a $\mathrm{N}(0,1)$ Gaussian random variable and $u > 0$, one has

$$P(|g| > u) \leqslant \mathrm{e}^{-u^2/2}\,. \tag{7}$$

Indeed, we have to prove that $\mathrm{P}(g > x) \leqslant \frac{1}{2}\,\mathrm{e}^{-x^2/2}$ when $x \geqslant 0$. We know already this inequality for $x \geqslant \sqrt{2/\pi}$ by (4), and the remaining values of x are obtained by checking the sign of the derivative of the function $f : x \mapsto \mathrm{e}^{-x^2/2} - 2\,\mathrm{P}(g > x)$ on the segment $[0, \sqrt{2/\pi}]$, namely, the easily understandable sign of

$$f'(x) = \bigl(\sqrt{2/\pi} - x\bigr)\,\mathrm{e}^{-x^2/2}\,.$$

Inequality (7) holds *a fortiori* for any centered Gaussian random variable Y having variance $\leqslant 1$: we can write $Y = \theta\,g$ with $\theta = (\mathrm{E}\,Y^2)^{1/2} \in [0,1]$ and with g being a $\mathrm{N}(0,1)$ variable, therefore

$$\mathrm{E}\,Y^2 \leqslant 1 \quad\Rightarrow\quad \mathrm{P}(|Y| > u) \leqslant \mathrm{P}(|g| > u) \leqslant \mathrm{e}^{-u^2/2}\,.$$

Lemma 1. *If $N \geqslant 2$ and if $g_1, g_2, \ldots, g_N$ are $\mathrm{N}(0,1)$ Gaussian random variables, independent or not, then*

$$\mathrm{E}\Bigl(\max_{1\leqslant i\leqslant N} |g_i|\Bigr) \leqslant \sqrt{2\ln N} + \frac{1}{\sqrt{2\ln N}}.$$

It follows that for every $N \geqslant 1$, we have

$$\mathrm{E}\Bigl(\max_{1\leqslant i\leqslant N} |g_i|\Bigr) \leqslant 2\sqrt{\ln(N+1)}. \tag{8}$$

If $\sigma \geqslant 0$ and if $X_1, X_2, \ldots, X_N$ are centered Gaussian random variables, then

$$\max_{1\leqslant i\leqslant N} \mathrm{E}\,X_i^2 \leqslant \sigma^2 \quad\Rightarrow\quad \mathrm{E}\Bigl(\max_{1\leqslant i\leqslant N} |X_i|\Bigr) \leqslant 2\sigma\sqrt{\ln(N+1)}. \tag{9}$$

Proof. Let $N \geqslant 2$, let $G_N^* = \max_{1\leqslant i\leqslant N} |g_i|$ and $x > 0$; by (7) we know that

$$P(|g_1| > x) = P(|g_2| > x) = \cdots = P(|g_N| > x) \leqslant \mathrm{e}^{-x^2/2},$$

and by the union bound inequality, we get

$$P(G_N^* > x) \leqslant N\,\mathrm{e}^{-x^2/2}\,.$$

Observe that if $x_0 = \sqrt{2\ln N}$, then $N\,\mathrm{e}^{-x_0^2/2} = 1$, and $x_0 \geqslant \sqrt{2\ln 2} > 0$. It follows that

$$\begin{aligned}\mathrm{E}\,G_N^* &= \int_0^\infty \mathrm{P}(G_N^* > x)\,\mathrm{d}x \leqslant x_0 + \int_{x_0}^\infty N\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x \\ &\leqslant x_0 + N\int_{x_0}^\infty \frac{x}{x_0}\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x = x_0 + N\,\frac{\mathrm{e}^{-x_0^2/2}}{x_0} = x_0 + \frac{1}{x_0}\cdot\end{aligned}$$

For the second inequality (8), we have first $\mathrm{E}\, G_1^* = \mathrm{E}\,|g_1| = \sqrt{2/\pi} < 1 < 2\sqrt{\ln 2}$, which covers the case $N = 1$. When $N \geqslant 2$, we use

$$\sqrt{2\ln y} + \frac{1}{\sqrt{2\ln y}} < 2\sqrt{\ln(y+1)} \quad \text{when} \quad y \geqslant 2,$$

or equivalently the fact that

$$y \geqslant 2 \quad \Rightarrow \quad f(y) = 4\ln(y+1) - 2\ln y - 2 - \frac{1}{2\ln y} > 0,$$

which is true because

$$f'(y) = \frac{4}{y+1} - \frac{2}{y} + \frac{1}{2y(\ln y)^2} > \frac{2y-2}{y(y+1)}$$

is > 0 when $y \geqslant 2$, and because $2\ln 2 > 1$ yields $f(2) > \ln(81/4) - 3 > 0$.

If $\sigma \geqslant 0$ and if $X_1, X_2, \ldots, X_N$ are centered Gaussian such that $\mathrm{Var}\, X_i \leqslant \sigma^2$, the sequence has the same distribution as a sequence $\sigma_1 g_1, \sigma_2 g_2, \ldots, \sigma_N g_N$ with $0 \leqslant \sigma_i \leqslant \sigma$ and g_i a $\mathrm{N}(0,1)$ variable, therefore

$$\mathrm{E} \max_{1 \leqslant i \leqslant N} |X_i| = \mathrm{E} \max_{1 \leqslant i \leqslant N} |\sigma_i g_i| \leqslant \sigma\, \mathrm{E} \max_{1 \leqslant i \leqslant N} |g_i|$$

and the result (9) follows. □

Inequality (8) applies equally well to *sub-gaussian variables* properly normalized, for example sequences $X_1, X_2, \ldots, X_N$ such that for each i and every $x > 0$, we have

$$\mathrm{P}(|X_i| > x) \leqslant \mathrm{e}^{-x^2/2},$$

as this is all that was used in the above proof. However, it would be more realistic to say that a sub-gaussian variable X is *normalized* when for every $x > 0$, we have

$$\mathrm{P}(|X| > x) \leqslant 2\, \mathrm{e}^{-x^2/2}\,. \tag{10}$$

A tiny change in the above proof leads to a bound $\sqrt{2\ln N} + 2/\sqrt{2\ln N}$ for the expectation $\mathrm{E}\, X_N^*$ of the maximum X_N^* of $N \geqslant 2$ sub-gaussian variables normalized by (10), writing now

$$\mathrm{E}\, X_N^* \leqslant x_0 + 2N \int_{x_0}^{\infty} \frac{x}{x_0}\, \mathrm{e}^{-x^2/2}\, \mathrm{d}x = x_0 + \frac{2}{x_0}, \quad \text{with} \quad x_0 = \sqrt{2\ln N}.$$

Lemma 2. *If $N \geqslant 1$ and if $g_1, g_2, \ldots, g_N$ are independent $\mathrm{N}(0,1)$ Gaussian random variables, one has*

$$\mathrm{E}\Bigl(\max_{1 \leqslant i \leqslant N} g_i\Bigr) \geqslant \frac{1}{2}\sqrt{\ln N}. \tag{11}$$

If $\sigma \geqslant 0$ and if $X_1, X_2, \ldots, X_N$ are centered independent Gaussian random variables, then

$$\min_{1 \leqslant i \leqslant N} \mathrm{E}\, X_i^2 \geqslant \sigma^2 \;\Rightarrow\; \mathrm{E}\Bigl(\max_{1 \leqslant i \leqslant N} X_i\Bigr) \geqslant \frac{\sigma}{2}\sqrt{\ln N}. \tag{12}$$

Furthermore, when the (g_i) are as above and when N tends to ∞, one has

$$\frac{\mathrm{E}\Bigl(\max_{1 \leqslant i \leqslant N} g_i\Bigr)}{\sqrt{2\ln N}} \xrightarrow[N]{} 1.$$

Proof. Let

$$g_N^* = \max_{1 \leqslant i \leqslant N} g_i.$$

We have $\mathrm{E}\, g_1^* = \mathrm{E}\, g_1 = 0 \geqslant (1/2)\sqrt{\ln 1}$, it can be shown (4) easily that

$$\mathrm{E}\, g_2^* = 1/\sqrt{\pi}, \qquad \text{and one checks that} \quad 1/\sqrt{\pi} > (1/2)\sqrt{\ln 2}.$$

Slightly less easy are the facts that

$$\begin{aligned} &\mathrm{E}\, g_3^* = 3/(2\sqrt{\pi}), && \text{and} \quad 3/(2\sqrt{\pi}) > (1/2)\sqrt{\ln 3}, \\ &\mathrm{E}\, g_4^* = 6 \arctan(\sqrt{2})/\pi^{3/2} > 1, && \text{and} \quad 1 > (1/2)\sqrt{\ln 4}. \end{aligned}$$

Our last effort has been to establish that

$$\mathrm{E}\, g_5^* = \frac{15}{\pi^{3/2}} \Bigl(\frac{\pi}{3} - \arcsin\Bigl(\frac{1}{\sqrt{3}} \Bigr) \Bigr) > 1.162 > 0.635 > \frac{1}{2}\sqrt{\ln 5}.$$

Taking this for granted, let $x_0 = 2\, \mathrm{E}\, g_5^* > 2.324$. Clearly $\mathrm{E}\, g_N^*$ increases with N, so we need only consider values of $N > 5$ such that

$$\frac{1}{2}\sqrt{\ln N} > \mathrm{E}\, g_5^* = \frac{x_0}{2} > 1.162,$$

or $\ln N > x_0^2 > 5.4$ and $N > 221$. Hence, we shall restrict our study to integers N such that $\sqrt{\ln N} \geqslant x_0 > 1$. The function $y \mapsto y - \ln(2y)$ is increasing when $y > 1$, thus

$$\ln N - \ln(2 \ln N) \geqslant x_0^2 - \ln(2x_0^2).$$

Let $u = x_0^2 - \ln(2x_0^2)$. One can check that $u > 3$. Let

$$s = \sqrt{2 \ln N - \ln(2 \ln N) - u}.$$

We see that

$$s \geqslant \sqrt{\ln N + x_0^2 - \ln(2x_0^2) - u} = \sqrt{\ln N}.$$

We have

$$x_0 \leqslant \sqrt{\ln N} \leqslant s \leqslant \sqrt{2 \ln N}. \tag{13}$$

Next we use (6), then we notice that $x_0^2 > 5$ and we obtain

$$\begin{aligned} \mathrm{P}(g_1 > s) &\geqslant \frac{\mathrm{e}^{-s^2/2}}{s\sqrt{2\pi}} \Bigl(1 - \frac{1}{s^2}\Bigr) \geqslant \frac{\mathrm{e}^{-s^2/2}}{\sqrt{2 \ln N}\sqrt{2\pi}} \Bigl(1 - \frac{1}{x_0^2}\Bigr) \\ &= \frac{1}{N} \frac{\mathrm{e}^{u/2}}{\sqrt{2\pi}} \Bigl(1 - \frac{1}{x_0^2}\Bigr) > \frac{1}{N} \frac{\mathrm{e}^{u/2}}{\sqrt{2\pi}} \frac{4}{5} > \frac{1.40}{N}. \end{aligned}$$

It follows that

$$\mathrm{P}(g_N^* < s) = \mathrm{P}(g_1 < s)^N \leqslant (1 - 1.4/N)^N \leqslant \mathrm{e}^{-1.4} < 1/4. \tag{14}$$

We need to take care of the rare but possible negative values of g_N^*. Let

$$\nu_N(\omega) = \min\bigl(g_N^*(\omega), 0\bigr) \leqslant 0, \quad \omega \in \Omega.$$

We see that

$$\nu_N \geqslant \min(g_N, 0)\, \mathbf{1}_{\{g_{N-1}^* < 0\}},$$

thus by independence

$$\mathrm{E}\, \nu_N \geqslant \Bigl(\mathrm{E} \min(g_N, 0)\Bigr) \Bigl(\mathrm{E}\, \mathbf{1}_{\{g_{N-1}^* < 0\}}\Bigr) = -\frac{1}{\sqrt{2\pi}}\, 2^{-N+1} > -2^{-N}.$$

We know that $N > 221$ and $x_0 > 1$, therefore

$$2^{-N} < 2^{-221} x_0 \leqslant 2^{-221} \sqrt{\ln N}.$$

Finally, using (13) and (14),

$$\begin{aligned}\mathrm{E}\, g_N^* &\geqslant s\,\mathrm{P}(g_N^* > s) + \mathrm{E}\,\nu_N \geqslant (1-\mathrm{e}^{-1.4})\sqrt{\ln N} - 2^{-N} \\ &> \Bigl(\frac{3}{4} - 2^{-221}\Bigr)\sqrt{\ln N} > \frac{1}{2}\sqrt{\ln N}.\end{aligned}$$

The claim about non $\mathrm{N}(0,1)$ variables is proved as before. Let us explain rapidly the last sentence. Lemma 1 implies that the limsup of the quotient $\mathrm{E}\, g_N^*/\sqrt{\ln N}$ is $\leqslant \sqrt{2}$. Now, fix u rather large and $\varepsilon > 0$ small. When N is large enough, we certainly have

$$\varepsilon \ln N - \ln(2\ln N) - u \geqslant 0.$$

Then

$$\sqrt{2\ln N} > s = \sqrt{2\ln N - \ln(2\ln N) - u} \geqslant \sqrt{(2-\varepsilon)\ln N}.$$

Because $s > 1$ for N large we may use (5) and write

$$\mathrm{P}(g_1 > s) \geqslant \mathrm{e}^{-3/2}\,\frac{\mathrm{e}^{-s^2/2}}{s\sqrt{2\pi}} \geqslant \frac{\mathrm{e}^{-(s^2+3)/2}}{\sqrt{2\ln N}\sqrt{2\pi}} = \frac{1}{N}\,\frac{\mathrm{e}^{(u-3)/2}}{\sqrt{2\pi}} =: \frac{\alpha}{N},$$

but now α can be made as large as we wish, and

$$\mathrm{E}\, g_N^* \geqslant s\,\mathrm{P}(g_N^* > s) - 2^{-N} \geqslant (1 - \mathrm{e}^{-\alpha} - 2^{-N})\sqrt{(2-\varepsilon)\ln N}$$

proves our claim. $\square$

Numerical experiments suggest that the quotient $\mathrm{E}\, g_N^*/\sqrt{\ln N}$ is actually increasing with $N \geqslant 2$. If this were true, the correct constant $c > 1/2$ in the inequality (11) for all $N \geqslant 2$ would simply be the value at $N = 2$, namely $c = 1/\sqrt{\pi \ln 2} > 0.67 > 2/3$.

1.3. *The comparison result*

The next comparison result plays a major rôle in what follows.

Proposition 1. *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two centered Gaussian processes indexed by the same set T. If we have*

$$\mathrm{E}(Y_s - Y_t)^2 \leqslant \mathrm{E}(X_s - X_t)^2$$

for all $s, t \in T$, we can conclude that

$$\mathrm{E}\Bigl(\sup_{t\in T} Y_t\Bigr) \leqslant \mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr).$$

The centering is necessary here, as the simple example $Y_t = 1 + X_t$ shows.

The proof of this result is not as elementary as what we have seen so far, it will not be given here (5). A first comparison result goes back to Slepian [Slep] in 1962, and it applies to comparing the *distributions* of the suprema: if in addition to the hypothesis of Proposition 1 one adds that $\mathrm{E}\, Y_t^2 = \mathrm{E}\, X_t^2$ for every $t \in T$, then for every x real one has

$$\mathrm{P}\Bigl(\sup_{t\in T} Y_t > x\Bigr) \leqslant \mathrm{P}\Bigl(\sup_{t\in T} X_t > x\Bigr).$$

Under this additional assumption, we see that Slepian's lemma implies the conclusion of Proposition 1.

The comparison result in Proposition 1 was announced in a Note [Sud_1] without proof by Sudakov (see Theorem 2 there); a complete proof is available in Sudakov's book [Sud_2]. The result was also approached by Simone Chevet [Che_1], and given by Fernique [Fer_2], [Fer_3] (6).

A more general version of Proposition 1 is due to Yehoram Gordon [Gord], and deals with a mixture of min and max; a reasonably simple proof can be found in Chap. 8 of the book by Daniel Li and Hervé Queffélec [LiQu]. Gordon's result is extremely useful for estimating the invertibility of Gaussian random maps between finite dimensional normed spaces; indeed, finding an estimate of the norm of the inverse of a Gaussian random map $T_\omega : E \to F$ involves estimating the *inf* of norms $\|T_\omega(x)\|_F$ of the images $T_\omega(x)$ of norm one vectors $x \in E$, where each norm $\|T_\omega(x)\|_F$ is a *sup* of Gaussian random variables of the form $< y^*, T_\omega(x) >$, and the y^* are all the norm one linear functionals on the target space F.

The Gordon comparison theorem can also be used to give another proof of a celebrated lemma due to William Johnson and Lindenstrauss [JoLi], first proved using concentration of measure on the Euclidean sphere in high dimension ([7]).

1.3.1. *The Sudakov entropy bound*

Suppose that $\mathrm{E}\sup_{t\in T} X_t$ is finite. It follows from Proposition 1 that for every $\varepsilon > 0$, we can cover T with a finite number of balls of radius ε: indeed, if $t_1, t_2, \ldots, t_n$ in T have mutual distances larger than ε, we shall compare the processes $(X_{t_j})_{j=1}^n$ and $(Y_{t_j})_{j=1}^n$ where the Y_{t_j} are independent centered Gaussian random variables of variance $\varepsilon^2/2$. We have when $i \neq j$

$$\mathrm{E}(Y_{t_j} - Y_{t_i})^2 = \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2} \leqslant d(t_j, t_i)^2 = \mathrm{E}(X_{t_j} - X_{t_i})^2,$$

hence by Proposition 1 and Inequality (12),

$$\mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr) \geqslant \mathrm{E}\Bigl(\sup_{1\leqslant i\leqslant n} X_{t_i}\Bigr) \geqslant \mathrm{E}\Bigl(\sup_{1\leqslant i\leqslant n} Y_{t_i}\Bigr) \geqslant \frac{\varepsilon}{2\sqrt{2}}\sqrt{\ln n}.$$

This gives a bound on n, and thus

$$\mathcal{N}(T, \varepsilon) \leqslant \exp\biggl(\frac{8}{\varepsilon^2}\Bigl(\mathrm{E}\sup_{t\in T} X_t\Bigr)^2\biggr).$$

It follows that when the expectation of $\sup_{t\in T} X_t$ is finite, the closure in $L^2(\Omega, \mathrm{P})$ of the set of Gaussian variables $(X_t)_{t\in T}$ is a compact subset of L^2.

Given $\delta > 0$, we say that a subset $S \subset T$ is *δ-separated* when any two of its points are δ-far apart,

$$s_1, s_2 \in S,\ s_1 \neq s_2 \quad\Rightarrow\quad d(s_1, s_2) \geqslant \delta.$$

Given a subset $A \subset T$, we call ([8]) *δ-packing-net for A* any maximal δ-separated subset S of A; we shall shorten it as "*δ-p-net*". Maximality implies that no point of A can be added to the set S and keep it δ-separated: for every $a \in A$, there is some $s \in S$ such that $d(a, s) < \delta$, in other words,

$$A \subset \bigcup_{s\in S} B(s, \delta).$$

Suppose that S is a δ-p-net for the set T; by the preceding remark, it implies that the balls $B(s, \delta)$, for $s \in S$, cover T, and this shows that $\mathcal{N}(T, \delta) \leqslant |S|$. If we denote by $\mathcal{N}_*(T, \delta)$ the smallest cardinality of a δ-p-net for T and by $\mathcal{N}^*(T, \delta)$ the largest, it follows that

$$\mathcal{N}(T, \delta) \leqslant \mathcal{N}_*(T, \delta) \leqslant \mathcal{N}^*(T, \delta).$$

Conversely, suppose that balls $B(t_i, \delta/2)$, for $i = 1, 2, \ldots, N$, cover T. If s_1, s_2 are two points in the δ-p-net S and $s_1 \neq s_2$, these two points cannot belong to the same open ball $B(t_i, \delta/2)$ since $d(s_1, s_2) \geqslant \delta$. This yields that $|S| \leqslant N$, thus

$$\mathcal{N}^*(T, \delta) \leqslant \mathcal{N}(T, \delta/2).$$

These two inequalities show that we can use $\mathcal{N}^*$ (or $\mathcal{N}_*$) instead of $\mathcal{N}$ in the definition of the Dudley integral,

$$I_D(T) \leqslant \int_0^\Delta \sqrt{\ln \mathcal{N}^*(T, \varepsilon)}\, \mathrm{d}\varepsilon \leqslant 2\, I_D(T).$$

We will need to consider the supremum of the process when the index t ranges, not only in the whole of T, but also in balls. For this we introduce when $s \in T$, $r > 0$, the quantity

$$\varphi_X(s, r) = \mathrm{E}\Bigl(\sup_{t \in B(s,r)} X_t \Bigr). \tag{15}$$

Under the assumption of Proposition 1, we have

$$\varphi_Y(s, r) \leqslant \varphi_X(s, r)$$

for all $s \in T$, $r > 0$: we just observe that the assumption of Proposition 1 holds for the two processes restricted to the ball $B(s, r)$. When only one process (X_t) appears in the discussion, we shall simply write $\varphi(s, r) = \varphi_X(s, r)$.

It is obvious that $\varphi(t, r)$ is non-decreasing in r, perhaps not continuous in r: if t_0 is isolated in T, with $B(t_0, r_0) = \{t_0\}$ and $X_{t_0} \neq 0$, and if there is a non zero random variable X_s in the process with $d(t_0, s) = r_0$, then we have when $0 < r \leqslant r_0$ that the ball $B(t_0, r)$ reduces to $\{t_0\}$, hence $\varphi(t_0, r) = \mathrm{E}\, X_{t_0}$ is equal to 0, but $\varphi(t_0, r)$ jumps when $r > r_0$ to a non-zero value larger than or equal to $\mathrm{E}\max(X_{t_0}, X_s) > 0$.

1.3.2. *A convex digression*

We may use *auto-indexation* for the process, by considering that the indexing set T is precisely the subset of $L^2(\Omega, P)$ consisting of the variables in the process, so that we have $X_t = t \in L^2(\Omega, P)$. This would not take care of situations where perhaps $s \neq t$ but $X_s = X_t$; they are anyway irrelevant when dealing with the supremum of a process. Working with subsets of $L^2(\Omega, \mathrm{P})$ is what Sudakov does in his book [Sud$_2$].

If we use auto-indexing, we understand that passing from $T \subset L^2(\Omega, P)$ to the convex hull $\mathrm{conv}(T)$ of T in $L^2(\Omega, \mathrm{P})$ will not change the supremum of the process: for every $\omega \in \Omega$, we have

$$\sup_{t \in T} X_t(\omega) = \sup_{s \in \mathrm{conv}(T)} X_s(\omega),$$

so that when dealing with suprema, we may assume that T is a convex set in $L^2(\Omega, \mathrm{P})$; however, when $T \subset L^2$ and $t \in T$ we will prefer writing X_t than just t, although $t = X_t$. If $s \in T$, if $\theta \in (0, 1)$ and $t \in B(s, r)$, then $(1 - \theta)\, s + \theta\, t \in B(s, \theta r)$ hence

$$\sup_{u \in B(s,\theta r)} X_u \geqslant (1 - \theta) X_s + \theta \sup_{t \in B(s,r)} X_t,$$

and since $\mathrm{E}\, X_s = 0$, we see that

$$\varphi(s, \theta r) \geqslant \theta\, \varphi(s, r).$$

If $s, t \in T$ are such that $d(s, t) < \delta$, then $B(t, r) \subset B(s, r + \delta)$, therefore

$$\varphi(t, r) \leqslant \varphi(s, r + \delta) \leqslant \frac{r + \delta}{r}\, \varphi(s, r).$$

It follows that $\varphi(t, r)$ is continuous in $t \in T$ in the convex case.

We come back now to Sudakov's work: Sudakov actually uses a geometric quantity that is proportional to the expectation of the supremum of a Gaussian process, where the process is seen as a subset K of a Gaussian subspace of $L^2(\Omega, \mathrm{P})$, and as we have said, we may and shall assume that K is convex. This approach of Sudakov uses a suitably normalized *mixed volume* (9).

Let us consider first the simplest case where K is a segment $[0, x] \neq \{0\}$ in the Euclidean space $\mathbb{R}^n$. Let B_n be the unit ball in $\mathbb{R}^n$, let $|A|_n$ denote the n-dimensional volume of $A \subset \mathbb{R}^n$ and set $v_n = |B_n|_n$. Then the n-dimensional volume of the Minkowski sum

$$B_n + u\,K = \{y + u\,z : y \in B_n,\, z \in [0, x]\}, \quad \text{where} \quad u > 0,$$

is the volume of the convex hull of the union of B_n and its translate $B_n + u\,x$. It is easy to see that $(B_n + u\,K) \setminus B_n$ consists of intervals of length $u\|x\|$ situated on the lines ℓ parallel to $[0, x]$ that intersect B_n. Let $x^\perp$ denote the hyperplane orthogonal to the segment $[0, x]$. The intersection points of those lines ℓ with $x^\perp$ fill the $(n-1)$-dimensional ball of that hyperplane, hence

$$|(B_n + u\,K) \setminus B_n|_n = v_{n-1} \,.\, u\|x\|$$

and

$$|B_n + u\,K|_n = v_n + v_{n-1} \,.\, u\|x\|, \tag{16}$$

so that the normalized expression

$$\frac{1}{v_{n-1}} \Bigl(\frac{|B_n + u\,K|_n - |B_n|_n}{u} \Bigr) = \|x\|$$

becomes independent of the dimension n of the Euclidean space into which the segment is embedded, and can be used for extending the notion to embeddings in an infinite dimensional Hilbert space.

When K is a k-dimensional convex compact subset of some $\mathbb{R}^n$, the expression in formula (16) —of degree one in u— transforms into a degree k polynomial in u, of which we can extract the "normalized" coefficient of u by looking at

$$h_1(K) := \lim_{u \to 0} \frac{1}{v_{n-1}} \Bigl(\frac{|B_n + u\,K|_n - |B_n|_n}{u} \Bigr).$$

Again, this does not depend upon the dimension n where K is embedded, and the definition of $h_1(K)$ can be extended to a compact convex subset K of a Hilbert space. Assuming that $K \subset L^2(\Omega, \mathrm{P})$ is the auto-indexing set of a centered Gaussian process $(X_t)_{t \in K}$, Sudakov as shown that

$$h_1(K) = \sqrt{2\pi}\, \mathrm{E}\Bigl(\sup_{t \in K} X_t \Bigr), \tag{17}$$

a fact that we can get immediately in our obvious example of a segment: if $x = 1 \in \mathbb{R}$ and $K = [0, 1]$, then the Gaussian variable associated to the point $1 \in K$ is a $\mathrm{N}(0, 1)$ variable g, and

$$h_1([0, 1]) = 1 = \|x\|, \qquad \sup_{t \in K} X_t = \max(g, 0),$$
$$\mathrm{E} \max(g, 0) = \int_0^\infty x\, \mathrm{e}^{-x^2/2}\, \frac{\mathrm{d}x}{\sqrt{2\,\pi}} = \frac{1}{\sqrt{2\pi}}.$$

Simone Chevet [Che$_2$] has obtained results that relate higher moments of the supremum to other normalized mixed volumes of K. The result for the second moment takes a fairly explicit form, let us express it for a finite index set T: given a Gaussian random vector $X = (X_1, X_2, \ldots, X_n)$, let $\sigma(\omega)$ denote the smallest index in $\{1, 2, \ldots, n\}$ such that

$$X_{\sigma(\omega)}(\omega) = \max_{1 \leqslant i \leqslant n} X_i(\omega), \quad \omega \in \Omega.$$

One has then

$$\begin{aligned} &\mathrm{E}\Big(\sup_{t\in K} X_t\Big)^2 - \mathrm{E}_\omega\, \mathrm{E}_{\omega'}(X_{\sigma(\omega)}(\omega'))^2 \\ =&\mathrm{E}\Big(\sup_{t\in K} X_t\Big)^2 - \sum_{i=1}^n \mathrm{P}(\sigma = i)\ \mathrm{E}\, X_i^2 = \frac{h_2(K)}{\pi}. \end{aligned}$$

1.4. *First applications of comparison*

We begin with a simple lemma.

Lemma 3. *Let (X_i), $i \in I$, and $(Y_{i,j})$, $i \in I$ and $j \in J_i$, be two independent families of integrable real random variables. One has that*

$$\mathrm{E}\Big(\sup_{i,j}\big(X_i + Y_{i,j}\big)\Big) \geqslant \mathrm{E}\Big(\sup_{i\in I} X_i\Big) + \inf_{i \in I} \mathrm{E}\Big(\sup_{j\in J_i} Y_{i,j}\Big),$$

where we must agree that $(+\infty) + (-\infty) = -\infty$ if it appears in the sum above.

Proof. Let

$$Y_* = \inf_{i\in I} \mathrm{E}\Big(\sup_{j \in J_i} Y_{i,j}\Big) \in [-\infty, \infty].$$

By independence, we may think of the X_i as functions of a variable u while the $Y_{i,j}$ are functions of a different variable v. We have

$$\mathrm{E}_v \sup_{i,j}\big(X_i + Y_{i,j}\big) = \mathrm{E}_v \sup_{i\in I}\sup_{j\in J_i}\big(X_i + Y_{i,j}\big) \geqslant \sup_{i \in I} \mathrm{E}_v \sup_{j\in J_i}\big(X_i + Y_{i,j}\big),$$

then

$$\mathrm{E}_v \sup_{j\in J_i}\big(X_i(u) + Y_{i,j}(v)\big) = X_i(u) + \mathrm{E} \sup_{j \in J_i} Y_{i,j} \geqslant X_i(u) + Y_*$$

and

$$\sup_{i\in I} \mathrm{E}_v \sup_{j\in J_i}\big(X_i + Y_{i,j}\big) \geqslant \sup_{i\in I} X_i(u) + Y_*.$$

Integrating in u concludes the proof. □

The next Lemma does most of the serious job in what follows(10).

Lemma 4. *Suppose that $(X_t)_{t\in A}$ is a centered Gaussian process, that ρ, δ, λ are positive real numbers such that $\rho^2 + \lambda^2 \leqslant \delta^2/2$ and that $(A_i)_{i=1}^N$ are subsets of the index set A satisfying*

– *we have $A_i \subset B(a_i, \rho)$ for some $a_i \in A$, and*

– *for all $t_i \in A_i$, $t_j \in A_j$ and $i \neq j$ we have $d(t_i, t_j) \geqslant \delta$ (we may say that the two sets A_i and A_j are δ-*separated*).*

It follows that

$$\mathrm{E}\Big(\sup_{t\in A} X_t\Big) \geqslant (\lambda/2)\sqrt{\ln N} + \min_{1\leqslant i\leqslant N} \mathrm{E}\Big(\sup_{t\in A_i} X_t\Big).$$

Proof. Let $(X_t^{(i)})_{i=1}^N$ be N independent copies of the process (X_t), let $g_1, g_2, \ldots, g_N$ be independent $\mathrm{N}(0,1)$ Gaussian variables, that are also independent from the $(X_t^{(i)})_{i=1}^N$, and set

$$A_* = \bigcup_{i=1}^N A_i \subset A.$$

Let us define another Gaussian process $(Y_t)_{t \in A_*}$ as

$$Y_t = X_t^{(i)} - X_{a_i}^{(i)} + \lambda g_i \quad \text{when} \quad t \in A_i.$$

We want to check that

$$\mathrm{E}(Y_s - Y_t)^2 \leqslant \mathrm{E}(X_s - X_t)^2$$

for all $s, t \in A_*$, in order to apply the Sudakov–Slepian lemma; there are two cases to consider: if there is an index i such that $s, t \in A_i$, we have

$$Y_s - Y_t = X_s^{(i)} - X_t^{(i)},$$

hence

$$\mathrm{E}(Y_s - Y_t)^2 = \mathrm{E}(X_s - X_t)^2$$

in this first case. If now $s \in A_i$, $t \in A_j$ and $i \neq j$, we see that

$$Y_s - Y_t = (X_s^{(i)} - X_{a_i}^{(i)}) + \lambda g_i - (X_t^{(j)} - X_{a_j}^{(j)}) - \lambda g_j\,;$$

the four random variables $X_s^{(i)} - X_{a_i}^{(i)}$, g_i, $X_t^{(j)} - X_{a_j}^{(j)}$ and g_j are centered and independent, hence orthogonal, and $d(s, a_i) \leqslant \rho$, $d(t, a_j) \leqslant \rho$, therefore

$$\mathrm{E}(Y_s - Y_t)^2 \leqslant 2\,\rho^2 + 2\,\lambda^2 \leqslant \delta^2 \leqslant \mathrm{E}(X_s - X_t)^2,$$

because we know that $d(s,t) \geqslant \delta$ since A_i and A_j are δ-separated. Using the Sudakov–Slepian lemma we obtain

$$\mathrm{E}\Bigl(\sup_{t \in A_*} Y_t\Bigr) \leqslant \mathrm{E}\Bigl(\sup_{t \in A_*} X_t\Bigr)$$

and as $A_* \subset A$ we have obviously that

$$\mathrm{E}\Bigl(\sup_{t \in A_*} X_t\Bigr) \leqslant \mathrm{E}\Bigl(\sup_{t \in A} X_t\Bigr).$$

It remains to apply Lemma 3, the lower bound (11) and get

$$\begin{aligned}
\mathrm{E}\Bigl(\sup_{t \in A} X_t\Bigr) &\geqslant \mathrm{E} \sup_{t \in A_*} Y_t = \mathrm{E} \sup_i \sup_{t \in A_i} \bigl((X_t^{(i)} - X_{a_i}^{(i)}) + \lambda g_i\bigr) \\
&\geqslant \mathrm{E} \max_{1 \leqslant i \leqslant N} (\lambda g_i) + \min_i \mathrm{E} \sup_{t \in A_i} (X_t^{(i)} - X_{a_i}^{(i)}) \\
&\geqslant (1/2)\,\lambda \sqrt{\ln N} + \min_i \bigl(\mathrm{E} \sup_{t \in A_i} X_t - \mathrm{E}\, X_{a_i}\bigr) \\
&= (\lambda/2) \sqrt{\ln N} + \min_i \mathrm{E} \sup_{t \in A_i} X_t. \qquad \square
\end{aligned}$$

Corollary 1. *Suppose that $(X_t)_{t \in T}$ is a centered Gaussian process, that $S \subset T$ is a finite 2δ-separated set contained in a ball $B(t_0, r)$, where $t_0 \in T$, $\delta, r > 0$. It follows that*

$$\mathrm{E}\Bigl(\sup_{t \in B(t_0, r + \delta/2)} X_t\Bigr) \geqslant (\delta/4) \sqrt{\ln |S|} + \min_{s \in S} \mathrm{E}\Bigl(\sup_{t \in B(s, \delta/2)} X_t\Bigr),$$

or, using Notation (15) and letting $N = |S|$ denote the cardinality of S,

$$\varphi(t_0, r + \delta/2) \geqslant (\delta/4) \sqrt{\ln N} + \min_{s \in S} \varphi(s, \delta/2).$$

Proof. We apply Lemma 4 with $\lambda = \rho = \delta/2$, $A = B(t_0, r + \delta/2)$ and $A_i = B(s_i, \rho)$, where $S = \{s_1, s_2, \ldots, s_N\}$. Then $\rho^2 + \lambda^2 = \delta^2/2$, and by the triangle inequality the balls $B(s_i, \rho) = B(s_i, \delta/2)$ are δ-separated and contained in A. □

1.5. *Trees and branches*

We shall deal with rooted trees $\mathcal{X}$; we consider that $\mathcal{X}$ is a set of *nodes*, and that to each node $x \in \mathcal{X}$ is associated a *finite* subset $C(x)$ of nodes in $\mathcal{X}$ that are the *children* of x, with $x \notin C(x)$ of course and

$$x \neq x' \ \Rightarrow\ C(x) \cap C(x') = \emptyset.$$

We assume that this *child relation* defines a rooted tree $\mathcal{X}$, whose root will be called x_0. For each child $y \in C(x)$ we say that x is the *parent* of y. The root has no parent. The *descendants* of $x \in \mathcal{X}$ are the elements of one of the sets $C^{(k)}(x)$, $k \geqslant 1$, inductively defined by letting $C^{(1)}(x) = C(x)$ and

$$C^{(k+1)}(x) = \bigcup_{y \in C(x)} C^{(k)}(y), \quad k \geqslant 1.$$

Every node $y \in \mathcal{X}$ different from the root x_0 is a descendant of the root. If y is a descendant of $x \in \mathcal{X}$, we say that x is an *ancester* of y: the root x_0 is an ancester of every node in the tree.

For $x \in \mathcal{X}$, we shall pay attention to the number of *siblings* of x: except if x is the root, this is the number of children of the parent of x —the "sisters and brothers" of x—; this number is set equal to 1 for the root. We shall indicate the successive levels in the tree by successive nonnegative integers: level 0 for the root, level 1 for the children of the root, 2 for the grandchildren of the root, and so on. The level of a node x in the tree will be denoted by $\ell(x) \in \mathbb{N}$,

$$\ell(x_0) = 0; \qquad y \in C(x) \ \Rightarrow\ \ell(y) = \ell(x) + 1.$$

The *maximal branches* of the tree are sequences $\mathbf{x} = (x_i)_{0 \leqslant i < L}$ of nodes in the tree, where $L = L(\mathbf{x})$ is finite or $L = +\infty$, and:

- the node x_0 in the branch $\mathbf{x}$ is the root of the tree,
- the node x_{i+1} is a child of x_i whenever $i \in \mathbb{N}$ and $i + 1 < L$,
- when L is finite, the node x_{L-1} has no child, it is a *leaf* of the tree.

We shall prefer avoiding leaves: the maximal branches of our main trees will all be infinite, this will help us to keep a somewhat unified treatment. A maximal branch $\mathbf{x}$ in $\mathcal{X}$ will thus look like

$$\mathbf{x} = (x_0,\ x_1,\ x_2,\ \ldots,\ x_i,\ \ldots).$$

A *path* in the tree will be a portion of a branch, namely, a finite sequence of nodes

$$x_j,\ x_{j+1},\ \ldots,\ x_k,$$

where x_{i+1} is a child of x_i for every i such that $j \leqslant i < k$.

Typically, our trees will arise from covering a compact subset T of a Hilbert space: a node would be somehow a ball B in T of some radius $r > 0$, and its children will be balls B' of smaller radius r', say $r' = r/3$, that cover B. In dimension n we can expect the covering of $T \subset \mathbb{R}^n$ by balls of radius r to have a cardinality of order r^{-n}, so that passing to children with radius $r/3$ would increase the covering number to some $3^n r^{-n}$ balls; one can then think that each node will have about 3^n children, a constant factor depending upon the given dimension n. But we work actually in infinite dimension, and the number of children, or of siblings of a given node, may increase dramatically from one level to the next in the tree.

We suppose that we have a rooted tree $\mathcal{X}$ such that all its maximal branches are infinite, or equivalently, such that for every $x \in \mathcal{X}$, the set $C(x)$ of children is never empty. We introduce the notion of a *3-control function* N, a function on the tree $\mathcal{X}$ that will control the evolution of the cardinality $|C(x)|$. We choose to say that $N(x)$ controls the number of siblings of $x \in \mathcal{X}$, rather than the number of children of x. Perhaps paradoxically, we do not try to keep this function small, but on the contrary, guarantee a huge growth. The first two conditions below define a *3-growing function* N on the tree; we shall then say that $N(x)$ is the *growth value* at $x \in \mathcal{X}$. Thus N is a 3-growing function on $\mathcal{X}$ if it satisfies the two following conditions $\mathbf{c_0}$ and $\mathbf{c_1}$:

$\mathbf{c_0}$ — We have $N(x_0) = 1$, and $N(y) \geqslant 2$ for every child $y \in C(x_0)$ of the root x_0.

$\mathbf{c_1}$ — For every node $x \in \mathcal{X}$, the value of N is the same for all children y of x, and $N(y) \geqslant N(x)^3$, that is to say (11)

$$y, y' \in C(x) \ \Rightarrow\ N(y) = N(y'); \qquad y \in C(x) \Rightarrow N(y) \geqslant N(x)^3.$$

Since $N(x) \geqslant 2$ for any node x at level $\ell(x) = 1$, and $N(y) \geqslant N(x)^3$ when y is a child of x, we see easily by induction that (12)

if $\ell(x) = k$, then $N(x) \geqslant 2^k$.

When $\ell(x) = 2$, one has actually $N(x) \geqslant 2^3 = 8 > 2^{\ell(x)}$, so that

if $\ell(x) = k > 1$, then $N(x) > 2^k$.

It follows that for any maximal branch $(x_i)_{i \geqslant 0}$ of the tree, we have

$$N(x_i) \geqslant 2^i \text{ for } i \geqslant 0, \quad \sum_{j=1}^{\infty} N(x_j)^{-1} < 1 \text{ thus } \sum_{j=0}^{\infty} N(x_j)^{-1} < 2. \tag{18}$$

We say that N is a *3-control function* on the rooted tree $\mathcal{X}$ if it is a 3-growing function, thus satisfying $\mathbf{c_0}$ and $\mathbf{c_1}$, and if in addition:

$\mathbf{c_2}$ — For every $x \in \mathcal{X}$ and $y \in C(x)$, the value of $N(y)$ is an upper bound for the number of siblings of y,

$$y \in C(x) \ \Rightarrow\ |C(x)| \leqslant N(y).$$

The growth condition in $\mathbf{c_1}$ will be used to exert a backward hold over the number of ancesters: if y is a child of x, we have that $N(x) \leqslant N(y)^{1/3}$. This will become clear in the next lemma.

Suppose that the tree $\mathcal{X}$ admits a 3-control function N, thus satisfying $\mathbf{c_0}$, $\mathbf{c_1}$ and $\mathbf{c_2}$. Let M be a number $\geqslant 1$ and consider the subtree of $\mathcal{X}$ defined by

$$\mathcal{X}_M = \{x \in \mathcal{X} : N(x) \leqslant M\}. \tag{19}$$

Because $N(y) > N(x)$ when $y \in C(x)$, we see that if $y \in \mathcal{X}_M$, then all ancesters of y in $\mathcal{X}$ belong to $\mathcal{X}_M$. It follows from the first inequality in (18) that all branches of $\mathcal{X}_M$ are finite, and since the set $C(x)$ is finite for every node $x \in \mathcal{X}$ by $\mathbf{c_2}$, we know that $\mathcal{X}_M$ is a finite tree (13). Note that the root x_0 of $\mathcal{X}$ belongs to $\mathcal{X}_M$, because $M \geqslant 1$ and because $N(x_0) = 1$ according to $\mathbf{c_0}$.

Lemma 5. *Suppose that the tree $\mathcal{X}$ admits a 3-control function N, and that the subtree $\mathcal{X}_M$ is defined by (19). The number of leaves of $\mathcal{X}_M$ is bounded above by $M^{3/2}$.*

Proof. When $M = 1$, the tree $\mathcal{X}_M$ reduces to the root $\{x_0\}$ of $\mathcal{X}$, and the number of leaves of $\mathcal{X}_M$ is clearly equal to $1 = M^{3/2}$. We assume now that $M > 1$, and we consider the subset $L_M \subset \mathcal{X}_M$ consisting of the leaves of $\mathcal{X}_M$. Set $h(x) = 0$ when the node $x \in L_M$ is a leaf of $\mathcal{X}_M$. Next, for $x \in \mathcal{X}_M \setminus L_M$, define inductively $h(x)$ by

$$h(x) = 1 + \max\{h(y) : y \in C_M(x)\} \quad \text{where} \quad C_M(x) = C(x) \cap \mathcal{X}_M.$$

For $x \in \mathcal{X}_M$, let $\lambda(x)$ denote the number of leaves $z \in L_M$ that are either equal to x, or are a descendant of x. When $h(x) = 0$, then x is a leaf of $\mathcal{X}_M$, has therefore no descendant in $\mathcal{X}_M$, thus $\lambda(x) = 1$. When $h(x) = 1$, the leaves of $\mathcal{X}_M$ that are a descendant of x are necessarily among the children of x in $\mathcal{X}$, and x itself is not a leaf of $\mathcal{X}_M$; this yields that $\lambda(x) \leqslant |C(x)|$; also, there is then (at least) one child y of x that is a leaf of $\mathcal{X}_M$, and $N(y) \leqslant M$ because $y \in \mathcal{X}_M$. It follows from $\mathbf{c_2}$ that

$$\lambda(x) \leqslant |C(x)| \leqslant N(y) \leqslant M.$$

As $M > 1$ and $\lambda = 1$ for leaves (when $h = 0$), this yields that

$$h(x) \leqslant 1 \;\; \Rightarrow \;\; \lambda(x) \leqslant M.$$

When $h(x) = 2$, there is a grandchild z of x, child of some $y \in C(x)$, that is a leaf of $\mathcal{X}_M$. We then have $N(z) \leqslant M$, and $N(y) \leqslant N(z)^{1/3}$ by $\mathbf{c_1}$; hence, using $\mathbf{c_2}$ we obtain

$$|C(x)| \leqslant N(y) \leqslant M^{1/3}.$$

Since $h(x) = 2$, we have $h(y) \leqslant 1$ for all the children y of x that are in $\mathcal{X}_M$, so $\lambda(y) \leqslant M$ and $x \notin L_M$, thus

$$\lambda(x) = \sum_{y \in C_M(x)} \lambda(y) \leqslant |C(x)|\, M \leqslant M^{1/3} M = M^{1+1/3}.$$

We deduce that

$$h(x) \leqslant 2 \;\; \Rightarrow \;\; \lambda(x) \leqslant M^{1+1/3}.$$

When $h(x) = 3$, we see in the same way that $|C(x)| \leqslant N(y) \leqslant M^{1/9}$ and $h(y) \leqslant 2$ for each one of the children $y \in C_M(x)$, hence $\lambda(x) \leqslant |C(x)|\, M^{1+1/3}$ and

$$h(x) \leqslant 3 \;\; \Rightarrow \;\; \lambda(x) \leqslant M^{1+1/3+1/9} < M^{3/2}.$$

We conclude easily inductively that when $x \in \mathcal{X}_M$ and $h(x) \leqslant k + 1$, we have

$$\lambda(x) \leqslant M^{1+1/3+1/9+1/27+\cdots+1/3^k} < M^{3/2},$$

in particular for the root x_0, so $M^{3/2}$ is a bound for the number of leaves of $\mathcal{X}_M$. □

2. The Fernique theorem

Let G be a compact abelian additive group and let $(X_g)_{g \in G}$ be a centered Gaussian process indexed by G. We suppose that this process is *invariant* under the group action, meaning that for every finite sequence $\{h, g_1, \ldots, g_k\}$ of elements of G, the two k-tuples of random variables

$$(X_{g_1}, X_{g_2}, \ldots, X_{g_k}) \quad \text{and} \quad (X_{g_1+h}, X_{g_2+h}, \ldots, X_{g_k+h})$$

have the same joint distribution. The official terminology says that the process $(X_g)_{g \in G}$ is a *stationary process*.

We assume of course that the family $(X_g)_{g \in G}$ is not reduced to a single random variable. We may slightly modify the point of view, and consider that the group acts on the set $\{X_g : g \in G\}$ of random variables by

$$h.X_g = X_{g+h}.$$

This action is transitive: given $g_0, g_1 \in G$, we have with $h = g_1 - g_0$ that

$$h.X_{g_0} = X_{g_1}.$$

We now arrive to the following setting: we assume that $(X_t)_{t\in T}$ is a centered Gaussian process, that G is a *multiplicative* group acting transitively on T, and that for every integer $k \geqslant 1$, every sequence $\{t_1, \ldots, t_k\}$ of elements of T and every $g \in G$, the two k-tuples

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \quad \text{and} \quad (X_{g.t_1}, X_{g.t_2}, \ldots, X_{g.t_k}) \tag{20}$$

have the same joint distribution. It follows that for $t_1, t_2 \in T$ and $g \in G$,

$$d(t_1, t_2) = d(g.t_1, g.t_2).$$

If $B(t, r)$ is a ball in T and $g \in G$, we see that

$$g.B(t, r) \subset B(g.t, r) \quad \text{therefore} \quad g.B(t, r) = B(g.t, r)$$

by letting the inverse g^{-1} of g act on $B(g.t, r)$. If $s_1, s_2, \ldots, s_n$ are arbitrary elements in $B(t, r)$, then $g.s_1, g.s_2, \ldots, g.s_n$ are in $B(g.t, r)$, and the distributional invariance yields that

$$\mathrm{E}\Bigl(\sup_{1\leqslant i\leqslant n} X_{s_i}\Bigr) = \mathrm{E}\Bigl(\sup_{1\leqslant i\leqslant n} X_{g.s_i}\Bigr).$$

Therefore, passing from finite subsets of G to infinite ones, we see that invariance and transitivity of the group action imply that for any t_0, t_1 in T, we have

$$\varphi(t_0, r) = \mathrm{E}\Bigl(\sup_{s\in B(t_0,r)} X_s\Bigr) = \mathrm{E}\Bigl(\sup_{s\in B(t_1,r)} X_s\Bigr) = \varphi(t_1, r). \tag{21}$$

In this invariant setting (14), we simply let

$$\varphi(r) = \varphi(t_0, r)$$

for an arbitrary fixed $t_0 \in T$ and for every $r > 0$.

An additional observation: suppose that $0 < r_1 < r_0$ and that S_0 is a r_1-p-net for the ball $B(t_0, r_0)$, for t_0 fixed in T. We can use the transitive group action in order to find for each $t \in T$ a r_1-p-net S for the ball $B(t, r_0)$, with $|S| = |S_0|$: we use $g \in G$ such that $g.t_0 = t$ and we let $S = g.S_0$. This shows that

$$\text{for } \ 0 < r' < r, \quad \mathcal{N}^*(B(t, r), r') \ \text{ does not depend upon } \ t \in T. \tag{22}$$

We shall prove the Fernique theorem under the form that follows.

Theorem 1. *Let $(X_t)_{t\in T}$ be a centered Gaussian process that satisfies the invariance conditions (21) and (22). If the expectation of the supremum $\sup_{t\in T} X_t$ is finite, then the Dudley integral is finite, and more precisely*

$$I_D(T) \ \leqslant\ \Sigma_1(T) \ \leqslant\ 432\,\mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr).$$

We can absolutely guarantee that the constant 432 above is not optimal. In fact, we shall not care about keeping the constants "right", we shall often and repeatedly replace for example an upper bound $\sqrt{3}$ by 2, in order to get simple but exaggerated integer values in the estimates.

2.1. *Looking for an entropy-like condition*

We assume that we have an invariant centered Gaussian process $(X_t)_{t\in T}$ such that

$$E^* := \mathrm{E}\Big(\sup_{t\in T} X_t\Big) < \infty.$$

Due to invariance, we can set $\varphi(r) = \varphi(t, r)$ for every $t \in T$ and every radius $r > 0$, and we know that $\varphi(r) \leqslant E^* < +\infty$. We can then rewrite Corollary 1.

Corollary 2. *Let $(X_t)_{t\in T}$ be a centered Gaussian process that satisfies the invariance condition (21). If $S \subset T$ is a finite 2δ-separated set that is contained in a ball $B(t_0, r)$, where $t_0 \in T$ and $\delta, r > 0$, one has that*

$$\varphi(r + \delta/2) \geqslant (\delta/4)\sqrt{\ln |S|} + \varphi(\delta/2).$$

We shall construct a rooted tree $\mathcal{X}$ and a 3-control function N for $\mathcal{X}$. The idea of associating a tree to the process can be traced in Chap. 7 of Fernique's work [Fer3]. The integer $N(x)$ will (in particular) be an upper bound for the number of siblings of x in $\mathcal{X}$. The maximal branches of this tree will all be infinite, at the cost of doing things in a way ([15]) that is slightly unnatural, but that helps us to present a moderately unified treatment, making no difference between the case of a *perfect* index set T or that of a finite set T.

Let us give a very short but very inexact description of the procedure: we will construct a tree $\mathcal{X}$ whose nodes correspond to smaller and smaller balls $B(t, r)$ in T. The "richness" of the tree will reflect the entropy of T, as the children of the node corresponding to $B(t, r)$ will arise from covering $B(t, r)$ by balls with a smaller radius. We shall apply Corollary 2 repeatedly when passing from a node to its children. Given a ball $B(t, r)$ associated to a node in $\mathcal{X}$, we shall find a suitable ρ-p-net S of N points in that ball, with $\rho < r$, and get from Corollary 2 an inequality that has the form

$$\varphi(r) \geqslant r\sqrt{\ln N} + \varphi(cr)$$

with $c < 1$ a fixed positive real number. Let us rather write it as

$$\varphi(r) - \varphi(cr) \geqslant r\sqrt{\ln N}.$$

Smaller balls of radius $r' = cr$ around the N points of the preceding ρ-p-net S will correspond to the children of the node associated to $B(t, r)$. Next, we would find for each of these children another analogous estimate

$$\varphi(cr) - \varphi(c^2 r) \geqslant r'\sqrt{\ln N'}.$$

Passing successively from r_i to $r_{i+1} = cr_i$ for all $i \geqslant 0$, a telescopic summation effect for the successive differences $\varphi(r_i) - \varphi(r_{i+1})$ will give us an upper bound $\varphi(r_0) = E^*$ for an expression

$$\sigma_1 := \sum_{i=0}^{\infty} r_i\sqrt{\ln N_{i+1}}$$

very similar to $\Sigma_1(T)$, and actually equivalent to it up to a universal multiplicative constant —one may prefer to say: an *absolute* constant—. The actual proof requires some more technicalities that will make the preceding summary simply vastly erroneous.

2.1.1. *Constructing a tree*

We shall construct a rooted tree $\mathcal{X}$, and a function N that will be a 3-control function on this tree. A node $x \in \mathcal{X}$ at level $i \geqslant 0$ will have the form ([16])

$$x = (t_0, t_1, \ldots, t_i) \ \text{ with } \ t_j \in T, \ \ j = 0, 1, \ldots, i,$$

where $t_0, t_1, \ldots, t_{i-1}$ are the successive *positions* of the ancesters of x, whereas t_i is the *present position*, or *latest position* in x. A radius $r_i = \rho(x) > 0$ will be associated to nodes at level i, we may consider that ρ is another function on the tree, but an extremely simple one: starting from the root x_0 with an inital radius $r_0 = \rho(x_0)$ precised below, the radius will simply be divided by 3 when moving from a level to the next,

$$r_{i+1} = r_i/3, \ \ i \geqslant 0.$$

In most of what will be done below, only the couple (t_i, r_i) will matter, where t_i is the latest position appearing in the node x and where $r_i = \rho(x) = 3^{-i} r_0$ is the radius associated to x. We shall use the notation

$$x \succ (t_i, r_i)$$

in order to mention that we have extracted from the whole node x that most relevant information (t_i, r_i). The preceding positions $t_0, t_1, \ldots, t_{i-1}$ in x will not appear explicitly in what follows, but they will implicitly enter in the evaluation of the control value $N(x)$. As it was explained briefly above, the expectation of the supremum of the values X_t of the process, for t in the ball $B(t_i, r_i) \subset T$, will be examined and used. Recall that under our invariance assumptions, we have

$$\varphi(r_i) = \varphi(t_i, r_i) = \mathrm{E}\Bigl(\sup_{t \in B(t_i, r_i)} X_t \Bigr).$$

Let $\Delta > 0$ denote the —finite([2])— diameter of the set T, let t_0 be an arbitrary point in T and let r_0 satisfy $r_0 > \Delta$, so that we have $B(t_0, r_0) = T$; assume in addition that $r_0 < 4\Delta/3$: that assumption will be used only later on. The root of the tree is $x_0 = (t_0)$, we let $\rho(x_0) = r_0$ be the radius associated to x_0, and the initial value of N is set to $N(x_0) = N_0 = 1$. The inductive construction of the tree and the function N goes as follows:

— suppose that a node $x_* \in \mathcal{X}$ and a value $N_* = N(x_*)$ for the function N at x_* have already been introduced. Let $t_* \in T$ be the latest position in x_* and $r_* = \rho(x_*) > 0$ be the radius associated to x_*, so that $x_* \succ (t_*, r_*)$. Consider

$$r = r_*/3 \ \text{ and let } \ S \subset B(t_*, r_*) \ \text{ be a } r\text{-p-net for } \ B(t_*, r_*).$$

The ball $B(t_*, r_*)$ contains at least the point t_* and thus S is not empty. The size of S will play a rôle in what follows: the case when $|S| > N_*^3$ corresponds to when $B(t_*, r_*)$ is "fairly large" and we shall call it the "rich case", the case $|S| \leqslant N_*^3$ being of course the "poor" one. In the rich case, we let

$$\widetilde{N} = |S|, \ \text{ and in the poor case } \ \widetilde{N} = N_*^3, \ \text{ so that } \ \widetilde{N} = \max(N_*^3, |S|).$$

For each $s \in S$, we define a new node x of the tree, that is declared to be a child of x_*, by appending the position s to the list of positions in x_*, as the latest position of x; we define N at x by $N(x) = \widetilde{N}$. We now repeat things a little differently, mentioning the successive generations in the tree: let $i \geqslant 0$ be an integer and let

$$x_* = x_i = (t_0, t_1, \ldots, t_i)$$

be a node at level i in the tree, with a control value $N_* = N_i = N(x_i)$ for N at x_i; let also $r_* = r_i = \rho(x_i) = 3^{-i} r_0$, so that $x_i \succ (t_i, r_i)$. Consider

$$r_{i+1} = r_i/3 \text{ and let } S_{i+1} \text{ be a } r_{i+1}\text{-p-net for } B(t_i, r_i).$$

Define $N_{i+1} = \max(N_i^3, |S_{i+1}|)$. Then, the children of x_i have the form

$$x_{i+1} = (t_0, t_1, \ldots, t_i, t_{i+1}),$$

where t_{i+1} can be any point in S_{i+1}. The new value of N is $N(x_{i+1}) = N_{i+1}$, and we let $\rho(x_{i+1}) = r_{i+1}$. There is at least one child for the node x_i —because the set S_{i+1} is not empty—, and there are at most N_{i+1} children of x_i, since $|C(x_i)| = |S_{i+1}| \leqslant N_{i+1}$.

Let us check that the function N is a 3-control function on the tree $\mathcal{X}$: for the root $x_0 \succ (t_0, r_0)$, we have set $N(x_0) = 1$; also, we have $\Delta < r_0 < 4\Delta/3$. It yields first that $B(t_0, r_0) = T$; next, we see that $\mathcal{N}(B(t_0, r_0), r_0/3) > 1$ since $2(r_0/3) < 8\Delta/9 < \Delta$, implying that $B(t, r_0/3) \neq T$ for any point $t \in T$. We have thus $|S_1| = N_1 \geqslant 2$, the condition $\mathbf{c_0}$ for a control function is satisfied. The function N was defined in a way that it has the same value $N(y)$ for all children y of a node x. In addition, we know that $N(y) \geqslant N(x)^3$, thus N satisfies $\mathbf{c_1}$; finally, we have $|C(x)| \leqslant N(y)$, condition $\mathbf{c_2}$.

A few comments are in order.

— We see that r_i is just $3^{-i} r_0$, but I find that r_i is better looking than $3^{-i} r_0$: I will keep r_i throughout. We have $r_{i+1} = r_i/3$ for $i \geqslant 0$; let us decide to set $r_{-1} = 3r_0$ so that we can say that $r_{i-1} = 3r_i$ for all $i \geqslant 0$.

— Being an r_{i+1}-p-net for $B(t_i, r_i)$, the set S_{i+1} is r_{i+1}-separated and

$$S_{i+1} \subset B(t_i, r_i) \subset \bigcup_{s \in S_{i+1}} B(s, r_{i+1}).$$

— In the poor case, the value $N_{i+1} = N_i^3$ is merely an *upper bound* for the number of children of the node x_i ([17]).

2.1.2. *Gaussian estimates*

Let us fix a node x_i in the ith generation, where we have $i \geqslant 0$, and let us perform the "extraction" $x_i \succ (t_i, r_i)$. Passing from x_i to all its children $x_{i+1} \succ (t_{i+1}, r_{i+1})$, we shall get from Corollary 2 a relation between the expectations of the suprema of the Gaussian process $(X_t)_{t \in T}$ over a ball centered at t_i on one hand, and over balls centered at the various t_{i+1} on the other. The radii of these balls will be fixed multiples of r_i, the (common) multiple for the children being smaller than that for their parent x_i. The set of points t_{i+1} that are the *latest positions* in the children x_{i+1} of x_i form the set S_{i+1}. Recall that $r_{i+1} = r_i/3$ and that by condition $\mathbf{c_1}$, the control value $N_{i+1} = N(x_{i+1})$ only depends upon the fixed node x_i. There are two cases to consider:

— suppose first that we are in the "rich" case: then we know that $N_{i+1} = |S_{i+1}|$, and S_{i+1} is a r_{i+1}-p-net for $B(t_i, r_i)$. Let $2\delta = r_{i+1}$, so that the set $S_{i+1} \subset B(t_i, r_i)$ is 2δ-separated. Notice that $\delta/2 = r_{i+1}/4 = r_i/12$; applying Corollary 2 we get

$$\varphi(r_i + r_i/12) \geqslant (r_i/24)\sqrt{\ln N_{i+1}} + \varphi(r_i/12),$$

and carelessly writing $r_i + r_i/12 < 2r_i$ we arrive at

$$\varphi(2r_i) \geqslant (r_i/24)\sqrt{\ln N_{i+1}} + \varphi(r_i/12); \tag{23}$$

— otherwise, in the "poor case" for x_i, we know that $N_{i+1} = N_i^3$. We essentially do nothing in this case, it is not the right time, we just observe that

$$r_i \sqrt{\ln N_{i+1}} = \sqrt{3}\, r_i \sqrt{\ln N_i} = \frac{\sqrt{3}}{3}\, r_{i-1} \sqrt{\ln N_i} \leqslant \frac{2}{3} r_{i-1} \sqrt{\ln N_i} \tag{24}$$

because $r_i = r_{i-1}/3$ and $\sqrt{3} < 2$, hence

$$(r_i/24) \sqrt{\ln N_{i+1}} \leqslant \frac{2}{3} (r_{i-1}/24) \sqrt{\ln N_i},$$

and obviously we have that

$$\varphi(r_i/12) \leqslant \varphi(2r_i).$$

Adding these two informations and letting $\kappa = 1/24$ we obtain

$$\varphi(2r_i) + \frac{2}{3} \kappa r_{i-1} \sqrt{\ln N_i} \geqslant \kappa r_i \sqrt{\ln N_{i+1}} + \varphi(r_i/12). \tag{25}$$

We now observe that this inequality is clearly valid also in the "rich" case, simply because we have then (23) and $\kappa r_{i-1} \sqrt{\ln N_i} \geqslant 0$.

Let $\mathbf{x} = (x_j)_{j \geqslant 0}$ denote a maximal branch in the tree $\mathcal{X}$: the node x_0 in the branch $\mathbf{x}$ is the root of $\mathcal{X}$, and x_{i+1} is a child of x_i for every $i \geqslant 0$. We may consider the elements in the branch as functions of $\mathbf{x}$ and write $x_i = x_i(\mathbf{x}) \succ (t_i(\mathbf{x}), r_i)$ for the node x_i at level i in the branch $\mathbf{x}$, with the control value $N_i(\mathbf{x}) = N(x_i(\mathbf{x}))$. Now, for any branch $\mathbf{x}$ and every $i \geqslant 0$, Equation (25) gives

$$\varphi(2r_i) + \frac{2}{3} \kappa r_{i-1} \sqrt{\ln N_i(\mathbf{x})} \geqslant \kappa r_i \sqrt{\ln N_{i+1}(\mathbf{x})} + \varphi(r_i/12). \tag{26}$$

Summing (26) from $i = 0$ to an arbitrary $k \geqslant 0$, using $r_{-1}\sqrt{\ln N_0} = 0$ because $N_0 = 1$, and reorganizing the root-of-log terms we get

$$\sum_{i=0}^{k} \varphi(2r_i) \geqslant \frac{\kappa}{3} \sum_{i=0}^{k} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} + \sum_{i=0}^{k} \varphi(r_i/12)$$

for any infinite branch $\mathbf{x}$. Observe that

$$\varphi(r_i/12) \geqslant \varphi(2r_{i+3}) \tag{27}$$

because $2r_{i+3} = 2r_i/27 < r_i/12$. Hence

$$\sum_{i=0}^{k} \varphi(r_i/12) \geqslant \sum_{j=3}^{k} \varphi(2r_j) \quad \text{thus} \quad 3\,\varphi(2r_0) \geqslant \sum_{i=0}^{2} \varphi(2r_i) \geqslant \frac{\kappa}{3} \sum_{i=0}^{k} r_i \sqrt{\ln N_{i+1}(\mathbf{x})}.$$

Finally, for every infinite branch $\mathbf{x}$ of $\mathcal{X}$ we have that

$$\sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{9}{\kappa}\, \varphi(2r_0) = \frac{9}{\kappa}\, E^* = 216\ \mathrm{E}\Big(\sup_{t \in T} X_t\Big). \tag{28}$$

The series above is very similar to the series $\Sigma_1(T)$ in (1). We let

$$\sigma_1(\mathbf{x}) := \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})},$$

a function of the branches $\mathbf{x}$, bounded by a multiple of the expectation of the supremum of the *invariant* process $(X_t)_{t \in T}$.

2.1.3. *Summary*

Let us review for future use the properties of our tree $\mathcal{X}$, its control function N and radius function ρ. For presenting these properties, let us denote by θ the function on $\mathcal{X}$ that associates to each node x its *latest position* $\theta(x) \in T$. For every node x in $\mathcal{X}$ with $x \succ (t, r) = (\theta(x), \rho(x))$, let us say that the ball $B(t, r)$ is the *ball associated to x*, and denote it by $\beta(x) = B(t, r)$. Recall that $C(x)$ denotes the set of children of x.

$\mathbf{a_0}$ — The root of the tree is $x_0 = (t_0)$, where $t_0 = \theta(x_0)$ is an arbitrary point in T. The radius $r_0 = \rho(x_0)$ is chosen such that $\Delta < r_0 < 4\Delta/3$, where Δ is the diameter of T; it implies that $\beta(x_0) = T$.

$\mathbf{a_1}$ — For every node $x \in \mathcal{X}$ with $x \succ (\theta(x), \rho(x))$, the ball $\beta(x) = B(\theta(x), \rho(x))$ associated to x is covered by the balls associated to the children $y \succ (\theta(y), \rho(y))$ of x,

$$\beta(x) \subset \bigcup_{y \in C(x)} \beta(y), \text{ and } \theta(y) \in B(\theta(x), \rho(x)),\ \rho(y) = \rho(x)/3 \text{ for } (18) \text{ any } y \in C(x).$$

$\mathbf{a_2}$ — The function $x \mapsto N(x)$ is a 3-control function on the tree $\mathcal{X}$, thus satisfies the properties $\mathbf{c_0}$, $\mathbf{c_1}$ and $\mathbf{c_2}$,

$$N(x_0) = 1, \quad y \in C(x_0) \Rightarrow N(y) \geqslant 2\,;$$
$$y, y' \in C(x) \ \Rightarrow\ N(y) = N(y')\,; \quad y \in C(x) \Rightarrow N(y) \geqslant N(x)^3\,;$$
$$y \in C(x) \Rightarrow |C(x)| \leqslant N(y).$$

We did not yet make use of the invariance condition (22). When defining the children of a node $x_* \in \mathcal{X}$ with $x_* \succ (t_*, r_*)$, we have set $r = r_*/3$ and we have then introduced a r-p-net of $\widetilde{N}$ points for the ball $B(t_*, r_*)$, where $\widetilde{N} = \widetilde{N}(x_*)$ could depend upon x_*, without using the fact that according to (22), the covering number $\mathcal{N}^*(B(t, r_*), r)$ *does not* depend upon $t \in T$. Using that condition, we can find a r-p-net with the *same* number $\widetilde{N}$ of points for every other ball $B(t, r_*)$ of that radius r_*. Doing this from the root x_0 on, we may replace the functions of branches $\mathbf{x} \mapsto N_i(\mathbf{x})$ by *constant values* N_i, for each $i \geqslant 0$, and write the conclusion of the tree construction as

$$\sigma_1 = \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} < +\infty. \tag{29}$$

2.2. *The entropy-like condition is sufficient*

Let $\mathcal{X}$ be a rooted tree associated to the index set T of an invariant centered Gaussian process $(X_t)_{t \in T}$, let N be a 3-control function for $\mathcal{X}$ and ρ the radius function. We assume that the nodes at level $i \geqslant 0$ in the tree $\mathcal{X}$ have the form $x_i = (t_0, t_1, \ldots, t_i)$ as in the previous sections, we set $r_i = \rho(x_i)$ and $N_i = N(x_i)$. We assume that $\mathcal{X}$, the functions N and ρ satisfy the three properties $\mathbf{a_0}$ to $\mathbf{a_2}$. In addition, we assume here that the values N_i do not depend on the branches. We shall explain that conversely, the condition $\sigma_1 < \infty$ in (29) allows one to bound the expectation of the supremum of the process $(X_t)_{t \in T}$.

We start from the root $x_0 \succ (t_0, r_0)$, that was chosen so that $B(t_0, r_0) = T$ (see the condition $\mathbf{a_0}$), and we move toward an arbitrary point $\tau \in T$ by successive steps in the tree, that will form an infinite branch $\mathbf{x}(\tau)$. To begin with, we set $x_0(\tau) = x_0$, we let $t_0(\tau) = t_0$, and we have that $\tau \in T = B(t_0(\tau), r_0)$. Next, the point τ is contained in at least one of the balls $B(t_1, r_1)$ corresponding to the last positions t_1 of the children of x_0, that cover $B(t_0, r_0)$ (this is condition $\mathbf{a_1}$); we let $x_1(\tau) \succ (t_1(\tau), r_1)$ be a child of $x_0(\tau) = x_0$ for which we have that $\tau \in B(t_1(\tau), r_1)$. Then, using $\mathbf{a_1}$ repeatedly, we go on choosing successive nodes $x_{i+1}(\tau) \succ (t_{i+1}(\tau), r_{i+1})$ such that $x_{i+1}(\tau)$ is a child of $x_i(\tau)$ for which we have again $\tau \in B(t_{i+1}(\tau), r_{i+1})$. We see that the sequence $(t_i(\tau))$ tends to τ in T because r_i tends to 0 ([19]) and $d(\tau, t_i(\tau)) < r_i$. We know that $t_{i+1}(\tau) \in B(t_i(\tau), r_i)$ using $\mathbf{a_1}$ again, and it yields that the step from $t_i(\tau)$ to $t_{i+1}(\tau)$ has size $< r_i$, we shall use it below.

The control function N takes the same value N_i on all the nodes in the ith generation of the tree $\mathcal{X}$. If we fix $i \geqslant 0$ and consider the finite subtree

$$\mathcal{X}_i = \{x \in \mathcal{X} : N(x) \leqslant N_i\}$$

we know by Lemma 5 that there are at most $N_i^{3/2}$ leaves in that subtree, and clearly, the leaves of $\mathcal{X}_i$ are precisely the nodes in the ith generation of $\mathcal{X}$. Hence, when τ varies in T and when $i > 0$ is fixed, the global number M_i of paths from x_0 to all the nodes $x_i(\tau)$ admits ([20]) the bound

$$M_i \leqslant N_i^{3/2} < N_i^2 \tag{30}$$

(if $i > 0$ then $N_i > 1$). Because $\operatorname{Var}(X_{t_{i+1}(\tau)} - X_{t_i(\tau)}) = d(t_{i+1}(\tau), t_i(\tau))^2 \leqslant r_i^2$ and using (9), we have for each integer $i \geqslant 0$ that

$$\mathrm{E} \sup_{\tau \in T} |X_{t_{i+1}(\tau)} - X_{t_i(\tau)}| \leqslant r_i \,.\, 2 \sqrt{\ln(M_{i+1} + 1)} \leqslant 2 r_i \sqrt{\ln N_{i+1}^2} < 3 r_i \sqrt{\ln N_{i+1}}.$$

Every $\tau \in T$ is the limit of a sequence $(t_i(\tau))_{i \geqslant 0}$ and $t_0(\tau) = t_0$, hence

$$X_\tau - X_{t_0} = \sum_{i=0}^{\infty} (X_{t_{i+1}(\tau)} - X_{t_i(\tau)}).$$

It follows that

$$\mathrm{E} \sup_{\tau \in T} |X_\tau - X_{t_0}| \leqslant \sum_{i=0}^{\infty} \mathrm{E} \sup_{\tau \in T} |X_{t_{i+1}(\tau)} - X_{t_i(\tau)}| \leqslant 3 \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}}$$

thus, knowing that $\mathrm{E}\, X_{t_0} = 0$, we conclude that

$$\mathrm{E}\Bigl(\sup_{t \in T} X_t\Bigr) \leqslant 3 \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} = 3 \sigma_1, \qquad \mathrm{E}\Bigl(\sup_{t \in T} |X_t|\Bigr) \leqslant \mathrm{E}\, |X_{t_0}| + 3 \sigma_1.$$

Of course what we just did is nothing but a variant of proving that the finiteness of the Dudley integral related to a Gaussian process implies that the expectation of its supremum is finite. We could actually have easily related directly the condition

$$\sigma_1 = \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} < +\infty$$

to Dudley's integral. Indeed, we have seen at (30) that there are $M_i \leqslant N_i^2$ paths from the root x_0 to the nodes of the ith generation in the tree; in other words, the set T_i of points $t_i \in T$ appearing in the nodes (t_i, r_i) of that ith generation contains at most N_i^2 points, and we know by iterating condition $\mathbf{a_1}$ that the balls $B(t_i, r_i)$ centered at those points $t_i \in T_i$ cover T. This means that

$$\mathcal{N}(r_i) \leqslant M_i \leqslant N_i^2,$$

where $\mathcal{N}(\varepsilon) = \mathcal{N}(T, \varepsilon)$ denotes the minimal number of open balls of radius $\varepsilon > 0$ needed to cover T. When $r_{i+1} \leqslant \varepsilon \leqslant r_i$ we thus have

$$\mathcal{N}(\varepsilon) \leqslant \mathcal{N}(r_{i+1}) \leqslant N_{i+1}^2,$$

hence

$$\int_{r_{i+1}}^{r_i} \sqrt{\ln \mathcal{N}(\varepsilon)}\, \mathrm{d}\varepsilon \leqslant r_i \sqrt{\ln(N_{i+1}^2)} < 2 r_i \sqrt{\ln N_{i+1}}$$

and because $r_0 > \Delta$ and $\lim_i r_i = 0$, we conclude using (28) that

$$I_D(T) = \int_0^\Delta \sqrt{\ln \mathcal{N}(\varepsilon)}\, \mathrm{d}\varepsilon < 2 \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} = 2\sigma_1 \leqslant 432\, \mathrm{E}\Bigl(\sup_{t \in T} X_t\Bigr).$$

These are pretty much the same arguments as those that we have seen before, when we discretized the Dudley integral.

3. *Beyond invariance*

Consider again a centered Gaussian process $(X_t)_{t \in T}$ such that

$$E^* = \mathrm{E}\Bigl(\sup_{t \in T} X_t\Bigr) < \infty.$$

If we give up invariance for the process $(X_t)_{t \in T}$, we lose Equality (21). Then, for each point $s \in T$ and $r > 0$, we can do nothing but consider

$$\varphi(s, r) = \mathrm{E}\Bigl(\sup_{t \in B(s,r)} X_t\Bigr) \leqslant E^*.$$

Without invariance and applying Corollary 1, it only remains from (23), when $i = 0$ for example, that

$$\begin{aligned} \varphi(t_0, 2r_0) = \mathrm{E}\Bigl(\sup_{t \in B(t_0, 2r_0)} X_t\Bigr) &\geqslant (r_0/24)\sqrt{\ln N_1} + \min_{s \in S_1} \mathrm{E}\Bigl(\sup_{t \in B(s, r_0/12)} X_t\Bigr) \\ &= (r_0/24)\sqrt{\ln N_1} + \min_{s \in S_1} \varphi(s, r_0/12). \end{aligned}$$

This cannot be used, unless we can make sure that the different values $\varphi(s, r_0/12)$, for $s \in S_1$, are "under some control", let say for simplicity that they are essentially the same. For this we need a quite natural extra step in the construction of the tree. After finding generation i and before defining the $(i{+}1)$th, we have to make divisions into "zones" where the function φ will be controlled in a suitable way. Now, the collection of children of a node x_i in the ith generation will possibly be split into different sub-collections corresponding to these "zones". Also, we shall need an improving control over φ as i increases, by locating values of φ in smaller and smaller intervals of the real line; in order to avoid multiplying unnecessarily the number of parameters of the construction, we shall use([21]) the rapidly growing numbers $N_i = N(x_i)$ to this end, locating values of φ at level i in small intervals of size N_i^{-1}.

We shall thus divide the successive sets obtained in the invariant case —there, they were the balls $B(t_i, r_i)$— into "homogeneous zones". The nodes at level $i \geqslant 0$ in the new tree $\mathcal{X}$ will now consist of sequences

$$x = (V, t_0, t_1, \ldots, t_i) \ \text{ where } \ t_j \in T, \ \ j = 0, 1, \ldots, i,$$

and where V is a non-empty subset of T, a "homogeneous zone" where a certain variation of φ will be controlled, and $t_0, t_1, \ldots, t_i$ are as before the successive positions of x and its ancesters. We say that V is the *region* associated to x, the point t_i is the *latest position* and we have as before a radius $r_i = \rho(x)$ associated to each node. It will be useful to define the function θ on $\mathcal{X}$ that associates to x its latest position $t_i = \theta(x)$, and the function R that indicates the region $V = R(x)$. We shall now denote the extraction of the main information V, t_i, r_i associated to x by writing

$$x \succ (V, t_i, r_i), \quad \text{again with} \quad r_i = 3^{-i} r_0.$$

We could (mostly uselessly) write $x \succ (R(x), \theta(x), \rho(x))$. The set V will be contained in the ball $B(t_i, r_i)$ of radius $r_i > 0$ and center $t_i \in T$. The new control function will be defined in two steps: we first find a bound N on the entropy of V, and then a bound in N^2 for the number of siblings of x, so that the actual control function will be N^2. The function N will be called the *growing function* for the tree, it will satisfy the conditions $\mathbf{c_0}$ and $\mathbf{c_1}$, and N^2 will be the *3-control function*, satisfying $\mathbf{c_2}$ in addition.

3.1. *A new tree*

The construction of the tree $\mathcal{X}$ and of the growing function N (and therefore of the control N^2) goes as follows: we suppose of course that T has at least two points, so that its diameter Δ is > 0, and finite(2). We choose r_0 such that $\Delta < r_0 < 4\Delta/3$ and t_0 a point in T, hence we have $T = B(t_0, r_0)$. The root of our tree is $x_0 = (V_0, t_0)$ with $V_0 = T$. We let $\rho(x_0) = r_0$ and $N_0 = N(x_0) = 1$. So far this is exactly as before —save for the addition of V_0 to the singleton (t_0)— and it is still convenient to let $r_{-1} = 3r_0$. Let us describe the modified construction step.

Suppose that a node x_* has been introduced in the tree $\mathcal{X}$, with $x_* \succ (V_*, t_*, r_*)$ and growth value $N_* = N(x_*)$. The first part of the construction step is essentially identical to what was done in the invariant case, except that the set V_* replaces now what was the ball $B(t_*, r_*)$ before, and that the control will be defined differently, in two steps, first the *growth value*, and later the control value: let

$$r = r_*/3 \text{ and let } S \text{ be a } r\text{-p-net for } V_*,$$

that is to say, the set $S \subset V_*$ is a r-separated set such that

$$V_* \subset \bigcup_{s \in S} B(s, r). \tag{31}$$

We let $\widetilde{N} = \max(N_*^3, |S|)$, this will be the growth value for all the children of x_*. Again, the *rich case* is when $\widetilde{N} = |S| > N_*^3$, and otherwise we have the *poor case* $\widetilde{N} = N_*^3$.

Here comes the difference: in the invariant case, the children of x_* corresponded to the points s in S and were thus directly associated to the balls $B(s, r)$; now, for each $s \in S$, we make a further splitting of the ball $B(s, r)$, arising from a splitting of T: for every integer $n \geqslant 1$ and α such that $0 \leqslant \alpha < n$, we consider the (possibly empty) subset of T defined by

$$W_\alpha(r, n) = \{v \in T : \varphi(v, r/12) \in Z(\alpha, n)\}$$

where

$$Z(\alpha, n) = [\alpha E^*/n, (\alpha + 1)E^*/n] \subset [0, E^*].$$

Notice that

$$T = \bigcup_{0 \leqslant \alpha < n} W_\alpha(r, n), \tag{32}$$

since the segments $Z(\alpha, n)$ cover $[0, E^*]$ and $0 \leqslant \varphi \leqslant E^*$.

Coming back to the construction, the sets $W_\alpha(r,\widetilde{N})$ will be the "homogeneous zones" that were mentioned earlier. We shall apply the further division induced by these sets $W_\alpha(r,\widetilde{N})$: the first component V of the children of x_* will be contained in one of the intersections $B(s,r)\cap W_\alpha(r,\widetilde{N})$. Precisely, the children x of $x_* \succ (V_*,t_*,r_*)$ will satisfy $x \succ (V,s,r)$, where

$$r = r_*/3, \quad V = V_* \cap B(s,r) \cap W_\alpha(r,\widetilde{N}) \neq \emptyset, \quad s \in S, \ 0 \leqslant \alpha < \widetilde{N}.$$

For every child $x \succ (V,s,r)$ of x_*, we have $s \in V_*$ and $V \subset V_*$ by construction, and we let $N(x) = \widetilde{N}$. The number of children of x_* is less than or equal to $|S|\widetilde{N} \leqslant \widetilde{N}^2$, we shall define $\widetilde{N}^2 = N^2(x)$ to be the control value for all those children x.

Intersecting the "coverings" in (31) and (32), we see that

$$V_* \subset \bigcup \{ B(s,r) \cap W_\alpha(r,\widetilde{N}) : s \in S, \ 0 \leqslant \alpha < \widetilde{N} \}, \tag{33}$$

which implies that the sets V corresponding to all children $x \succ (V,s,r)$ of x_* cover V_*.

Let us check that the function N is a 3-growing function on the tree $\mathcal{X}$: for the root $x_0 \succ (V_0,t_0,r_0)$, we have $V_0 = T$, $N(x_0) = 1$ and $\Delta < r_0 < 4\Delta/3$. It yields that $\mathcal{N}^*(V_0,r_0/3) = \mathcal{N}^*(T,r_0/3) > 1$, since the inequality $2(r_0/3) < \Delta$ implies that $B(t,r_0/3) \neq T$ for any point $t \in T$. We have thus $N_1 \geqslant 2$, the condition $\mathbf{c_0}$ for a growing function is satisfied. The function N was defined in a way that it takes on the same value $N(y)$ for all children y of a node x and $N(y) \geqslant N(x)^3$, thus N satisfies the condition $\mathbf{c_1}$, and is therefore a 3-growing function. It is obvious that N^2 also satisfies $\mathbf{c_0}$ and $\mathbf{c_1}$ and furthermore, we have $|C(x)| \leqslant N^2(y)$ when $y \in C(x)$, the condition $\mathbf{c_2}$ for N^2 to be a control function is satisfied.

Let us repeat things with a mention to the level $i \geqslant 0$ in the new tree of the parent node

$$x_* = x_i = (V_i, t_0, t_1, \ldots, t_i)$$

with growth value $N_i = N(x_i)$. We let $r_i = \rho(x_i) = 3^{-i} r_0$ and $r_{i+1} = r = r_i/3$, we have a set $S_{i+1} = S$ that is a r_{i+1}-p-net for V_i, and we let $N_{i+1} = \widetilde{N} = \max(N_i^3, |S_{i+1}|)$. Then the set V_i is divided into homogeneous zones V_{i+1}, the children x_{i+1} of x_i have the form

$$x_{i+1} = (V_{i+1}, t_0, t_1, \ldots, t_i, s),$$

with

$$V_{i+1} = V_i \cap B(s, r_{i+1}) \cap W_\alpha(r_{i+1}, N_{i+1}) \neq \emptyset, \quad s \in S_{i+1}, \ 0 \leqslant \alpha < N_{i+1},$$

and we have that

$$R(x_{i+1}) = V_{i+1}, \quad \theta(x_{i+1}) = s, \quad \rho(x_{i+1}) = r_{i+1} = \rho(x_i)/3), \quad N(x_{i+1}) = N_{i+1}.$$

The number of children of x_i is bounded by the control value N_{i+1}^2. By Equation (33), we know that the sets V_{i+1} corresponding to all the children of x_i cover V_i. For every child $x_{i+1} \succ (V_{i+1}, t_{i+1}, r_{i+1})$ of $x_i \succ (V_i, t_i, r_i)$, we have

$$V_{i+1} \subset V_i, \quad t_{i+1} \in V_i, \quad V_{i+1} \subset B(t_{i+1}, r_{i+1}).$$

By construction, when v runs in the set V_{i+1}, the values of $\varphi(v, r_{i+1}/12)$ stay in a segment of length E^*/N_{i+1}. For each node $x \succ (V,t,r)$ in $\mathcal{X}$ let

$$z(x) = \inf \{ \varphi(v, r/12) : v \in V \} \geqslant 0. \tag{34}$$

From what we have just said, we know when $x_{i+1} \succ (V_{i+1}, t_{i+1}, r_{i+1})$ that

$$v \in V_{i+1} \ \Rightarrow \ z(x_{i+1}) \leqslant \varphi(v, r_{i+1}/12) \leqslant z(x_{i+1}) + E^*/N_{i+1},$$

and we also have

$$v \in V_0 \;\Rightarrow\; z(x_0) \leqslant \varphi(v, r_0/12) \leqslant z(x_0) + E^*/N_0,$$

because $N_0 = 1$ and $0 \leqslant \varphi \leqslant E^*$.

In the invariant case, the function $t \mapsto \varphi(t, r_{i+1}/12)$ is constant on T and in that situation, the above procedure produces exactly *one* homogeneous zone inside the ball $B(s, r_{i+1})$, for each $s \in S_{i+1}$, namely, the ball $B(s, r_{i+1})$ itself: there is only one set $V_{i+1} = V_i \cap B(s, r_{i+1})$ for each given point $s \in S_{i+1}$.

3.1.1. *Estimates*

Suppose that $i \geqslant 0$ and that x_i is a given node in the ith generation of the new tree $\mathcal{X}$, with $x_i \succ (V_i, t_i, r_i)$ and growth value $N_i = N(x_i)$. Recall that r_{i+1}-p-net S_{i+1} for V_i with $r_{i+1} = r_i/3$ has been introduced, and that $N_{i+1} = \max(N_i^3, |S_{i+1}|)$ is the growth value for all the children of x_i in the next generation $(i{+}1)$. There are as before two possibilities:

— in the “rich case”, we have $N_{i+1} = |S_{i+1}| > N_i^3$ and we know that the set S_{i+1} is r_{i+1}-separated, contained in $B(t_i, r_i)$ because $S_{i+1} \subset V_i \subset B(t_i, r_i)$; therefore, by Corollary 1 applied with $2\delta = r_{i+1}$ —and thus $\delta/2 = r_{i+1}/4 = r_i/12$— we obtain

$$\begin{aligned}\varphi(t_i, r_i{+}r_i/12) &= \mathrm{E}\Bigl(\sup_{t \in B(t_i, r_i+r_i/12)} X_t\Bigr)\\ &\geqslant (r_i/24)\sqrt{\ln N_{i+1}} + \min_{s \in S_{i+1}} \mathrm{E}\Bigl(\sup_{t \in B(s, r_i/12)} X_t\Bigr)\\ &\geqslant \kappa r_i \sqrt{\ln N_{i+1}} + \inf_{v \in V_i} \varphi(v, r_i/12) = \kappa r_i \sqrt{\ln N_{i+1}} + z(x_i),\end{aligned}$$

by the definition (34) of $z(x_i)$; we just lazily write and keep in mind that

$$\varphi(t_i, 2r_i) \geqslant \kappa r_i \sqrt{\ln N_{i+1}} + z(x_i)\,;$$

— otherwise, we are in the poor case, thus $N_{i+1} = N_i^3$. On one hand, we use as before $\frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i} \geqslant \kappa r_i \sqrt{\ln N_{i+1}}$ —see Equation (24)—; on the other hand, given $v \in V_i$, we have

$$z(x_i) \leqslant \varphi(v, r_i/12) \leqslant \varphi(t_i, r_i + r_i/12) \leqslant \varphi(t_i, 2r_i)$$

because $B(v, r_i/12) \subset B(t_i, r_i{+}r_i/12)$ since $v \in V_i \subset B(t_i, r_i)$.

In both cases, we conclude that

$$\varphi(t_i, 2r_i) + \frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i} \geqslant \kappa r_i \sqrt{\ln N_{i+1}} + z(x_i). \tag{35}$$

This replaces Inequality (25) from the invariant case.

3.1.2. *Adding up*

Let us consider a branch $\mathbf{x} = (x_i)_{i \geqslant 0}$ of the tree, with $x_i = x_i(\mathbf{x}) \succ (V_i(\mathbf{x}), t_i(\mathbf{x}), r_i)$ and growth value $N_i(\mathbf{x})$ for each $i \geqslant 0$, where

$$V_i(\mathbf{x}) = R(x_i(\mathbf{x})),\;\; t_i(\mathbf{x}) = \theta(x_i(\mathbf{x})),\;\; r_i = \rho(x_i(\mathbf{x})),\;\; N_i(\mathbf{x}) = N(x_i(\mathbf{x})).$$

Let us write the values $x_i(\mathbf{x})$, $t_i(\mathbf{x})$ and $N_i(\mathbf{x})$ without their $\mathbf{x}$ variable, but remember that there is a branch $\mathbf{x}$ that remains fixed in the following lines. Adding (35) from $i = 0$ to an arbitrary $k \geqslant 0$ and reorganizing as before the root-of-log terms we get

$$\sum_{i=0}^{k} \varphi(t_i, 2r_i) \geqslant \frac{\kappa}{3} \sum_{i=0}^{k} r_i \sqrt{\ln N_{i+1}} + \sum_{i=0}^{k} z(x_i). \tag{36}$$

Inequality (27) has to be revised, we argue now as follows: let $x_i \succ (V_i, t_i, r_i)$ be the node of $\mathbf{x}$ at level $i \in \mathbb{N}$; the set $V_i \subset B(t_i, r_i)$ is an "homogeneous" subset, meaning precisely that $z(x_i) \leqslant \varphi(v, r_i/12) \leqslant z(x_i) + E^*/N_i$ for all $v \in V_i$. One then has that

$$\varphi(t_{i+3}, 2r_{i+3}) \leqslant z(x_i) + E^*/N_i\,; \tag{37}$$

indeed, we see that $2r_{i+3} = 2r_i/27 < r_i/12$, and we know that $t_{i+3} \in V_{i+2} \subset V_i$ (22), therefore

$$\varphi(t_{i+3}, 2r_{i+3}) \leqslant \varphi(t_{i+3}, r_i/12) \leqslant z(x_i) + E^*/N_i.$$

It follows from (37) that

$$\sum_{i=0}^{k} z(x_i) + \sum_{i=0}^{k} E^*/N_i \geqslant \sum_{j=3}^{k} \varphi(t_j, 2r_j)$$

and using (36) we get

$$\sum_{i=0}^{k} \varphi(t_i, 2r_i) + \sum_{i=0}^{k} E^*/N_i \geqslant \frac{\kappa}{3} \sum_{i=0}^{k} r_i \sqrt{\ln N_{i+1}} + \sum_{j=3}^{k} \varphi(t_j, 2r_j).$$

Because $\varphi \leqslant E^*$, then recalling (18) we obtain

$$5E^* \geqslant 3\varphi(t_0, 2r_0) + 2E^* > \sum_{i=0}^{2} \varphi(t_i, 2r_i) + \sum_{i=0}^{k} E^*/N_i \geqslant \frac{\kappa}{3} \sum_{i=0}^{k} r_i \sqrt{\ln N_{i+1}}.$$

We conclude that

$$\sigma_1(\mathbf{x}) := \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{15}{\kappa}\, E^* = 360\, \mathrm{E}\Bigl(\sup_{t \in T} X_t\Bigr) \tag{38}$$

for every branch $\mathbf{x}$ of the tree $\mathcal{X}$.

For every node $x \succ (R(x), \theta(x), \rho(x))$ in $\mathcal{X}$, we said that the set $R(x)$ is the *region associated to the node x*. Let us review the properties of the tree and the functions ρ and N.

$\mathbf{a_0^*}$ — (compare to $\mathbf{a_0}$) The root of the tree is $x_0 = (V_0, t_0)$, where $V_0 = T$ and where $t_0 = \theta(x_0)$ is a point in T. The radius $r_0 = \rho(x_0)$ is such that $\Delta < r_0 < 4\Delta/3$, where Δ is the diameter of T. It implies that $B(t_0, r_0) = T = V_0$.

$\mathbf{a_1^*}$ — (see $\mathbf{a_1}$) For every node $x \succ (R(x), \theta(x), \rho(x)) \in \mathcal{X}$, the region $R(x)$ associated to x is the union of the regions associated to the children $y \succ (R(y), \theta(y), \rho(y))$ of x:

$$R(x) = \bigcup_{y \in C(x)} R(y), \quad \text{and} \quad \theta(y) \in R(x) \subset B(\theta(x), \rho(x)),\ \ \rho(y) = \rho(x)/3.$$

In particular, we see that $R(x_{i+1}(\mathbf{x})) \subset R(x_i(\mathbf{x}))$ for every branch $\mathbf{x} = (x_i(\mathbf{x}))_{i \in \mathbb{N}}$ and every integer $i \geqslant 0$.

$\mathbf{a_2^*}$ — (see $\mathbf{a_2}$) The function N is a 3-growing function for $\mathcal{X}$, and the function N^2 is a 3-control function for $\mathcal{X}$:

$$N(x_0) = 1, \quad y \in C(x_0) \Rightarrow N(y) \geqslant 2\,;$$
$$y, y' \in C(x) \ \Rightarrow\ N(y) = N(y')\,; \qquad y \in C(x) \Rightarrow N(y) \geqslant N(x)^3\,;$$
$$y \in C(x) \Rightarrow |C(x)| \leqslant N^2(y).$$

For every branch $\mathbf{x} = (x_i(\mathbf{x}))_{i \in \mathbb{N}}$ and every $i \geqslant 0$, it follows that $N(x_i(\mathbf{x})) \geqslant 2^i$ (see (18)).

3.2. *Suppose we have that nice tree*

Let $\mathcal{X}$ be a tree with nodes $x \succ (V, t, r)$ where $r > 0$, t belongs to the index set T of a centered Gaussian process $(X_t)_{t \in T}$, and V is a non-empty subset of T. Suppose that the tree and the growth function N on $\mathcal{X}$ satisfy the properties $\mathbf{a_0^*}$ to $\mathbf{a_2^*}$. If $\mathbf{x} = (x_j(\mathbf{x}))_{j \geqslant 0}$ is a branch in the tree and $i \geqslant 0$, let

$$V_i(\mathbf{x}) = R(x_i(\mathbf{x})), \ \ t_i(\mathbf{x}) = \theta(x_i(\mathbf{x})), \ \ r_i = \rho(x_i(\mathbf{x})), \ \ N_i(\mathbf{x}) = N(x_i(\mathbf{x})).$$

For a node x_i at level $i \geqslant 0$, we also write $x_i \succ (V_i, t_i, r_i)$ and $N_i = N(x_i)$. Assume that for some constant K_1 and for every branch $\mathbf{x}$ in the tree $\mathcal{X}$, we have

$$\sigma_1(\mathbf{x}) = \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant K_1.$$

We want to use this information in order to bound the expectation of the supremum of the process $(X_t)_{t \in T}$.

When τ is a point in T, we can again find a branch $\mathbf{x}$ such that $t_i(\mathbf{x})$ tends to τ, and more precisely, such that $\tau \in V_i(\mathbf{x}) \subset B(t_i(\mathbf{x}), r_i)$ for every integer $i \geqslant 0$, where we write $x_i(\mathbf{x}) \succ (V_i(\mathbf{x}), t_i(\mathbf{x}), r_i)$: we let $x_0(\mathbf{x}) = x_0$ be the root of the tree; then we have $\tau \in V_0 = T$. Assuming that $x_i = x_i(\mathbf{x})$ has been found such that $\tau \in V_i(\mathbf{x}) = R(x_i)$, we can find a child x_{i+1} of x_i such that $\tau \in R(x_{i+1})$, because by condition $\mathbf{a_1^*}$, these regions $R(x_{i+1})$ cover $R(x_i)$ when x_{i+1} varies in $C(x_i)$. We then let $x_{i+1}(\mathbf{x}) = x_{i+1}$, going on with the construction of the branch $\mathbf{x} = \mathbf{x}(\tau)$.

Let us fix a level $i \in \mathbb{N}$ in the tree. Consider a node $x_i \succ (V_i, t_i, r_i)$ from the ith generation. Let us first estimate the probability of a (relatively) *large jump* when passing from x_i to a specific fixed child $x_{i+1} \succ (V_{i+1}, t_{i+1}, r_{i+1})$ of x_i, that is to say, the probability of having a big absolute value for the difference $X_{t_{i+1}} - X_{t_i}$ of the two values of the process $(X_t)_{t \in T}$ at t_{i+1} and t_i. Consider $u > 0$ and let the size of the jump be

$$s_i(u, x_i) = \bigl(u + \sqrt{8 \ln N_{i+1}}\,\bigr) r_i,$$

where $N_{i+1} = N(x_{i+1})$ depends on the parent x_i only, by condition $\mathbf{a_2^*}$. We shall recall it by writing $N(x_{i+1}) = N_{i+1}[x_i]$. We know that $t_{i+1} \in V_i \subset B(t_i, r_i)$ by condition $\mathbf{a_1^*}$. We have therefore $d(t_{i+1}, t_i) < r_i$, so the variance of $(X_{t_{i+1}} - X_{t_i})/r_i$ is < 1, yielding

$$\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > \bigl(u + \sqrt{8 \ln N_{i+1}}\,\bigr) r_i\bigr) \leqslant \mathrm{P}\bigl(|g| > u + \sqrt{8 \ln N_{i+1}}\,\bigr)$$

where g is a $\mathrm{N}(0,1)$ Gaussian random variable, and by (7) we have

$$\begin{aligned}\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > s_i(u, x_i)\bigr) &\leqslant \mathrm{P}\bigl(|g| > u + \sqrt{8 \ln N_{i+1}}\,\bigr) \leqslant \exp\bigl(-\bigl(u + \sqrt{8 \ln N_{i+1}}\,\bigr)^2/2\bigr) \\ &\leqslant \exp\bigl(-u^2/2 - 4 \ln N_{i+1}\bigr) = \mathrm{e}^{-u^2/2}\, N_{i+1}^{-4}.\end{aligned}$$

Note that $N_{i+1} \geqslant 2^{i+1}$ by condition $\mathbf{a_2^*}$. We rewrite the last two lines as

$$\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > s_i(u, x_i)\bigr) \leqslant \mathrm{e}^{-u^2/2}\, N_{i+1}^{-1} N_{i+1}^{-3} \leqslant \mathrm{e}^{-u^2/2}\, 2^{-i-1} N_{i+1}^{-3} = c_i(u) N_{i+1}^{-3}$$

where we have set $c_i(u) = 2^{-i-1}\, \mathrm{e}^{-u^2/2}$. The total probability $p_i^{(i+1)}(x_i)$ of having a jump of that size $s_i(u, x_i)$ when passing from x_i to *any one* of its children in the $(i+1)$th generation (there are at most $N_{i+1}^2[x_i]$ of them by $\mathbf{a_2^*}$) is thus bounded by

$$p_i^{(i+1)}(x_i) \leqslant N_{i+1}^2 \,.\, c_i(u) N_{i+1}^{-3} = c_i(u) N_{i+1}^{-1}[x_i].$$

Using $N_{k+1} \geqslant N_k^3$ for every integer $k \geqslant 0$ (condition $\mathbf{a}_3^*$ again) we conclude that

$$p_i^{(i+1)}(x_i) \leqslant c_i(u)\, N_i^{-3}[x_{i-1}], \tag{39}$$

where x_{i-1} is the parent of x_i. For $0 \leqslant j \leqslant i$, let $p_j^{(i+1)}(x_j)$ denote the probability that starting from $x_j \in \mathcal{X}$ at level j along any possible path in the tree of the form

$$x_j, x_{j+1}, \ldots, x_i, x_{i+1},$$

ending at an arbitrary descendant x_{i+1} of x_j in the $(i+1)$th generation, we have a jump of size $\geqslant s_i(u, x_i)$ between the levels i and $(i+1)$ in one of those paths. By induction, starting from the inequality (39) that corresponds to $k = i$, we shall estimate $p_k^{(i+1)}(x_k)$ for any node x_k at level k, for k going down from $k = i - 1$ to $k = 1$.

Fix k such that $1 \leqslant k < i$. Suppose that $p_{k+1}^{(i+1)}(x_{k+1}) \leqslant c_i(u)\, N_{k+1}^{-3}[x_k]$ for every node $x_k \in \mathcal{X}$ at level k and any child $x_{k+1} \in C(x_k)$. Knowing by $\mathbf{a_2^*}$ that x_k has at most $N_{k+1}^2[x_k]$ children, we see that

$$p_k^{(i+1)}(x_k) \leqslant N_{k+1}^2[x_k] . c_i(u)\, N_{k+1}^{-3}[x_k] = c_i(u)\, N_{k+1}^{-1}[x_k] \leqslant c_i(u)\, N_k^{-3}[x_{k-1}],$$

where x_{k-1} is the parent of x_k. In this way, we arrive at $p_1^{(i+1)}(x_1) \leqslant c_i(u)\, N_1^{-3}[x_0]$ for every $x_1 \in C(x_0)$, and finally with $N_1 = N_1[x_0] \geqslant 2$ we have

$$p_0^{(i+1)}(x_0) \leqslant N_1^2 . c_i(u) N_1^{-3} = c_i(u) N_1^{-1} < c_i(u) = 2^{-i-1}\, \mathrm{e}^{-u^2/2} .$$

This is a bound for the probability of a jump of size $s_i(u, x_i)$ between the ith generation and the next, for any path from the root to a node in the $(i+1)$th generation. Starting from the root, the first jump to consider is $s_0(u, x_0)$ between x_0 and its children x_1, so the probability $p(u)$ of finding for some $i \geqslant 0$ a jump of size $s_i(u, x_i)$ between the levels i and $(i+1)$ in the tree is bounded above by

$$p(u) \leqslant \sum_{i=0}^{\infty} p_0^{(i+1)}(x_0) \leqslant \mathrm{e}^{-u^2/2} \sum_{i=0}^{\infty} 2^{-i-1} = \mathrm{e}^{-u^2/2} .$$

Except for a set $S(u) \subset \Omega$ having probability $\leqslant p(u)$, we obtain that when $\omega \notin S(u)$, *all* steps

$$|X_{t_{i+1}(\mathbf{x})}(\omega) - X_{t_i(\mathbf{x})}(\omega)|, \qquad i \geqslant 0,$$

along *any* branch $\mathbf{x}$ are less than $s_i(u, \mathbf{x}) := s_i(u, x_i(\mathbf{x}))$, so that outside $S(u)$, we have

$$\begin{aligned} \sum_{i=0}^{\infty} |X_{t_{i+1}(\mathbf{x})} - X_{t_i(\mathbf{x})}| &\leqslant \sum_{i=0}^{\infty} s_i(u, \mathbf{x}) \\ &\leqslant \Bigl(\sum_{i=0}^{\infty} r_i\Bigr) u + \sum_{i=0}^{\infty} r_i \sqrt{8 \ln N_{i+1}(\mathbf{x})} \\ &< (3/2) r_0\, u + 3K_1 \leqslant 2\Delta u + 3K_1. \end{aligned} \tag{40}$$

We know that every $\tau \in T$ is the limit of the sequence $(t_i(\mathbf{x}))_{i \geqslant 0}$ for a certain branch $\mathbf{x}$, and $t_0(\mathbf{x}) = t_0$, hence

$$X_\tau - X_{t_0} = \sum_{i=0}^{\infty} (X_{t_{i+1}(\mathbf{x})} - X_{t_i(\mathbf{x})}).$$

It follows from (40) that for every $u > 0$, we have

$$\mathrm{P}\bigl(\sup_{\tau \in T} |X_\tau - X_{t_0}| > 2\Delta u + 3K_1\bigr) \leqslant \mathrm{P}(S(u)) \leqslant \mathrm{e}^{-u^2/2} .$$

This certainly implies that

$$\mathrm{E}\Bigl(\sup_{t \in T} |X_t - X_{t_0}|\Bigr) < \infty,$$

and more precisely, letting $X^* = \sup_{t \in T} |X_t - X_{t_0}|$, we obtain that

$$\begin{aligned} \mathrm{E}\, X^* &= \int_0^\infty \mathrm{P}(X^* > v)\,\mathrm{d}v \leqslant 3K_1 + \int_{3K_1}^\infty \mathrm{P}(X^* > v)\,\mathrm{d}v \\ &= 3K_1 + \int_0^\infty \mathrm{P}(X^* > 3K_1 + v)\,\mathrm{d}v = 3K_1 + 2\Delta \int_0^\infty \mathrm{P}(X^* > 3K_1 + 2\Delta u)\,\mathrm{d}u \\ &\leqslant 3K_1 + 2\Delta \int_0^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u = 3K_1 + \sqrt{2\pi}\,\Delta. \end{aligned}$$

Also,

$$\Delta\sqrt{\ln 2} < r_0\sqrt{\ln N_1} < K_1$$

so that

$$\mathrm{E}\, X^* \leqslant 3K_1 + \sqrt{2\pi}\,\Delta \leqslant \Bigl(3 + \sqrt{\frac{2\pi}{\ln 2}}\Bigr) K_1 < 7\,K_1.$$

3.2.1. *Only the rich really count*

— This is just a remark in passing. Consider a branch $\mathbf{x}$ in $\mathcal{X}$ and suppose that a path

$$x_{j+1}, x_{j+2}, \ldots, x_k$$

in that branch $\mathbf{x}$ consists only of *poor nodes*. By Equation (24) we have

$$r_i\sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{2}{3} r_{i-1}\sqrt{\ln N_i(\mathbf{x})}$$

for every i between $j+1$ and k, thus

$$\sum_{i=j+1}^k r_i\sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \Bigl(\sum_{m=1}^\infty \Bigl(\frac{2}{3}\Bigr)^m\Bigr) r_j\sqrt{\ln N_{j+1}(\mathbf{x})} = 2r_j\sqrt{\ln N_{j+1}(\mathbf{x})}.$$

This shows that for every branch $\mathbf{x}$,

$$\sum_{i=0}^\infty r_i\sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant 3 \sum_{x_j \text{ rich}} r_j\sqrt{\ln N_{j+1}(\mathbf{x})}.$$

— We could have tried to relate the construction of the "new tree" in section 3 to the successive balls introduced in the invariant case. Indeed, we can construct the tree $\mathcal{X}$ of Section 2.1.1 regardless of invariance. That would provide us with a system of balls $B(t_i, r_i)$ with the properties $\mathbf{a_0}$, $\mathbf{a_1}$ and $\mathbf{a_2}$. Next, we can try to build the new tree $\mathcal{X}^*$ by observing that the balls for $\mathcal{X}$ give coverings of the sets V_i —but the centers may be outside V_i—. There is a difficulty here: suppose that several balls $B(t_i, r_i)$ related to $\mathcal{X}$ meet an homogeneous zone V for $\mathcal{X}^*$ at points s_i; then it may be impossible to choose these s_i to be well separated. Also, the number N_i in section 2.1.1 refers to the global covering of $B(t_{i-1}, r_{i-1})$ by balls $B(t_i, r_i)$, while in the new tree $\mathcal{X}^*$ the value N_i^* refers to a single set V_{i-1}.

3.3. *Majorizing measure*

We continue with the tree $\mathcal{X}$, growth function N and control function N^2 specified at the beginning of section 3.2, where the nodes at level $i \geqslant 0$ have the form

$$x_i = (V_i, t_0, t_1, \ldots, t_i)$$

and $x_i \succ (V_i, t_i, r_i)$ with $r_i = 3^{-i} r_0$. The construction of the successive regions V_i associated to the nodes could be seen as the introduction of a sequence of increasing finite fields $(\mathcal{F}_i)_{i \geqslant 0}$ of subsets of T of which the V_i are the atoms, and for justifying this we better make the sets V_i *disjoint* by a slight modification of the procedure. Let us come back to the construction step of the children of a given node $x_* \succ (V_*, t_*, r_*)$: first, we have set $r = r_*/3$ and introduced a r-p-net S for V_*; then, the sets V for the children $x \succ (V, t, r)$ of x_* were constructed by breaking into *homogeneous pieces* V each one of the sets $V_* \cap B(s, r)$, for s varying in S. The sets V obtained in this way for a fixed s *cover* $V_* \cap B(s, r)$, but we did not use that fact: we only used that V is *contained* in $B(s, r)$ when $x \succ (V, s, r)$, and that the various sets V corresponding to *all* the children of x_* cover V_*. Thus, instead of defining the regions V starting from the covering of V_* by the balls $B(s, r)$, that overlap in general and may thus produce overlapping regions V when breaking the sets $V_* \cap B(s, r)$ into pieces, we could for example consider the covering of V_* by the Voronoi cells that are associated to the r-p-net $S = \{s_1, s_2, \ldots, s_p\}$ for V_*, namely, the covering of V_* by the sets

$$W_k = \{t \in V_* : \operatorname{dist}(t, S) = d(t, s_k)\}, \quad k = 1, 2, \ldots, p,$$

and in the usual way, make out of those W_k a covering of V_* by the *disjoint sets*

$$\beta_k = W_k \setminus \bigcup_{j<k} W_j, \quad k = 1, 2, \ldots, p.$$

The set β_k is not empty: since S is a r-net, we have $\operatorname{dist}(t, S) < r$ for every $t \in T$, hence $W_j \subset B(s_j, r)$ for each j, thus $s_k \notin W_j$ when $j \neq k$ because $d(s_k, s_j) \geqslant r$, and it follows that $s_k \in \beta_k$. We complete the "disjointification procedure" by an inconsequential modification of the sets $Z(\alpha, n)$, that we define now for every $n \geqslant 1$ by

$$Z(0, n) = [0, E^*/n], \quad \text{and} \quad Z(\alpha, n) = (\alpha E^*/n, (\alpha+1) E^*/n] \ \text{ when } \ 1 \leqslant \alpha < n,$$

a collection of $(n-1)$ half-open intervals; then, for each fixed $n \geqslant 1$, the new sets $Z(\alpha, n)$ are disjoint. Letting $N_* = N(x_*)$ be the growth value for the parent node x_*, we introduce the integer $\widetilde{N} = \max(N_*^3, |S|)$ for the growth value at each child x of x_*; finally, the children $x \succ (V, t, r)$ of x_* are constructed from the elements

$$V = V_* \cap \beta_k \cap Z(\alpha, \widetilde{N}) \neq \emptyset, \quad 1 \leqslant k \leqslant p = |S|, \ \ 0 \leqslant \alpha < \widetilde{N}, \ \ t = s_k \in S, \ \ r = r_*/3,$$

and the different sets V, that clearly cover V_*, are therefore pairwise disjoint. We still have that $V \subset B(t, r)$, knowing that $\beta_k \subset W_k \subset B(s_k, r) = B(t, r)$. As before, we only kept the *non-empty* sets V of the form above. ([23])

We will use auto-indexing, and assume here that $T \subset L^2(\Omega, \mathrm{P})$ is *closed* in L^2. Then T is a compact subset of L^2 (by the Sudakov bound, because $E^* < \infty$), and for every branch $\mathbf{x}$ of $\mathcal{X}$, where we have $x_i(\mathbf{x}) \succ (V_i(\mathbf{x}), t_i(\mathbf{x}), r_i)$ for $i \geqslant 0$, the (Cauchy) sequence $(t_i(\mathbf{x})) \subset T$ tends to a point in T. The set $\mathbf{X}$ of all branches $\mathbf{x}$ of $\mathcal{X}$ is a compact space for its natural tree-topology ([24]), and it projects on T via the continuous ([24]) mapping π that associates a limit point in T to each branch $\mathbf{x}$,

$$\pi(\mathbf{x}) = \lim_i t_i(\mathbf{x}) \in T.$$

On the other hand, because the sets V_i of a same level i are disjoint, each point τ in T determines a *unique* branch $\mathbf{x} = \mathbf{x}(\tau)$ with the property that $\tau \in V_i(\mathbf{x}) \subset B(t_i(\mathbf{x}), r_i)$ for every $i \geqslant 0$; it follows that $d(\tau, t_i(\mathbf{x})) < r_i$ tends to 0, hence

$$\tau = \pi(\mathbf{x}(\tau)), \quad \text{and} \quad T = \pi(\mathbf{X}). \tag{41}$$

Each set V_i in the ith generation is an atom of the finite field $\mathcal{F}_i$ of subsets of T. The field $\mathcal{F}_0$ consists of the sole atom $V_0 = T$. The condition $\mathbf{a}_1^*$ implies that each atom V_i

of $\mathcal{F}_i$ is split into atoms V_{i+1} in the next field $\mathcal{F}_{i+1}$ —here, we might make the side remark that the growth function N_{i+1} is *previsible*: its value depends only upon the preceding field $\mathcal{F}_i$—.

This construction of fields on T could go along with defining a probability measure on the space $(T, \mathcal{F})$: this is what Fernique was looking for, a *mesure majorante* for every centered Gaussian process with a finite expectation for its supremum (see Chap. 6 in [Fer3]). The results in the next Section 3.4 imply that this measure exists ([3]), however, the form given there in Equation (44) proved more manageable and useful than the existence of a majorizing measure. We say with Fernique that a *majorizing measure* for the centered Gaussian process $(X_t)_{t \in T}$ is a probability measure μ on T such that there is a constant K for which

$$J(\tau) := \int_0^\Delta \sqrt{\ln \frac{1}{\mu(B(\tau, r))}} \, \mathrm{d}r \leqslant K$$

for every $\tau \in T$. Fernique showed that the existence of a majorizing measure implies that the expectation of the supremum of the associated process is finite. It should not be a big surprise to learn that the invariant probability measure m on the torus $\mathbb{T}$ is a majorizing measure for any invariant centered Gaussian process on $\mathbb{T}$ whose supremum has a finite expectation: one could say that it is just another form of the Fernique theorem for the torus —note that here the metric on $\mathbb{T}$ is the metric of the process, not the usual one—. As it is the case for Dudley's integral, the integral $J(\tau)$ above can be discretized in the form of an "equivalent" series

$$S(\tau) := \sum_{i=0}^{\infty} r_i \sqrt{\ln \frac{1}{\mu(B(\tau, r_{i+1}))}},$$

that satisfies $(2/9)\, S(\tau) \leqslant J(\tau) \leqslant S(\tau)$ for every $\tau \in T$.

Knowing what we know now, the most sensible thing to do seems to consider the "natural" probability measure ν on $\mathbf{X}$ associated to the tree $\mathcal{X}$ —or to the filtration—. Let $i \geqslant 0$ and let V_i be an atom of the field $\mathcal{F}_i$; there is a unique node $\overline{x}_i$ at level i such that $\overline{x}_i \succ (V_i, t_i, r_i)$ for some $t_i \in T$; we associate to V_i the closed-open subset of $\mathbf{X}$ defined by

$$V_i^* = \{\mathbf{x} \in \mathbf{X} : x_i(\mathbf{x}) = \overline{x}_i \succ (V_i, t_i, r_i)\}.$$

We set

$$\nu(V_0^*) = \nu(\mathbf{X}) = 1, \quad \text{and} \quad \nu(V_{i+1}^*) = \frac{\nu(V_i^*)}{s_{i+1}}$$

for every atom V_{i+1} of $\mathcal{F}_{i+1}$, where V_i is the atom in $\mathcal{F}_i$ containing the atom V_{i+1} and where $s_{i+1} = s(x_{i+1}) \in \{1, \ldots, N_{i+1}^2\}$ is the number of siblings of the node x_{i+1} that corresponds to the atom V_{i+1}, that is to say, the unique node x_{i+1} at level $(i+1)$ such that $x_{i+1} \succ (V_{i+1}, t_{i+1}, r_{i+1})$. It follows that $\nu(V_i^*)$ is equal to the sum of all the values $\nu(V_{i+1}^*)$ corresponding to the siblings of x_{i+1}, so that these coherent values do define a probability measure ν on $\mathbf{X}$. For each node $x_i \succ (V_i, t_i, r_i)$ in $\mathcal{X}$, the measure of V_i^* is equal to

$$\nu(V_i^*) = \prod_{j=1}^{i} \frac{1}{s(x_j)} \geqslant \prod_{j=1}^{i} \frac{1}{N^2(x_j)} =: \frac{1}{M(x_i)},$$

where $x_{i-1}, \ldots, x_j, \ldots, x_1$ are the ancesters of x_i. We see as before in Lemma 5 that

$$M(x_i) = N^2(x_i)\, N^2(x_{i-1}) \ldots N^2(x_1)\, N_0^2 \leqslant N(x_i)^2\, N(x_i)^{2/3} N(x_i)^{2/9} \ldots < N(x_i)^3.$$

Let μ be the image measure of ν by the projection π from $\mathbf{X}$ onto T. If V is an atom, the image by π of $V^* \subset \mathbf{X}$ is a compact subset of T that contains V (see (41)) and is contained in the closure of V in T: indeed, if $\mathbf{x} \in V_i^*$, we have $t_{j+1}(\mathbf{x}) \in V_j(\mathbf{x}) \subset V_i(\mathbf{x}) = V_i$ for every integer $j \geqslant i$, and $t_{j+1}(\mathbf{x})$ tends to $\pi(\mathbf{x})$. If $V \subset B(t, r_1)$ and $r_1 < r_2$, we get that

$$\pi(V^*) \subset \{s \in T : d(t, s) \leqslant r_1\} \subset B(t, r_2)$$

hence

$$\nu(V^*) \leqslant \mu(\pi(V^*)) \leqslant \mu(B(t, r_2)).$$

Let $\tau \in T$ and let $\mathbf{x} = \mathbf{x}(\tau)$ be the branch such that $\tau \in V_i(\mathbf{x})$ for each $i \geqslant 0$. Let us write $x_i(\mathbf{x}) = x_i \succ (V_i, t_i, r_i)$. We know that $\tau \in V_i \subset B(t_i, r_i)$ hence $V_i \subset B(\tau, 2r_i)$, and because $2r_i = 2r_{i-1}/3 < r_{i-1}$ we have

$$\mu(B(\tau, r_{i-1})) \geqslant \nu(V_i^*) \geqslant \frac{1}{M(x_i)} \geqslant \frac{1}{N(x_i)^3}.$$

Observing that $\mu(B(\tau, r_0)) = \mu(T) = 1$, then writing $N(x_i) = N_i(\mathbf{x})$, we obtain

$$\sum_{i=0}^{\infty} r_i \sqrt{\ln \frac{1}{\mu(B(\tau, r_{i+1}))}} = \sum_{i=1}^{\infty} r_{i-2} \sqrt{\ln \frac{1}{\mu(B(\tau, r_{i-1}))}}$$
$$\leqslant \sum_{i=1}^{\infty} r_{i-2} \sqrt{\ln N_i(\mathbf{x})^3} < 6 \sum_{i=1}^{\infty} r_{i-1} \sqrt{\ln N_i(\mathbf{x})} = 6\,\sigma_1(\mathbf{x}).$$

Hence, the probability measure μ on T is a majorizing measure for the process. $\square$

3.4. *Changing the variable*

We continue with the same tree $\mathcal{X}$, growth function N and control function N^2, as in section 3.2. We shall perform a "change of variable" identical to the one that has led us from the first *Dudley series* $\Sigma_1(T)$ in (1) to the series $\Sigma_2(T)$ in (2). We introduced a number b such that $b > 1/(\ln 2) > 1$. Let us consider a branch $\mathbf{x}$ in the tree $\mathcal{X}$; for every integer $k \geqslant 0$, let $i_k(\mathbf{x})$ be the smallest integer $i \geqslant 0$ for which([25]) we have that $b^{k-1} < \ln N_{i+1}(\mathbf{x})$; when $k = 0$ we obtain that $i_0(\mathbf{x}) = 0$, because $\ln N_0 = 0 < b^{-1}$ and $\ln N_1 \geqslant \ln 2 > b^{-1}$. Let us define for every branch $\mathbf{x}$ a "variable analog" $\sigma_2(\mathbf{x})$ of $\Sigma_2(T)$ by setting

$$\sigma_2(\mathbf{x}) = \sum_{k=0}^{\infty} r_{i_k(\mathbf{x})}\, b^{k/2}. \tag{42}$$

We repeat exactly the computations that we have done for $\Sigma_2(T)$: let $I(\mathbf{x}) \subset \mathbb{N}$ be the set of values $i_k(\mathbf{x})$, $k \geqslant 0$, and for every $i \in I(\mathbf{x})$, let $k(i) = k(i, \mathbf{x})$ be the largest k such that $i_k(\mathbf{x}) = i$. First, adding up geometric progressions, then observing when $i_k(\mathbf{x}) = i$ that $b^{k-1} < \ln N_{i+1}(\mathbf{x})$ and thus $b^{k(i)} < b \ln N_{i+1}(\mathbf{x})$, we get

$$\sigma_2(\mathbf{x}) = \sum_{i \in I(\mathbf{x})} \Bigl(\sum_{i_k(\mathbf{x}) = i} r_i\, b^{k/2} \Bigr) \leqslant \sum_{i \in I(\mathbf{x})} r_i \frac{\bigl(\sqrt{b}\bigr)^{k(i)+1} - 1}{\sqrt{b} - 1} < \frac{b^{1/2}}{b^{1/2} - 1} \sum_{i \in I(\mathbf{x})} r_i\, b^{k(i)/2}$$
$$< \frac{b}{b^{1/2}-1} \sum_{i \in I(\mathbf{x})} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{b}{b^{1/2}-1} \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})},$$

therefore

$$\sigma_2(\mathbf{x}) \leqslant \frac{b}{b^{1/2}-1}\, \sigma_1(\mathbf{x}). \tag{43}$$

Let us fix $k \geqslant 0$. By the definition of $i_k(\mathbf{x})$, we know that

$$\ln N_{i_k(\mathbf{x})}(\mathbf{x}) \leqslant b^{k-1} < \ln N_{i_k(\mathbf{x})+1}(\mathbf{x}).$$

Consider the family X_k of nodes $x_{i_k(\mathbf{x})}(\mathbf{x})$, where $\mathbf{x}$ varies in the set of branches of $\mathcal{X}$. The nodes $x \in X_k$ are characterized by

$$N(x) \leqslant \exp(b^{k-1}) \quad \text{and} \quad y \in C(x) \;\Rightarrow\; N(y) > \exp(b^{k-1}).$$

Thus, the nodes in X_k are the leaves of the subtree $\mathcal{X}_M$ from Equation (19), where we should let $M = M_k = \exp(b^{k-1})^2$ —remember that the 3-control function here is given by $x \in \mathcal{X} \mapsto N(x)^2$—. By Lemma 5, we obtain that

$$|X_k| \leqslant M_k^{3/2} = \exp(3\,b^{k-1}).$$

When $k = 0$, we have seen that $i_0(\mathbf{x}) = 0$ for every branch $\mathbf{x}$, hence $x_{i_0(\mathbf{x})}(\mathbf{x}) = x_0$: the subtree $\mathcal{X}_{M_0}$ is reduced to the root $x_0 \in \mathcal{X}$, and therefore $X_0 = \{x_0\}$.

For every index $k \geqslant 0$, let T_k denote the set of *latest positions* $\theta(x) \in T$ of all the nodes $x \in X_k$, namely

$$T_k = \{\theta(x) : x \in X_k\}; \quad \text{we have that} \;\; \ln|T_k| \leqslant \ln|X_k| < 3\,b^k.$$

We see that $T_0 = \{t_0\}$ since $X_0 = \{x_0\} = \{(V_0, t_0)\}$. If the expectation E^* of the supremum of the process $(X_t)_{t\in T}$ is finite, we know by (38) that the function $\mathbf{x} \mapsto \sigma_1(\mathbf{x})$ is bounded by some value K_1 that is a universal multiple of E^*; it follows by Equation (43) that $\mathbf{x} \mapsto \sigma_2(\mathbf{x})$ is bounded by a multiple $K_2 = c(b)\,K_1$ of K_1, with a factor $c(b)$ that only depends upon the choice of b.

Let τ be an arbitrary point in T; there is a branch $\mathbf{x} = \mathbf{x}(\tau)$ such that τ belongs to the region $R(x)$ for each node x in $\mathbf{x}$; if $k \geqslant 0$ is given, consider the index $i = i_k(\mathbf{x})$ associated to that branch. Then $x_i(\mathbf{x}) = x_{i_k(\mathbf{x})}(\mathbf{x})$ is an element of X_k. This means that if we write $x_i(\mathbf{x}) \succ (V_i, t_i, r_i)$, we have $t_i \in T_k$. As always we know that

$$V_i \subset B(t_i, r_i), \quad \text{and} \quad \tau \in V_i \subset B(t_i, r_i).$$

It yields that $\mathrm{dist}(\tau, T_k) \leqslant \mathrm{dist}(\tau, t_i) < r_i = r_{i_k(\mathbf{x})}$, we thus conclude that

$$\sum_{k=0}^{\infty} \mathrm{dist}(\tau, T_k)\, b^{k/2} < \sum_{k=0}^{\infty} r_{i_k(\mathbf{x})}\, b^{k/2} = \sigma_2(\mathbf{x}) \leqslant K_2.$$

After such a long time spent together, we finally abandon the tree $\mathcal{X}$. Here is the final word: the sets (T_k) are finite subsets of T such that

$$\left\{\begin{array}{l} |T_0| = 1; \qquad \ln|T_k| < 3\,b^k \;\; \text{and} \\[2mm] \text{for every } \tau \in T, \quad \displaystyle\sum_{k=0}^{\infty} \mathrm{dist}(\tau, T_k)\, b^{k/2} \leqslant K_2. \end{array}\right. \tag{44}$$

That this implies a bound on the expectation of the supremum is slightly easier to check than before: to each $\tau \in T$ and $k \geqslant 0$ we can associate a point $t_k(\tau) \in T_k$ such that $d(\tau, t_k(\tau)) = \mathrm{dist}(\tau, T_k)$; the number of couples $(t_k(\tau), t_{k+1}(\tau))$ that can appear when τ varies in T is less that $q_k := |T_k|\,|T_{k+1}|$, so that $\ln q_k$ is less than $6\,b^{k+1}$. When going from $t_k(\tau)$ to $t_{k+1}(\tau)$ we shall look for jumps of order $\mathrm{dist}(\tau, T_k)\, b^{k/2}$ for the process. We can then apply the Gaussian bounds of Lemma 1 as we did before, in section 2.2 or in section 3.2, and obtain

$$\mathrm{P}\Bigl(\frac{|X_{t_{k+1}(\tau)} - X_{t_k(\tau)}|}{\mathrm{dist}(\tau, T_k) + \mathrm{dist}(\tau, T_{k+1})} > c\,b^{k/2}\Bigr) \leqslant \exp(-c^2 b^k/2).$$

We just have to choose $c > \sqrt{12\,b}$ in order to compensate the size of $q_k \leqslant \exp(6\,b.b^k)$ when applying the union bound inequality. We may observe that in doing so, we shall only use the sub-gaussian character: we could as well deal here with any centered process $(X_t)_{t \in T}$ satisfying (44) and such that

$$\mathrm{P}\big(|X_t - X_s|/d(t,s) > u\big) \leqslant 2\,\mathrm{e}^{-u^2/2}, \quad s,t \in T,\ d(t,s) > 0,\ u > 0.$$

We can at last state Talagrand's theorem ([26]).

Theorem 2. *The expectation of the supremum of a centered Gaussian process $(X_t)_{t \in T}$ is finite if and only if there exists a family $(T_k)_{k \geqslant 0}$ of subsets of T that satisfies (44).*

4. Back to norms

We will finally come back to some Functional Analysis, with normed spaces and bounded linear maps, we also say *operators*. We start easily, with $\mathbb{R}^n$ equipped with the usual Euclidean norm

$$\|x\|_2 = \Big(\sum_{i=1}^n x_i^2 \Big)^{1/2}, \quad x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n.$$

Now and later below we shall deal with transforming, via a bounded linear map a, a *uniform scalar estimate* obtained in a normed linear domain space E, into a *norm estimate* in the normed range space F. Let for example $a : E = \mathbb{R}^n \to F$ be a linear map from our normed space $\mathbb{R}^n$ to a normed linear space F. Consider the standard Gaussian probability measure γ_n on $\mathbb{R}^n$,

$$\mathrm{d}\gamma_n(x) = \frac{1}{(2\pi)^{n/2}}\,\mathrm{e}^{-\|x\|_2^2/2}\,\mathrm{d}x.$$

The uniform scalar estimate we have in mind is this: for every linear functional ξ on $\mathbb{R}^n$, the image measure $\gamma_\xi := \xi(\gamma_n)$ on the line is a centered Gaussian probability measure such that

$$\int_{\mathbb{R}} u^2\,\mathrm{d}\gamma_\xi(u) = \|\xi\|^2,$$

hence

$$\sup_{\|\xi\| \leqslant 1} \int_{\mathbb{R}^n} |\langle \xi, x \rangle|^2\,\mathrm{d}\gamma_n(x) = \sup_{\|\xi\| \leqslant 1} \int_{\mathbb{R}} u^2\,\mathrm{d}\gamma_\xi(u) \leqslant 1.$$

This is what we mean by a *uniform scalar estimate* for a measure on the domain space, here γ_n on $\mathbb{R}^n$. This is just one example, implicitly involving the L^2-norm, but we could also use the other L^p-norms and even beyond, as we shall see later.

We can view the normed space $\mathbb{R}^n$ with the probability measure γ_n as a probability space $(\Omega, \mathrm{P}) = (\mathbb{R}^n, \gamma_n)$. Then each linear functional ξ on $\mathbb{R}^n$ can be viewed as a centered Gaussian random variable defined on Ω, and the preceding quantities can be seen as *variances*,

$$\mathrm{Var}_{\gamma_n}(\xi) = \mathrm{E}_{\gamma_n}(\xi - \mathrm{E}_{\gamma_n}\xi)^2 = \mathrm{E}_{\gamma_n}\xi^2 = \int_{\mathbb{R}} u^2\,\mathrm{d}\gamma_\xi(u).$$

Now, we may be interested in the image probability measure $\nu := a(\gamma_n)$ of γ_n on F —in other words, the *pushforward* $\nu = a_\#\gamma_n$ of γ_n—, and look for a *norm estimate* on ν, rather than a mere scalar one, namely an estimate about

$$\int_F \|y\|_F\,\mathrm{d}\nu(y) = \int_{\mathbb{R}^n} \|a\,x\|_F\,\mathrm{d}\gamma_n(x).$$

Here, we chose an L^1-norm that will better fit our immediate purpose. In the Gaussian case, this choice is not crucial, as it is known that all moments of Gaussian measures on normed spaces are equivalent (Shepp–Landau–Fernique, see the simpler proof in [Fer_1]), but it will be different soon.

The norm of a vector $y \in F$ is the supremum of the values $\langle y^*, y\rangle$ at y of the linear functionals y^* in the unit ball B_{F^*} of the dual F^* of F. Let us introduce the adjoint map a^* of a, that maps from F^* to $(\mathbb{R}^n)^* \simeq \mathbb{R}^n$. It is defined by the equation

$$\langle a^* y^*, x\rangle = \langle y^*, a\,x\rangle, \quad x \in \mathbb{R}^n, \;\; y^* \in F^*.$$

We have therefore

$$\|a\,x\|_F = \sup_{y^* \in B_{F^*}} \langle y^*, a\,x\rangle = \sup_{y^* \in B_{F^*}} \langle a^* y^*, x\rangle.$$

Letting $T = a^*(B_{F^*})$ be the image of the unit ball B_{F^*} of F^* under a^*, we can rewrite

$$\|a\,x\|_F = \sup_{\xi \in T} \langle \xi, x\rangle.$$

We said that each linear functional ξ can be viewed as a (centered) Gaussian random variable defined on $(\mathbb{R}^n, \gamma_n)$,

$$\xi(\omega) = \langle \xi, \omega\rangle, \quad \text{with} \quad \omega = x \in \mathbb{R}^n.$$

Then the family of $\xi = t \in T$ is a centered Gaussian process $(X_t)_{t\in T}$ with $X_t = t = \xi$, and

$$\sup_{t\in T} X_t(\omega) = \|a\,\omega\|_F.$$

Finally,

$$\int_F \|y\|_F \,\mathrm{d}\nu(y) = \int_{\mathbb{R}^n} \|a\,x\|_F \,\mathrm{d}\gamma_n(x) = \mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr).$$

This is the connection to what we have seen in the preceding sections.

4.1. *Radonifying*

The words “application radonifiante” were used by Laurent Schwartz, I couldn’t say whether he invented this *radonifiant* or borrowed it somewhere. It is by attending the 1969 seminar [Sch] on the subject that I had my first experiences in “live” mathematics.

Let us start by giving an idea of the so-called *Gaussian cylindrical measure* γ_H on a Hilbert space H; our Hilbert space will be $\ell^2(\mathbb{N})$. The standard way of building a model for an infinite sequence $X_0, X_1, \ldots, X_n, \ldots$ of independent $\mathrm{N}(0,1)$ Gaussian random variables is to use an infinite product of copies of the probability space $(\mathbb{R}, \gamma_1)$, where γ_1 is the distribution of the $\mathrm{N}(0,1)$ Gaussian random variables. We let

$$\Omega = \mathbb{R}^{\mathbb{N}}, \quad \Gamma = \bigotimes_{i=0}^{\infty} \gamma^{(i)}, \quad \gamma^{(i)} = \gamma_1.$$

Then, the formulas

$$X_i(\omega) = \omega_i \in \mathbb{R}, \quad i \geqslant 0,$$

where $\omega = (\omega_i)_{i\geqslant 0} \in \Omega = \mathbb{R}^{\mathbb{N}}$, define a sequence of independent $\mathrm{N}(0,1)$ variables $(X_i)_{i\geqslant 0}$ on the probability space (Ω, Γ). Consider now $H = \ell^2(\mathbb{N})$ as a subset of $\mathbb{R}^{\mathbb{N}} = \Omega$ via the formal identical injection $(x_i)_{i\in\mathbb{N}} \in \ell^2(\mathbb{N}) \mapsto (x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$. If we set on Ω the product topology, we can easily check that the closed unit ball B_H of H, namely

$$B_H = \bigl\{x = (x_i)_{i\in\mathbb{N}} : \sum_{i=0}^{\infty} x_i^2 \leqslant 1\bigr\},$$

is a compact subset of Ω, as well as its multiple rB_H of an arbitrary radius $r > 0$. Hence, the subset H of Ω is a K_σ-set, thus a Borel subset of Ω. The ball rB_H is contained in the hypercube

$$C_r = [-r, r]^{\mathbb{N}}, \quad \text{and} \quad \Gamma(C_r) = \prod_{i=0}^{\infty} \gamma^{(i)}([-r, r]) = 0,$$

as $\gamma^{(i)}([-r, r]) = \gamma_1([-r, r]) < 1$. It follows that $\Gamma(rB_H) = 0$, and H is a Γ-null set in Ω: the probability measure Γ does not induce a meaningful measure on H. However, if ξ is a bounded linear functional on H, we can make sense of ξ as a random variable on (Ω, Γ); indeed, we can see the action of $\xi = (\xi_i)_{i \geqslant 0} \in H^* \simeq \ell^2(\mathbb{N})$ as a series of multiples of the coordinate functions on Ω. For every integer $n \geqslant 0$, the function

$$\omega \mapsto \sum_{i=0}^{n} \xi_i X_i(\omega)$$

is a centered Gaussian variable of variance $\sum_{i=0}^{n} \xi_i^2$ on (Ω, P); if $\xi \in H^*$, the convergent series

$$\xi(\omega) := \sum_{i=0}^{\infty} \xi_i X_i(\omega)$$

defines a Gaussian variable of variance $\|\xi\|_{H^*}^2$, and we can introduce its distribution Γ_ξ on the line. From this we can also consider, for every finite dimensional quotient H_0 of H, a probability measure Γ_{H_0} on H_0 that is the distribution of the vector valued random variable P_{H_0}, the quotient map from H onto H_0; this distribution is actually the $\mathrm{N}(0, \mathrm{Id})$ Gaussian measure of that Euclidean space H_0. In addition, if H_1 is a further quotient of H_0, then Γ_{H_1} is the image measure of Γ_{H_0} by the quotient map from H_0 onto H_1. Thus, the Gaussian cylindrical measure γ_H on H is not a measure on H, but a *projective system* of measures on the system of finite dimensional quotients of H. This "canonical" object γ_H is also called *white noise*.

This notion of projective system of probability measures applies as well to any Banach space, but for the Hilbert space it is not easy to distinguish between a finite dimensional quotient H_0 and the finite dimensional subspace H_0^f orthogonal to the kernel of the quotient map P_{H_0}: one can thus also think that the cylindrical measure γ_H is in some sense the limit of the *inductive* system of $\mathrm{N}(0, \mathrm{Id}_{H^f})$ Gaussian probability measures on the finite dimensional subspaces H^f of H. In the case of $H = \ell^2(\mathbb{N})$, we can simplify and consider the sequence $(\gamma_n)_{n \geqslant 1}$ of measures on the specific n-dimensional subspaces H_n of H defined by

$$H_n = \{(x_i)_{i \geqslant 0} : x_j = 0,\ j \geqslant n\}$$

and see the cylindrical measure $\gamma_{\ell^2(\mathbb{N})}$ as a sort of limit of the sequence (γ_n).

Consider now for every $u > 0$ the product $K_u \subset \mathbb{R}^{\mathbb{N}}$ defined by

$$K_u = \prod_{i=0}^{\infty} \bigl[-\bigl(u + \sqrt{4 \ln(i+3)}\bigr), u + \sqrt{4 \ln(i+3)}\,\bigr].$$

It follows easily from (7) that

$$\Gamma(\Omega \setminus K_u) \leqslant \Bigl(\sum_{i=0}^{\infty} \frac{1}{(i+3)^2}\Bigr) \mathrm{e}^{-u^2/2}, \quad \text{hence} \quad \lim_{u \to \infty} \Gamma(K_u) = 1.$$

Let us introduce the diagonal map

$$\beta : \mathbb{R}^{\mathbb{N}} \longrightarrow \mathbb{R}^{\mathbb{N}}, \qquad \beta\big((x_i)_{i\in\mathbb{N}}\big) = (\beta_i\, x_i)_{i\in\mathbb{N}}$$

where the diagonal coefficients $(\beta_i)_{i\in\mathbb{N}}$ decrease as $1/\sqrt{\ln i}$, say

$$\beta_i = \frac{1}{\sqrt{\ln(i+3)}} < 1, \qquad i \geqslant 0.$$

The images $\beta(K_u)$ of the different products K_u are contained in hypercubes $[-c_u, c_u]^{\mathbb{N}}$, because

$$\frac{u + \sqrt{4\,\ln(i+3)}}{\sqrt{\ln(i+3)}} < u + 2 =: c_u.$$

The image measure $\beta(\Gamma)$ is therefore supported on a family of hypercubes, that can be seen as bounded subsets in the Banach space $\ell^\infty(\mathbb{N})$. We can consider $\beta(\Gamma)$ as a probability measure on $\ell^\infty(\mathbb{N})$ —we should add: with respect to the weak*-Borel sets, or in other words, the Borel sets induced on $\ell^\infty(\mathbb{N})$ by those of $\mathbb{R}^{\mathbb{N}}$—.

We can also view β as a linear map from $\ell^2(\mathbb{N})$ to $\ell^\infty(\mathbb{N})$, and we then have an example of a *radonifying fact*: the somewhat abstract Gaussian cylindrical measure γ_H defined "on" the linear space $H = \ell^2(\mathbb{N})$ is transformed by the map β into a "true" measure $\beta(\gamma_H) = \beta(\Gamma)$ on the range space $\ell^\infty(\mathbb{N})$.

4.2. *p-summing maps*

Given $p \in (0, \infty)$, a bounded linear map a from a normed space E to another normed space F is *p-summing* when there is a constant C such that for every finite nonnegative measure μ on E, one has

$$\int_F \|y\|_F^p \,\mathrm{d}a(\mu)(y) = \int_E \|a\,x\|_F^p \,\mathrm{d}\mu(x) \leqslant C^p \sup_{x^*\in B(E^*)} \int_E \big|\langle x^*, x\rangle\big|^p \,\mathrm{d}\mu(x). \tag{45}$$

The p-summing norm (27) $\pi_p(a)$ of a is defined to be the smallest possible C in the above inequality. The notion of factorization of linear operators is important is this context: it is easy to see that given $a_0 : E_0 \to E$ and $a_1 : F \to F_1$, we have

$$\pi_p(a_1 \circ a \circ a_0) \leqslant \|a_0\|\, \pi_p(a)\, \|a_1\|,$$

so that the p-summing maps form an *operator ideal*. This property applies in particular when E_0 is a subspace of E, with the induced norm, and when a_0 is the isometric injection from E_0 into E; then $a \circ a_0$ is the restriction of a to E_0: the p-summing maps are stable by restriction —a fact that is is clear directly from the definition and the Hahn–Banach extension theorem for functionals—. Also, we can prove that a given map is p-summing if we manage to factor it through another map that is known to be p-summing.

These concepts were thoroughly studied by Pietsch [Pie$_2$]. In [Pie$_1$], he introduced a notion that became known as the *Pietsch measure* for the p-summing map a: the unit ball B_{E^*} of the dual E^* equipped with the weak* topology is a compact space; the bounded linear map a from E to F is p-summing with $\pi_p(a) \leqslant C$ if and only if there exists a probability measure P_a on that compact B_{E^*} such that

$$\|a\,x\|_F^p \leqslant C^p \int_{B_{E^*}} \big|\langle x^*, x\rangle\big|^p \,\mathrm{d}\mathrm{P}_a(x^*), \quad x \in E. \tag{46}$$

Deducing Inequality (45) from this is a simple matter of appying the Fubini theorem, followed by an easy $L^1(\mathrm{P}_a)$–$L^\infty(\mathrm{P}_a)$ bound,

$$\begin{aligned}
\int_F \|y\|_F^p \,\mathrm{d}a(\mu)(y) &= \int_E \|a\,x\|_F^p \,\mathrm{d}\mu(x) \\
&\leqslant C^p \int_{B_{E^*}} \Bigl(\int_E \bigl|\langle x^*, x\rangle\bigr|^p \,\mathrm{d}\mu(x)\Bigr) \,\mathrm{d}\mathrm{P}_a(x^*) \\
&\leqslant C^p \sup_{x^* \in B_{E^*}} \Bigl(\int_E \bigl|\langle x^*, x\rangle\bigr|^p \,\mathrm{d}\mu(x)\Bigr).
\end{aligned} \tag{47}$$

Proving the existence of the Pietsch measure requires a clever application of the Hahn–Banach separation theorem —and of the fact that the space of measures on a compact space is the dual of the space of continuous functions on that compact—. A sketch of proof is given below in a more general setting.

The existence of the Pietsch measure leads to a factorization. Let K_E denote the compact space B_{E^*} equipped with the weak* topology. Classically, one thinks of E as the subspace of $C(K_E)$ (the space of continuous functions on K_E) consisting of the functions

$$x^* \in K_E \;\mapsto\; \langle x^*, x\rangle, \qquad x \in E,$$

with the Hahn–Banach theorem implying that E is *isometrically* embedded in $C(K_E)$ in that way. The preceding chain of inequalities (47) can easily be modified and used to show that the formal "injection"

$$i_E : C(K_E) \;\longrightarrow\; L^p(K_E, \mathrm{P}_a)$$

is p-summing and $\pi_p(i_E) = 1$. If we consider E as a subspace of $C(K_E)$, we may look at its image $i_E(E)$ in $L^p(K_E, \mathrm{P}_a)$ and call E_p the closure of $i_E(E)$ in $L^p(\mathrm{P}_a)$. Then the Pietsch measure inequality says that a can be factored as

$$a : E \;\underset{j_E}{\longrightarrow}\; E_p \;\underset{a_1}{\longrightarrow}\; F$$

where j_E is the restriction of i_E to E. Indeed, Inequality (46) means that

$$\|a\,x\|_F \leqslant C\,\|i_E(x)\|_{L^p(\mathrm{P}_a)}, \qquad x \in E,$$

so that we can define a bounded operator a_1 from the image $j_E(E)$ to the space F by letting $a_1(j_E(x)) = a(x)$ for $x \in E$, then extend a_1 to the closure E_p of $j_E(E)$ in $L^p(\mathrm{P}_a)$. In this factorization, the first map j_E is p-summing as restriction of i_E and

$$\pi_p(j_E) \leqslant 1, \qquad \|a_1\| \leqslant C.$$

Note that we may find P_a for which $C = \pi_p(a)$. We can sum up everything by recalling that $a = a_1 \circ j_E$ and drawing the following diagram:

$$\begin{array}{ccccc}
C(K_E) & \underset{i_E}{\longrightarrow} & L^p(K_E, \mathrm{P}_a) & & \\
\bigcup & & \bigcup & & \\
E & \underset{j_E}{\longrightarrow} & E_p & \underset{a_1}{\longrightarrow} & F
\end{array}$$

The early works of Alexandre Grothendieck [Gro$_1$], [Gro$_2$] in Functional Analysis contain factorizations and measures similar to the Pietsch measure and factorization, in particular in the case where $p = 2$. Lindenstrauss and Pełczyński [LiPe] have translated the now famous *Grothendieck inequality* from [Gro$_2$] in the language of p-summing operators: *every linear operator from an L^1 space to a Hilbert space H is 1-summing.*

The usual definition [Pie$_1$] of a p-summing map from E to F uses finite sequences of vectors in E rather than measures, asking that for some C and every $n \geqslant 1$ we have

$$\sum_{j=1}^{n} \|a\,x_j\|_F^p \leqslant C^p \sup_{x^* \in B_{E^*}} \sum_{j=1}^{n} \bigl|\langle x^*, x_j\rangle\bigr|^p, \qquad (x_j)_{j=1}^n \subset E. \tag{48}$$

Replacing each vector x_j with $\lambda_j^{1/p} x_j$, we see that (48) is directly equivalent to restricting our definition (45) to finitely supported nonnegative measures $\mu = \sum_{j=1}^n \lambda_j \, \delta_{x_j}$ on E.

Suppose now that a linear mapping $a : E \to F$ satisfies

$$\int_E \|a\,x\|_F \,\mathrm{d}\mu(x) \leqslant C \sup_{x^* \in B(E^*)} \Bigl(\int_E \bigl|\langle x^*, x\rangle\bigr|^p \,\mathrm{d}\mu(x) \Bigr)^{1/p}, \tag{49}$$

for some C and every *probability* measure μ on E. This is formally weaker that our definition of p-summing maps (the $L^1(\mu)$ norm on the left is smaller), but we will check easily that it is equivalent. Playing with the $L^q(\mu)$ norms on the right-hand side of (49), this implies that *when the mapping a is p-summing, it is also q-summing for all $q \geqslant p$.*

We show now that (49) implies (48). Let $(x_j)_{j=1}^n$ be given and let us try to prove (48). We may discard those vectors x_j with $a\,x_j = 0$. Then, for every index j in the remaining set $R \subset \{1, 2, \ldots, n\}$, we define

$$\lambda_j = \|a\,x_j\|_F^p, \quad x'_j = \frac{1}{\|a\,x_j\|_F} x_j \in E, \quad \lambda'_j = \frac{\lambda_j}{\sum_{i \in R} \lambda_i} \cdot$$

We observe that $\|a\,x'_j\|_F = 1$, and (49) applied to $\mu = \sum_{j \in R} \lambda'_j \, \delta_{x'_j}$ gives the result,

$$\begin{aligned} 1 = \sum_{j \in R} \lambda'_j = \sum_{j \in R} \lambda'_j \, \|a\,x'_j\|_F &= \int_E \|a\,x\|_F \,\mathrm{d}\mu(x) \\ &\leqslant C \sup_{x^* \in B_{E^*}} \Bigl(\sum_{j \in R} \lambda'_j \, \bigl|\langle x^*, x'_j\rangle\bigr|^p \Bigr)^{1/p} \\ &= C \sup_{x^* \in B_{E^*}} \Bigl(\frac{\sum_{j \in R} \bigl|\langle x^*, x_j\rangle\bigr|^p}{\sum_{j \in R} \|a\,x_j\|_F^p} \Bigr)^{1/p}. \end{aligned} \qquad \square$$

4.3. Φ-*summing maps*

One can go beyond the class of L^p-spaces, and introduce *Φ-summing maps* for any *Orlicz function* Φ, especially for functions Φ that grow more rapidly at infinity than any power-type function. An Orlicz function Φ is an increasing convex function on $[0, \infty)$ with $\Phi(0) = 0$. If (S, ν) is a measure space, one says that a function f on S belongs to the space $L^\Phi(\nu)$ when there is $a > 0$ such that the integral of $\Phi(a|f|)$ with respect to ν is finite; hence, Φ and $u \mapsto \Phi(cu)$ define the same space of functions, for every $c > 0$. We extend Φ to $\mathbb{R}$ by letting $\Phi(t) = \Phi(|t|)$ when $t < 0$.

Of special interest are the functions $u \mapsto \mathrm{e}^{c\,u^2} - 1$, that are closely related to the Gaussian distribution. Fernique examined those functions and their *conjugate functions* in Chap. 5 of [Fer$_3$]. We make here a bizarre choice of c and set

$$\Phi_2(u) = \mathrm{e}^{3\,u^2/8} - 1, \quad u \in \mathbb{R}.$$

We say that a random variable X has norm $\leqslant 1$ with respect to Φ_2 when

$$\mathrm{E}\,\Phi_2(X) \leqslant 1, \quad \text{that is, when} \quad \mathrm{E}\exp(3\,X^2/8) \leqslant 2.$$

By Markov's inequality, this implies an estimation on the tail of the distribution of X,

$$\mathrm{P}(|X| > u) \leqslant 2\,\mathrm{e}^{-3u^2/8}, \quad u > 0,$$

and conversely, this bound on the tail yields for example that $\mathrm{E}\,\Phi_2(\frac{1}{2}X)$ is finite. It is easy to see that having $f \in L^{\Phi_2}$ is equivalent to saying that f belongs to all the L^p spaces with $p < \infty$ and that for some C, one has ([28])

$$\|f\|_{L^p} \leqslant C\sqrt{p}, \quad p \geqslant 2.$$

If X is a $\mathrm{N}(0,1)$ Gaussian variable, we get

$$\mathrm{E}\,\Phi_2(X) + 1 = \int_{\mathbb{R}} \mathrm{e}^{3u^2/8}\,\mathrm{e}^{-u^2/2}\,\frac{\mathrm{d}u}{\sqrt{2\pi}} = \int_{\mathbb{R}} \mathrm{e}^{-u^2/8}\,\frac{\mathrm{d}u}{\sqrt{2\pi}} = 2,$$

hence X has norm 1 in L^{Φ_2} —this is the reason for our strange normalization of Φ_2—.

We say that a linear map $a : E \to F$ is Φ-summing if it transforms measures μ on E with a uniform scalar L^Φ-estimate into measures on F with a norm estimate, in the form (49) for example: for some C and for every *probability* measure μ on E, one has

$$\sup_{\xi \in B(E^*)} \big\|x \mapsto \langle \xi, x\rangle\big\|_{L^\Phi(\mu)} \leqslant 1 \quad \Rightarrow \quad \int_E \|a\,x\|_F \,\mathrm{d}\mu(x) \leqslant C. \tag{50}$$

Suppose that $a \neq 0$ satisfies (50) with $C = 1$, and consider the set S of functions f on the unit ball B_{E^*} of the dual E^* of E that have the form

$$f(\xi) = -1 + \sum_{i=1}^n \lambda_i\,\Phi(\langle \xi, y_i\rangle), \quad \xi \in B_{E^*},$$

where $n \geqslant 1$, $\lambda_i \geqslant 0$ and $\sum_j \lambda_j = 1$, and $\|a\,y_i\|_F > 1$ for every $i = 1, 2, \ldots, n$. This set S is a convex set of functions that are continuous on the compact $K_E = B_{E^*}$ equipped with the weak-$*$ topology. One sees from (50) applied to $\mu = \sum_i \lambda_i\,\delta_{y_i}$ that S is disjoint from the open convex cone Ω consisting of functions < 0 on K_E. It follows from the Hahn–Banach separation theorem that there exists a probability measure P_a on K_E such that $\mathrm{P}_a(f) \geqslant 0$ for every $f \in S$, in particular,

$$\int_{K_E} \Phi(\langle \xi, y\rangle)\,\mathrm{d}\mathrm{P}_a(\xi) \geqslant 1$$

whenever $\|a\,y\|_F > 1$, and it yields that

$$\int_{K_E} \Phi\Big(\frac{\langle \xi, x\rangle}{\|a\,x\|_F}\Big)\,\mathrm{d}\mathrm{P}_a(\xi) \geqslant 1 \ \text{ when } \ a\,x \neq 0, \ \text{ or } \ \|a\,x\|_F \leqslant \big\|\xi \mapsto \langle \xi, x\rangle\big\|_{L^\Phi(\mathrm{P}_a)} \tag{51}$$

for every $x \in E$. The probability measure P_a is a Pietsch measure for the Φ-summing character of a, let us say a Φ-*Pietsch measure*.

Going from the inequality (51) with the Pietsch measure, back to the definition (50), is less pleasant than in the p-summing case, due to the lack of homogeneity here. Since Φ is convex and $\Phi(0) = 0$, we know that $\Phi(t)/t$ is non-decreasing for $t > 0$, hence when $\|a\,x\|_F \geqslant 1$ we may write

$$\|a\,x\|_F\,\Phi\Big(\frac{\langle \xi, x\rangle}{\|a\,x\|_F}\Big) = \Phi\Big(\frac{\langle \xi, x\rangle}{\|a\,x\|_F}\Big)\Big/\Big(\frac{1}{\|a\,x\|_F}\Big) \leqslant \Phi\big(\langle \xi, x\rangle\big)$$

and using (51), we get

$$\|a\,x\|_F \leqslant \|a\,x\|_F \int_{K_E} \Phi\Big(\frac{\langle \xi, x\rangle}{\|a\,x\|_F}\Big)\,\mathrm{d}\mathrm{P}_a(\xi) \leqslant \int_{K_E} \Phi\big(\langle \xi, x\rangle\big)\,\mathrm{d}\mathrm{P}_a(\xi).$$

Suppose that μ is a probability measure on E such that

$$\sup_{\xi \in K_E} \int_E \Phi\bigl(\langle \xi, x\rangle\bigr)\, \mathrm{d}\mu(x) \leqslant 1.$$

It follows that

$$\int_E \mathbf{1}_{\{\|a\,x\|_F \geqslant 1\}}\, \|a\,x\|_F \,\mathrm{d}\mu(x) \leqslant \int \Phi\bigl(\langle \xi, x\rangle\bigr)\,\mathrm{d}\mathrm{P}_a(\xi)\,\mathrm{d}\mu(x) \leqslant \int \mathrm{d}\mathrm{P}_a(\xi) = 1,$$

and

$$\int_E \|a\,x\|_F \,\mathrm{d}\mu(x) \leqslant 1 + \int_E \mathbf{1}_{\{\|a\,x\|_F \geqslant 1\}}\, \|a\,x\|_F \,\mathrm{d}\mu(x) \leqslant 2.$$

Starting with a Pietsch measure with constant 1, we got Inequality (50) with $C = 2$, which is somewhat unsatisfactory, but was easily obtained.

If the given Orlicz function Φ is our function Φ_2, we see that for any p finite, the $L^{\Phi_2}(\mu)$-norm is larger than some multiple of the $L^p(\mu)$-norm: it then follows, comparing (49) and (50), that being Φ_2-summing is weaker than having any of the p-summing properties with $p < \infty$. A prototypical example of a Φ_2-summing map is the diagonal map $\beta : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ given by

$$\beta\bigl((x_i)_{i\in\mathbb{N}}\bigr) = (\beta_i\, x_i)_{i\in\mathbb{N}}, \quad \beta_i = \frac{1}{\sqrt{\ln(i+3)}},$$

seen previously in Section 4.1. Indeed, suppose that a diagonal map $\delta : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ is given, with diagonal coefficients $(\delta_i)_{i\in\mathbb{N}}$ that decrease at least as rapidly as the $(\beta_i)_{i\in\mathbb{N}}$ do, say

$$|\delta_i| \leqslant \beta_i < 1, \quad i \geqslant 0.$$

Then δ is a Φ_2-summing mapping: let μ be a probability measure on $\ell^\infty(\mathbb{N})$ satisfying a uniform Φ_2-scalar estimate; let us suppose for instance that for every coordinate function $x \in \ell^\infty \mapsto x_i$ with $i \geqslant 0$, we have

$$\int \Phi_2(x_i)\,\mathrm{d}\mu(x) \leqslant 1.$$

By Markov again, it implies for each coordinate function that

$$\mu(x : |x_i| > u) \leqslant 2\,\mathrm{e}^{-3\,u^2/8}, \quad i \geqslant 0.$$

Then

$$\begin{aligned} \mu(x : \|\delta\,x\|_\infty > 2+u) &= \mu\bigl(x : \sup_{i\geqslant 0} |\delta_i\,x_i| > 2+u\bigr) \\ &\leqslant \sum_{i\geqslant 0} \mu\bigl(x : |x_i| > (2+u)\sqrt{\ln(i+3)}\bigr) \\ &\leqslant \sum_{i\geqslant 0} \mu\bigl(x : |x_i| > \sqrt{4\,\ln(i+3)} + u\bigr) \\ &\leqslant 2\Bigl(\sum_{i\geqslant 0} \frac{1}{(i+3)^{3/2}}\Bigr)\,\mathrm{e}^{-3\,u^2/8}. \end{aligned}$$

This yields of course that the norm function $x \in \ell^\infty \mapsto \|\delta\,x\|_\infty$ belongs to $L^{\Phi_2}(\mu)$. In particular, every linear mapping $a : E \to F$ that factors through a "sub-object" of β has the property that

$$\sup_{\xi \in B(E^*)} \bigl\| x \mapsto \langle \xi, x\rangle \bigr\|_{L^\Phi(\mu)} \leqslant 1 \quad \Rightarrow \quad \Bigl\| x \mapsto \|a\,x\|_F \Bigr\|_{L^\Phi(\mu)} \leqslant C \tag{52}$$

for some C and every probability measure μ on E. By such a “sub-object factorization”, we mean that there is a mapping $\alpha : E \to \ell^\infty$, linear and bounded, such that

$$\|a\,x\|_F \leqslant \|\beta \circ \alpha\, x\|_\infty, \qquad x \in E.$$

We will give a Φ_2-Pietsch measure for β. Let $(e_n)_{n\geqslant 0}$ denote the standard unit vector basis for $\ell^1(\mathbb{N})$, and consider the measure on the unit ball of $\ell^1(\mathbb{N})$ defined by

$$\mathrm{P}_\beta = \rho \sum_{i=0}^{\infty} \frac{1}{(i+3)^2}\, \delta_{e_i},$$

where $\rho > 0$ is chosen to make P_β a probability measure, namely

$$\rho^{-1} = \sum_{i=0}^{\infty} \frac{1}{(i+3)^2} = \frac{\pi^2}{6} - 1 - \frac{1}{4} < \frac{1}{2}, \quad \text{thus} \quad \rho > 2.$$

The probability measure P_β is in particular a measure on the unit ball of the dual of $\ell^\infty(\mathbb{N})$, and we shall see that

$$\|\beta x\|_\infty \leqslant 3 \,\bigl\|\xi \mapsto \langle \xi, x\rangle\bigr\|_{L^{\Phi_2}(\mathrm{P}_\beta)}, \qquad x \in \ell^\infty(\mathbb{N}).$$

Suppose that $\|\beta x\|_\infty > 3$: there exists an index $j \geqslant 0$ such that $\beta_j\,|x_j| > 3$, thus we have that $|x_j| > 3\sqrt{\ln(j+3)} > \sqrt{(16/3)\ln(j+3)}$. Then

$$\int \Phi_2(\langle \xi, x\rangle)\,\mathrm{d}\,\mathrm{P}_\beta(\xi) = -1 + \rho \sum_{i=0}^{\infty} \exp(3x_i^2/8)\,\frac{1}{(i+3)^2}$$

$$\geqslant -1 + \rho\,\frac{\exp(3x_j^2/8)}{(j+3)^2} > -1 + \rho\,\frac{\exp(2\,\ln(j+3))}{(j+3)^2} = -1 + \rho \geqslant 1. \qquad \square$$

4.4. *More old memories*

In his “radonifying works”, Schwartz introduced Bernoulli random variables and p-stable random variables, and it had an impact on the emphasis that was given some time later to random aspects in Banach space theory. One can trace it too in the isomorphic characterization (**29**) of Hilbert spaces by Stanisław Kwapień [Kw$_1$] and in the very influential results of Haskell Rosenthal [Ros] on reflexive subspaces of L^1. Not forgetting of course Jean-Pierre Kahane and his normed space valued Khintchine inequalities [Kah$_1$], a result to be also found in his important book [Kah$_2$]. Let us describe Kahane's result.

Consider a sequence $(\varepsilon_n)_{n\geqslant 1}$ of independent Bernoulli random variables, symmetric (meaning that $\mathrm{P}(\varepsilon_n = 1) = P(\varepsilon_n = -1) = 1/2$; for example, the Rademacher functions on $[0,1]$). The classical Khintchine inequalities state that for every $p \in (0,\infty)$, there are constants A_p and B_p such that

$$A_p \Bigl(\sum_{n=1}^{\infty} a_n^2\Bigr)^{1/2} \leqslant \Bigl(\mathrm{E}\bigl|\sum_{n=1}^{\infty} a_n\,\varepsilon_n\bigr|^p\Bigr)^{1/p} \leqslant B_p \Bigl(\sum_{n=1}^{\infty} a_n^2\Bigr)^{1/2}$$

for all real coefficients $(a_n)_{n\geqslant 1}$, with of course $A_2 = B_2 = 1$. It has been conjectured for a long time that the optimal constant A_1 is equal to $1/\sqrt{2}$ (what one gets for the sum $\varepsilon_1 + \varepsilon_2$, for which $\mathrm{E}\,|\varepsilon_1 + \varepsilon_2| = 1$) but it had to wait 1976 and Stanisław Szarek for a proof [Szar]. In 1978, Uffe Haagerup found the values of A_p and B_p for every $p \in (0,\infty)$ —the full version appeared in 1981 in [Haa$_1$]—; that particular result is a rather singular point in Haagerup's impressive work, some of which has crossed the path that we are following here, we shall mention an example below.

The Khintchine inequalities imply that all the moments of linear combinations of Bernoulli variables are equivalent, for $0 < p < \infty$. Kahane [Kah₁] proved that one can extend that equivalence to the case when the a_n are vectors in a normed space E and the modulus of real numbers is replaced by the norm in E: for every p, q with $0 < p < q < \infty$, there is a universal (30) constant $C_{p,q}$ such that

$$\Bigl(\mathrm{E}\bigl\|\sum_{n=1}^{\infty} \varepsilon_n\, a_n\bigr\|_E^q\Bigr)^{1/q} \leqslant C_{p,q}\,\Bigl(\mathrm{E}\bigl\|\sum_{n=1}^{\infty} \varepsilon_n\, a_n\bigr\|_E^p\Bigr)^{1/p}.$$

The real crux of the matter is Kahane's exponential estimate on the tail of the norm of those vector valued sums, that can be expressed as follows: assuming that

$$\mathrm{P}\Bigl(\bigl\|\sum_{n=1}^{\infty} \varepsilon_n\, a_n\bigr\|_E > 1\Bigr) \leqslant 1/4, \quad \text{it follows that} \quad \mathrm{P}\Bigl(\bigl\|\sum_{n=1}^{\infty} \varepsilon_n\, a_n\bigr\|_E > u\Bigr) \leqslant 2\,\mathrm{e}^{-c\,u},$$

for some universal $c > 0$ and every $u \geqslant 0$. The exponential behaviour given by Kahane is not optimal, the behaviour is actually sub-gaussian, as shown by Kwapień [Kw_2], and cannot be better, according to the central limit theorem.

Kwapień, Rosenthal and also Jørgen Hoffmann-Jørgensen [Hoff] are jointly partially responsible for the introduction of the notion of *type* of a Banach space. The notion made its way into the second volume of the main book for Banach space theory in the '80s, the *Classical Banach Spaces* by Lindenstrauss and Lior Tzafriri [LiTz]: let again $(\varepsilon_n)_{n\geqslant 1}$ denote a sequence of independent symmetric Bernoulli variables and let $p \in (1, 2]$. A Banach space E is a *type-p Banach space* if for some C and for every sequence of vectors (a_n) in E, one has (31)

$$\Bigl(\mathrm{E}\Bigl\|\sum_n \varepsilon_n\, a_n\Bigr\|_E^p\Bigr)^{1/p} \leqslant C\,\Bigl(\sum_n \|a_n\|_E^p\Bigr)^{1/p}.$$

Kahane's inequalities [Kah₁] imply that the pth moment on the left can be replaced by any other qth moment, $0 < q < \infty$ (with a possible change in C of course). We exclude the trivial case $p = 1$, which always holds true by the triangle inequality in E.

It is a fairly trivial matter to check that when $1 < p < \infty$, the Banach space L^p has type $\min(p, 2)$. The non-obvious case of the Schatten classes was given by Nicole Tomczak-Jaegermann [Tom], and was one of the first signs of the intrusion of non-commutativity in the so-called *local theory of Banach spaces*. A couple of years later, the (non-commutative) C^*-algebras came in, in attempts to extend some theorems involving the classical $C(K)$ spaces. And with it, the necessity of dealing with both $a\,a^*$ and a^*a when adapting the real "square function" $(\sum |a_n|^2)^{1/2}$, a change fundamental in Pisier's extension to C^*-algebras [Pis₃] of the Grothendieck theorem, as well as in the improvement (32) obtained a few years later by Haagerup [Haa₂].

Rosenthal [Ros] showed that any reflexive subspace R of $L^1(0, 1)$ can be "lifted" to $L^p(0, 1)$ for some $p > 1$ by a simple *change of density*: there is a function $\varphi > 0$ such that $\varphi \in L^q(0, 1)$, $1/q + 1/p = 1$ and

$$\Bigl(\int_0^1 \Bigl|\frac{x(u)}{\varphi(u)}\Bigr|^p \,\mathrm{d}u\Bigr)^{1/p} \leqslant \|x\|_{L^1}, \qquad x \in R.$$

For a Banach space E and $1 < p \leqslant 2$, Rosenthal used inequalities of the form

$$\mathrm{E}\bigl\|\sum f_n\, a_n\bigr\|_E \leqslant C\,\bigl(\sum \|a_n\|_E^p\bigr)^{1/p}$$

where (f_n) denotes a sequence of independent p-stable random variables and (a_n) any sequence of vectors in E. That E satisfies these inequalities was proved equivalent to E being of type p, not very long time after.

For $0 < p \leqslant 2$, a random variable f is p-stable when for some $c > 0$ one has

$$\mathrm{E}\, \mathrm{e}^{\mathrm{i}\, t f} = \mathrm{e}^{-c\,|t|^p}, \quad t \in \mathbb{R}.$$

Gaussian variables are an exceptional case of stable variables. When $p < 2$, a p-stable variable can be obtained by an integral combination of Poisson variables, and can thus be approximated by a series of Poisson variables. One can identify an abstract version of this construction mechanism in the work [Kriv] of Jean-Louis Krivine. Krivine proved that given an infinite dimensional Banach space E, there is at least one $p \in [1, \infty]$ such that E contains for every integer $n \geqslant 1$ almost isometric copies of ℓ^p_n, namely, of the space $\mathbb{R}^n$ equipped with the ℓ^p-norm. Some time before, Boris Tsirelson [Tsir] had disproved the corresponding infinite dimensional long-standing conjecture, by showing that: *not every Banach space contains ℓ^p or c_0*, exhibiting an example of a Banach space that fails to contain c_0 or any ℓ^p. Tsirelson constructed that example by a totally novel method that had a tremendous influence for finding out several well hidden phenomenons about infinite dimensional Banach spaces. One of several instances is the —negative— solution of the distortion problem (**33**) by Ted Odell and Thomas Schlumprecht [OdS]. After spending years in the field, I am tempted to say: if you think of a pleasant structural infinite dimensional property that a general Banach space might satisfy, then most probably you will see someone come with an example of a space that fails to have that nice property.

A year before Tsirelson's discovery, Per Enflo [Enf_2] had ruined the hope that every Banach space could have the *approximation property*, a notion that was thoroughly investigated by Grothendieck [Gro_3] around 1955. As above, Enflo has *constructed* an example of a space failing the approximation property, not a space that you would deal with every other day. However, Andrzej (Tomek) Szankowski [Szan] showed some years later that the classical space $B(H)$ of bounded operators on a separable Hilbert space H also fails to have the approximation property.

Enflo had another result at that time, notably more important in my view for the development of the theory. Robert C. James [Jam] had introduced the notion of a *super-reflexive* Banach space (**34**), and related it to a notion of *tree* in a vector space: for a *linear binary tree*, each node is a vector that has two children of which it is the midpoint. Enflo [Enf_1] showed that the super-reflexive spaces are precisely those that admit an equivalent uniformly convex norm. It was easy to connect the trees of James to the probabilistic notion of a vector valued *martingale*, and Pisier [Pis_1] was able to improve Enflo's theorem by extending results from "classical" martingale theory, in addition to the Gurarii–James theorem and a very nice modification of Enflo's renorming idea (**35**). Since then, *"Martingales in Banach Spaces"* has been a significant theme of that field.

Around 1976, Jean Bourgain begins his incredible career. The list of his main results would be longer than this already long report. After several of his stunning achievements, some will start calling him "God"! (**36**) I chose quasi-randomly to mention here a single result, one about $\Lambda(p)$-sets (**37**): the work of Rudin [Rud] has provided an example of a $\Lambda(4)$-set in $\mathbb{Z}$ that is not a $\Lambda(q)$-set for any $q > 4$, but no example was known of a $\Lambda(p)$-set with $2 < p < 4$ that is not $\Lambda(q)$ for any $q > p$. Bourgain [Bour] gives such examples by a delicate analysis of the properties of randomly selected subsets of a finite set of characters, showing once more his amazing power of breaking through walls that nobody else could.

4.5. *Closing the path*

The young people in the (very) little group of researchers interested in Functional Analysis and Banach Spaces in the '70 s at the "Centre de Mathématiques" directed by Laurent Schwartz at École Polytechnique, Paris, were not afraid of formulating rather bold hypotheses, I do not dare to call them *conjectures*. We had almost no reasonable support for them, and so it was just as probable to win twice at the lottery than see two of them proved true in the following years. One was about the *K-convexity* ([38]), and was solved by Pisier [Pis_4] ([39]). The second one was a strong belief in an affirmative answer to the following question: *is every γ-summing linear operator a Φ_2-summing one?* Here, being γ-summing simply means that the image of the standard Gaussian cylindrical measure γ_H of the Hilbert space will be a true measure on the range space; it is clear that every Φ_2-summing map is γ-summing.

That the answer to the question is positive is a consequence of the result given in the previous section 3. Indeed, an immediate reason for γ-summing maps to be Φ_2-summing as well is that the estimation of the expectation of the supremum of a Gaussian process satisfying the conclusion (44) only uses (7), which is also true —by definition— for sub-gaussian variables. And having a uniform scalar Φ_2-estimate for a measure μ on a normed space E amounts to having a sub-gaussian process indexed by $x^* \in B_{E^*}$. But we will work in the rest of the section to give in addition a sort of Pietsch factorization for γ-summing maps, a factorization through a somewhat canonical Φ_2-summing map.

Let $a : H = \ell^2(\mathbb{N}) \to F$ be a γ-summing linear bounded operator; then the image of the cylindrical measure γ_H on H is a Gaussian Radon measure μ on F, that we can describe as the limit of the images $a(\gamma_n)$, where γ_n is the standard Gaussian probability measure on the n-dimensional subspace H_n of $\ell^2(\mathbb{N})$ consisting of all vectors $(x_i)_{i\in\mathbb{N}}$ with $x_i = 0$ when $i \geqslant n$. We then have, for example by [Fer_1], that

$$\int_F \|y\|_F \,\mathrm{d}\mu(y) < \infty, \quad \text{and} \quad \int_F \|y\|_F \,\mathrm{d}\mu(y) = \lim_n \int_{H_n} \|a\,x\|_F \,\mathrm{d}\gamma_n(x),$$

therefore, introducing the Gaussian measure Γ on $\mathbb{R}^{\mathbb{N}}$ from section 4.1, we have

$$\begin{aligned} &\lim_n \int_{H_n} \|a\,x\|_F \,\mathrm{d}\gamma_n(x) \\ &= \lim_n \int_{H_n} \sup_{y^*\in B_{F^*}} \langle a^*y^*, \omega\rangle \,\mathrm{d}\gamma_n(\omega) \\ &= \int_{\mathbb{R}^{\mathbb{N}}} \sup_{y^*\in B_{F^*}} \langle a^*y^*, \omega\rangle \,\mathrm{d}\Gamma(\omega) = \int_F \|y\|_F \,\mathrm{d}\mu(y) < \infty. \end{aligned} \tag{53}$$

We recall that when $\xi = (\xi_i) \in \ell^2(\mathbb{N})^* \simeq \ell^2(\mathbb{N})$, the series $\langle \xi, \omega\rangle = \sum_{i\in\mathbb{N}} \xi_i\omega_i$ is convergent for Γ-almost all $\omega = (\omega_i) \in \Omega = \mathbb{R}^{\mathbb{N}}$, and actually defines a centered Gaussian random variable on (Ω, Γ) with variance $\sum \xi_i^2$.

Let $T = a^*(B_{F^*}) \subset H^*$ be the image of the unit ball of the dual F^* by the adjoint mapping a^*. For each $t \in T$, define $X_t : \omega \mapsto \langle t, \omega\rangle \in L^2(\Omega, \Gamma)$. This is a Gaussian process, and saying that the mapping a is γ-summing implies that

$$\mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr) < \infty,$$

a mere restatement of (53). We shall use (44) to see that a admits a factorization through a somehow "canonical" model: we decide quite arbitrarily to declare *canonical* the diagonal map β already seen twice before, that has diagonal coefficients

$$\beta_n = \frac{1}{\sqrt{\ln(n+3)}}, \quad n \geqslant 0.$$

We choose a countable dense subset $\{d_j\}_{j \geqslant 0}$ in the unit ball of $H^* \simeq \ell^2(\mathbb{N})$, and we know by (44) that there is a constant $K_2 = K_2(T)$ and a sequence of finite sets $T_k \subset T$, with $k \geqslant 0$, such that

$$|T_0| = 1, \quad \ln|T_k| < 3\,b^k, \quad \sum_{k \geqslant 0} \mathrm{dist}(t, T_k)\, b^{k/2} \leqslant K_2, \quad t \in T, \tag{54}$$

where distances are evaluated in $\ell^2(\mathbb{N})$; note that

$$\|X_s - X_t\|_{L^2(\Gamma)} = \|s - t\|_{\ell^2(\mathbb{N})}.$$

We may decide to set $t_0 = 0 \in T$, so $T_0 = \{0\}$. For every couple $(s, t) \in T_k \times T_{k+1}$ such that $d(t, s) > 0$, consider the norm one functional

$$\xi(s, t) = \frac{1}{d(t, s)}\,(t - s) \in H^*$$

and complete the definition with $\xi(s, t) = 0$ when $d(t, s) = 0$. For $k \geqslant 0$, let

$$\Xi_k = \{\xi(s, t) : (s, t) \in T_k \times T_{k+1}\} \cup \{d_k\} \subset B_{H^*}.$$

We have $|\Xi_k| \leqslant 1 + |T_k|\,.\,|T_{k+1}|$ and thus

$$\ln|\Xi_k| \leqslant 1 + 3\,b^k + 3\,b^{k+1} < (1 + 3 + 3\,b)\,b^k < 7\,b.b^k \tag{55}$$

because $\ln(1 + u) < 1 + \ln u$ when $u \geqslant 1$, and $b > 1$. We consider a listing $(x_n^*)_{n \geqslant 0}$ of all elements in the union

$$\bigcup_{k \geqslant 0} \Xi_k$$

where we may have to repeat functionals and have $x_m^* = x_n^*$, with $m \neq n$, if that same functional happens to belong to two different sets Ξ_k, and where the elements of Ξ_k are listed before those of Ξ_{k+1}. Let

$$I_k = \{n : x_n^* \in \Xi_k\}.$$

The sets I_k are disjoint and cover $\mathbb{N}$. The sequence (x_n^*) is contained in the unit ball B_{H^*} and contains the (d_j) as subsequence, hence it is dense in B_{H^*}. We define a first linear mapping f from $\ell^1(\mathbb{N})$ to H^* by letting

$$f(e_n) = x_n^*,$$

where $(e_n)_{n \geqslant 0}$ is the standard unit vector basis for $\ell^1(\mathbb{N})$. The adjoint map f^* from H to $\ell^\infty(\mathbb{N})$ acts on vectors $x \in H$ by

$$f^*(x) = \bigl(\langle x_n^*, x\rangle\bigr)_{n \geqslant 0}, \quad \text{and} \quad \|f^*(x)\|_\infty = \sup_n \bigl|\langle x_n^*, x\rangle\bigr| \geqslant \sup_j \bigl|\langle d_j, x\rangle\bigr| = \|x\|_H \tag{56}$$

because the sequence (d_j) was chosen dense in the unit ball of H^*, and $\|f^*(x)\|_\infty \leqslant \|x\|_H$ because the x_n^* belong to the dual unit ball. The mapping f^* is therefore an isometry from H into $\ell^\infty(\mathbb{N})$.

Next, we need find a diagonal map $\delta = (\delta_n)_{n \in \mathbb{N}}$ on $\ell^\infty(\mathbb{N})$ such that $\sqrt{\ln(n+3)}\,\delta_n$ is bounded and such that for some C, we have

$$\|a\,x\|_F \leqslant C\,\|\delta\,f^* x\|_\infty, \quad x \in H.$$

That would allow us to factorize the mapping $a : H \to F$ through the restriction to the subspace $f^*(H) \subset \ell^\infty(\mathbb{N})$ of the Φ_2-summing map $\delta : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$. By construction, we know that

$$\mathbb{N} = \bigcup_{k=0}^{\infty} I_k,$$

where the I_k are pairwise disjoint subsets. We define the coefficients $(\delta_n)_{n\in\mathbb{N}}$ for the required diagonal map δ by

$$\delta_n = b^{-k/2} \quad \text{when} \quad n \in I_k,$$

and we define $\delta : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ by

$$\delta\big((x_n)_{n\in\mathbb{N}}\big) = (\delta_n x_n)_{n\in\mathbb{N}}.$$

Let us denote by $\delta^{(1)} : \ell^1(\mathbb{N}) \to \ell^1(\mathbb{N})$ the restriction of δ to $\ell^1(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$. Clearly, the map δ is the adjoint $(\delta^{(1)})^*$ to $\delta^{(1)}$. For the unit vector basis (e_n) in $\ell^1(\mathbb{N})$, we have

$$\delta^{(1)}(e_n) = \delta_n\, e_n, \quad n \in \mathbb{N}.$$

Given any point $\tau \in T$, we can find a sequence $(t_k(\tau))_{k\geqslant 0}$ such that $t_k(\tau) \in T_k$ and such that $\operatorname{dist}(\tau, t_k(\tau)) = \operatorname{dist}(\tau, T_k)$ for each $k \geqslant 0$. We know by (54) that $\operatorname{dist}(\tau, T_k)$ tends to 0. It follows that

$$\tau = \sum_{k=0}^{\infty} \big(t_{k+1}(\tau) - t_k(\tau)\big)$$

in H^* (remember that we chose $t_0 = t_0(\tau) = 0$). Then observe that

$$t_{k+1}(\tau) - t_k(\tau) = d(t_k(\tau), t_{k+1}(\tau))\, x^*_{n_k}(\tau)$$

for a certain $n_k \in I_k \subset \mathbb{N}$, where

$$x^*_{n_k}(\tau) = \xi(t_k(\tau), t_{k+1}(\tau)) \in \Xi_k.$$

Now $x^*_{n_k}(\tau) = f(e_{n_k})$ and $\delta^{(1)}(e_{n_k}) = b^{-k/2} e_{n_k}$. Hence $x^*_{n_k}(\tau) = b^{k/2} (f \circ \delta^{(1)})(e_{n_k})$ and

$$\begin{aligned}\tau = \sum_{k=0}^{\infty} \big(t_{k+1}(\tau) - t_k(\tau)\big) &= \sum_{k=0}^{\infty} d(t_k(\tau), t_{k+1}(\tau))\, x^*_{n_k}(\tau) \\ &= \sum_{k=0}^{\infty} d(t_k(\tau), t_{k+1}(\tau))\, b^{k/2} (f \circ \delta^{(1)})(e_{n_k}).\end{aligned}$$

So, we see that $\tau \in T \subset H^*$ is the image by $f \circ \delta^{(1)} : \ell^1(\mathbb{N}) \to H^*$ of the element

$$m(\tau) = \sum_{k=0}^{\infty} d(t_k(\tau), t_{k+1}(\tau))\, b^{k/2} e_{n_k} \in \ell^1(\mathbb{N})$$

that sits in a fixed ball in $\ell^1(\mathbb{N})$, because

$$\|m(\tau)\|_1 = \sum_{k=0}^{\infty} d(t_k(\tau), t_{k+1}(\tau))\, b^{k/2} \leqslant 2 \sum_{k=0}^{\infty} d(\tau, T_k)\, b^{k/2} \leqslant 2\, K_2.$$

For $x \in H$ we have

$$\langle \tau, x \rangle = \langle (f \circ \delta^{(1)})\, m(\tau), x \rangle = \langle m(\tau), \delta\, f^* x \rangle \leqslant 2\, K_2\, \|\delta\, f^* x\|_\infty.$$

It follows that

$$\|a\, x\|_F = \sup_{\tau \in T} \langle \tau, x \rangle \leqslant 2\, K_2\, \|\delta\, f^* x\|_\infty, \tag{57}$$

and this indicates a possible factorization: indeed, we can draw the following diagram:

$$\begin{array}{ccccccc} & & \ell^\infty & \xrightarrow[\delta]{} & \ell^\infty & & \\ & & \cup & & \cup & & \\ H & \xrightarrow[f^*]{} & E_\infty & \xrightarrow[\delta_1]{} & F_\infty & \xrightarrow[a_1]{} & F \end{array} \tag{58}$$

where f^* is the isometric embedding of H into $\ell^\infty(\mathbb{N})$ given at (56), E_∞ is the image of H under f^*, δ_1 is the restriction to E_∞ of the diagonal map δ, F_∞ is the closure of the image of E_∞ under δ_1, and a_1 is defined so that $a = a_1 \circ \delta_1 \circ f^*$. Equation (57) shows that

$$\|a_1\| \leqslant 2\,K_2.$$

It remains to check that in the above diagram, the map δ is Φ_2-summing —this will follow from the calculations of section 4.3 on diagonal operators—. The map δ_1 will then be Φ_2-summing, as restriction of the Φ_2-summing map δ, and we shall get the wanted factorization

$$a = a_1 \circ \delta_1 \circ f^*.$$

For proving that δ is Φ_2-summing, we will show that the diagonal coefficients (δ_n) decrease at least as fast as those of the Φ_2-summing map β. We defined $\delta_n = b^{-k/2}$ when $n \in I_k$, we thus need to see that for some C and whenever $n \in I_k$, we have

$$\delta_n = b^{-k/2} \leqslant C\,\beta_n = \frac{C}{\sqrt{\ln(n+3)}}, \quad \text{or} \quad \ln(n+3) \leqslant C^2\,b^k, \quad k \geqslant 0. \tag{59}$$

Knowing that $n \in I_k$ and letting $c = 7\,b$, we have by (55) that

$$n \leqslant \sum_{s=0}^{k} |I_s| = \sum_{s=0}^{k} |\Xi_s| \leqslant \sum_{s=0}^{k} \exp(c\,b^s).$$

Because $b > 1/(\ln 2) > 1$ we can check (40) that $c\,(b-1) > \ln 2$, and when $s \geqslant 1$,

$$b^s - b^{s-1} \geqslant b - 1, \quad \exp(c\,b^{s-1}) \leqslant \mathrm{e}^{c(1-b)} \exp(c\,b^s) < (1/2)\exp(c\,b^s).$$

It follows that

$$n \leqslant \Bigl(\sum_{n=0}^{k} 2^{-n}\Bigr) \exp(c\,b^k) < 2\exp(7b\,.\,b^k)$$

and from this, we get

$$\ln(n+3) \leqslant \ln\bigl(3 + 2\exp(7b\,.\,b^k)\bigr) \leqslant 3 + \ln 2 + 7b\,.\,b^k < (11\,b)\,.\,b^k,$$

that is to say, Inequality (59) with $C = \sqrt{11\,b}$. □

If we insist on introducing the “canonical” object β, then because $|\delta_n| \leqslant C\,\beta_n$, we can factor $\delta : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ as $\delta = \alpha \circ \beta$ where $\alpha : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ is diagonal with coefficients $\alpha_n = \delta_n/\beta_n$ bounded by C. We may rewrite (57) in the form

$$\|a\,x\|_F \leqslant 2\,K_2\|\delta\,f^*x\|_\infty = 2\,K_2\|\alpha\beta\,f^*x\|_\infty \leqslant 2C\,K_2\|\beta\,f^*x\|_\infty,$$

then modify the diagram (58) and factor a through a “sub-object” β_1 of our canonical diagonal mapping β,

$$\begin{array}{ccccccc} & & \ell^\infty & \xrightarrow[\beta]{} & \ell^\infty & & \\ & & \cup & & \cup & & \\ H & \xrightarrow[f^*]{} & E_\infty & \xrightarrow[\beta_1]{} & G_\infty & \xrightarrow[a_2]{} & F \end{array}$$

where β_1 is the restriction to E_∞ of the diagonal map β, and G_∞ is the closure of the image of E_∞ under β_1. We get $a = a_2 \circ \beta_1 \circ f^*$ with $\|a_2\| \leqslant 2CK_2$.

Notes

(1) There are obvious and well known difficulties when one is trying to consider

$$\sup_{t\in T} X_t(\omega),$$

where T is *uncountable* but each X_t is merely a *class* of random variables (an uncountable union of negligible sets is no longer negligible...). This is not our main concern here; we shall be happy enough to be able to deal with countable index sets T.

(2) We shall only be interested in the situation where the expectation of the supremum of the (centered) process is finite,

$$E^* := \mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr) < \infty,$$

and this implies that T is bounded for the distance d: indeed, if $s, t \in T$, then

$$E^* \geqslant \mathrm{E}\max(X_t, X_s) = \mathrm{E}\Bigl(\max(X_t, X_s) - X_s\Bigr) = \mathrm{E}\max(X_t - X_s, 0);$$

this is the expectation of the positive part of a centered Gaussian random variable of variance $d(s,t)^2$,

$$\mathrm{E}\max(X_t - X_s, 0) = d(s,t) \int_0^\infty u\, \mathrm{e}^{-u^2/2} \,\frac{\mathrm{d}u}{\sqrt{2\pi}} = \frac{d(s,t)}{\sqrt{2\pi}} \leqslant E^*,$$

and it follows that the diameter Δ of T satisfies $\Delta \leqslant \sqrt{2\pi}\, E^*$.

(3) in *Regularity of Gaussian processes*, Acta Math. 159 (1987), 99–149.

(4) One can check rather easily by running a pseudo-random estimation that the inequality $\mathrm{E}\, g_5^* > 1.162$ is likely to hold true. It is quite possible that *Mathematica* or another software of the kind can immediately give the formulas claimed here; I posted my own calculations at

https://webusers.imj-prg.fr/~bernard.maurey/articles/MaxiGauss.pdf

I was not able to find a closed-form expression for $\mathrm{E}\, g_6^*$ (nor for the higher orders, needless to say). A numerical computation leads to $\mathrm{E}\, g_6^* > 1.267206$, thus $\mathrm{E}\, g_6^*/\sqrt{\ln 6} > 0.946$. The next computed value $\mathrm{E}\, g_7^* > 1.352178$, with $\mathrm{E}\, g_7^*/\sqrt{\ln 7} > 0.969$, does not contradict the hypothesis that $\mathrm{E}\, g_n^*/\sqrt{\ln n}$ is actually increasing with $n \geqslant 2$.

(5) Not easy, but not awfully difficult. What one has to do is compute the derivative of the function

$$u \;\mapsto\; \mathrm{E}\Bigl(\sup_{t\in T}\bigl((1-u)Y_t + uX_t\bigr)\Bigr)$$

for $u \in [0,1]$, in the case when (Y_t) and (X_t) are independent and T is a finite set. The Slepian–Sudakov hypothesis allows one to see that the derivative is non-negative.

(6) Simone Chevet in [Che_1] does not study the *expectation* of the supremum, but its *distribution*, and she writes a proof for the Slepian lemma; however, Fernique in [Fer_2]

claims to see between the lines of Chevet's article a proof for Proposition 1; he himself writes on page 63 a one line justification for that Proposition, namely:

$$4\,\frac{d}{d\alpha}\big[\{\sup_{t\in T} Z_\alpha(t)\}\big] \;=\; \sum_{\substack{s,t\in T\times T\\ s\neq t}} \frac{d}{d\alpha}[\Delta_{Z_\alpha}(s,t)] \int \frac{dx}{dx_s\,dx_t} \int g_\alpha(x)\,du,$$

only precising that the last integral is done on the domain $x_s = x_t = \sup x_i = u$. If I try be a little understandable, I have to add that

$$Z_\alpha = \sqrt{\alpha}\,X + \sqrt{1-\alpha}\,Y, \quad 0 \leqslant \alpha \leqslant 1,$$

where X and Y are independent copies of the two processes from Proposition 1, and say that Δ_Z denotes the L^2-metric associated to a process Z. Also, T is supposed to be finite here, and g_α is the function on $\mathbb{R}^T$ equal to the density (supposed to exist) of the distribution of Z_α.

It seems that Fernique meant to have an expectation in the left-hand side of the main equality above. He gave the details of the proof in Chap. 2 of [Fer$_3$].

(7) Paul Lévy's isoperimetric inequality for the sphere $S^{n-1} \subset \mathbb{R}^n$ implies that the measure of the ε-enlargement A_ε of a set A of probability $1/2$ on that sphere is larger than the probability of the ε-enlargement of a halfsphere. The latter value can be computed explicitely and is very close to 1 when n is large; the isoperimetric result therefore implies that for some $c > 0$, one has

$$\sigma_{n-1}(A_\varepsilon) \geqslant 1 - c\,\mathrm{e}^{-n\varepsilon^2/2}\,.$$

The enlargement

$$A_\varepsilon = \{x \in S^{n-1} : \operatorname{dist}(x, A) < \varepsilon\}$$

is taken with respect to the geodesic distance d on the sphere, and here, σ_{n-1} is the invariant *probability* measure on S^{n-1}. This is a *concentration of measure phenomenon*: in a high dimension n, for any given set $A \subset S^{n-1}$ of measure $1/2$, most of the mass of σ_{n-1} sits around the boundary of A, and by a union bound estimate for the complements, it allows one to see that the intersection of many such enlargements will not be empty. It was used by Vitali Milman for giving a proof of the Dvoretzky theorem on the existence of Euclidean sections of convex sets in high dimension [Mil]. Milman's approach was extended in an important paper by Tadeusz Figiel, Lindenstrauss and Milman [FLM].

Proofs using concentration of measure on the sphere can often be replaced by Gaussian proofs, because the standard Gaussian probability measure γ_n on $\mathbb{R}^n$ is essentially supported on a sphere (of radius $\sqrt{n}$), and because nice concentration results exist for the Gaussian measure.

(8) The *packing problem* for solid balls consists of finding an optimal arrangement of these balls, so as to fit a maximal number of them in a given space. One can say that a finite set S is *δ-packing* when it is δ-separated: then, the balls of radius $\delta/2$ centered at the points of S are disjoint and thus satisfy the requirement of the packing problem. A *δ-net* S for a set A is a subset $S \subset A$ such that the balls $B(s,\delta)$ that are centered at the points $s \in S$ cover A. One sees easily that a *maximal* δ-packing set S for a set A is at the same time a δ-net for A. Note that most often in the literature, the covering balls are supposed to be *closed* and the δ-separation *strict*. We found more convenient to turn it around, with *open* covering balls and δ-separation defined by $d(s,t) \geqslant \delta$.

(9) The general notion of mixed volume involves n convex sets $K_1, K_2, \ldots, K_n$ in $\mathbb{R}^n$; the mixed volume $V(K_1, K_2, \ldots, K_n)$ is a number $\geqslant 0$ obtained from the coefficients that appear in the expansion as a polynomial in the —non-negative— variables $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the n-dimensional volume

$$\Bigl|\sum_{i=1}^{n} \lambda_i K_i\Bigr|_n$$

of the Minkowski sum $\sum_{i=1}^{n} \lambda_i K_i$. In our account of the cited work of Sudakov, only the special case $V(K, B_n, B_n, \ldots, B_n)$ appears, with B_n the Euclidean unit ball in $\mathbb{R}^n$.

(10) This lemma, a rather simple application of the Slepian–Sudakov comparison result, does not appear in the original proofs by Fernique. I learned about it by attending lectures about the so-called *generic chaining* method of Talagrand, given at Marne-la-Vallée around 2002 by (by then) young researchers there. One could claim with a bit of exaggeration and some *mauvaise foi* that once the lemma and its Corollary 2 are set in place, the proof of the Fernique theorem reduces to an almost mechanical tree-manipulation.

(11) There is nothing magical about this cube N^3. We could rewrite everything with another power N^α, as long as $\alpha > 1$.

(12) Actually, it is the *logarithm* of N_i that grows like a power: we have that

$$\ln N_{i+1} \geqslant 3 \ln N_i,$$

and N_i is thus enormously larger than 2^i, of order at least $\exp(c3^i)$ when $i \geqslant 1$, where we can set $c = (\ln 2)/3$, according to the fact that $N_1 \geqslant 2$.

(13) This is known as *König's infinity lemma*, 1927, an easy but logically interesting result.

(14) We do not actually need a group action; all we need is that for any two balls $B(t_1, r)$ and $B(t_2, r)$ in T, considered as subset of L^2, there is an onto affine L^2-isometry between them, sending t_1 to t_2. Of course this family of isometries will generate a group, but we may ignore it completely.

(15) If T has no isolated point, all branches produced by our process will be "naturally" infinite. We could arrange to work with a set T with no isolated point, for example by replacing T by its convex hull in $L^2(\Omega, \mathrm{P})$, but the added complexity would ruin the small simplification of not having to deal with isolated points —we could then say that the nodes are just balls—.

(16) In a previous version, I described the tree $\mathcal{X}$ as a collection of couples $x = (t, r)$, with $t \in T$ and $r > 0$, corresponding to what is now the couple of the latest position t in x and of the radius $r = \rho(x)$ associated to the node x. This was not correct for the following reason: I want to keep the *construction step* of the children of a given node x_* as simple as possible, and in doing so, it becomes possible that the same couple (t, r) will appear in two different descendants of x_*, thus ruining the tree structure. We could remedy this problem and be able to consider that the couples (t, r) are indeed the nodes of the tree, in different ways that will each complicate too much the construction step. We could use a tiny variation of each radius r in order to encode the positions of the ancesters of the node (t, r): then the *radii* of the descendants will never coincide. Or, instead of using a δ-p-net in the ball $B(t_*, r_*)$ for defining the positions of the children of x_*, we could have defined inductively a subset $A_* \subset B(t_*, r_*)$ so that the sets A_*

corresponding to a given level in the tree be disjoint, yet cover T, and we could have defined the children of x^* using a δ-p-net in A_*: then the *positions* of the descendants will never coincide; this is what we do in a later situation far below, where a little more complication would not be so noticeable.

(17) We may encounter degenerate cases $x \succ (t, r)$ where the ball $B(t, r)$ is a finite set with less than N^3 points: this may be said "very poor". Note that in this situation, each point $s \in B(t, r)$ is isolated in T. However, this fact does not play a rôle in the further discussion.

(18) Remember that we decide to set $r_{-1} = 3r_0$. In the construction of the tree we insisted that $r_{i+1} = r_i/3$, but this will not appear in the reverse direction that comes next: only $\lim_i r_i = 0$ will be used.

(19) When T is finite, we have of course $t_i(t) = t$ when $i \geqslant i_0$. Then the successive nodes in a branch have the form

$$x_j = (t_0, t_1, \ldots, t_{i_0-1}, t, t, \ldots, t), \quad j \geqslant i_0,$$

and are only distinguished by the length of the sequence x_j. Our "unnatural" treatment was meant to avoid considering the finite case separately.

(20) It is a remarkable and very important fact that, due to the absolutely huge growth of the constants N_i, the logarithm of the number M_i of points in the ith generation is comparable to the logarithm of the number N_i of children of a *single node* in the $(i-1)$th generation.

(21) Any sequence n_i with $\sum n_i < \infty$ could be used in place of $n_i = N_i$.

(22) In the invariant case, we did not insist that the ball $B(t_{i+1}, r_{i+1})$ associated to a child x_{i+1} of $x_i = (t_i, r_i, N_i)$ be *contained* in the ball $B(t_i, r_i)$ associated to its parent x_i. But here, we need to know that for $k > 1$, the points in the regions V_{i+k} of the next generations will still satisfy the homogeneity condition that was set before for the points of V_i. We ensure $V_{i+k} \subset V_i$ by imposing that $V_{i+1} \subset V_i$ for every $i \geqslant 0$.

(23) Now that the sets V at a given level in the tree are disjoint, we could define the nodes of the tree simply as couples

$$(V, t), \quad \text{with} \quad t \in T, \;\; r > 0, \;\; \emptyset \neq V \subset B(t, r).$$

Indeed, if $\xi = (V_i, t_i)$ is such a "new node" at level $i > 0$, its parent (V_{i-1}, t_{i-1}) would be the *unique* node at level $(i-1)$ such that V_{i-1} contains V_i, and we might inductively retrieve without ambiguity the entire past $(t_0, t_1, \ldots, t_i)$ of that node ξ.

(24) For each $i \geqslant 0$, let X_i be the finite set of nodes at level i in the tree $\mathcal{X}$, where X_i is equipped with the discrete topology. The product space $P = \prod_{i \geqslant 0} X_i$ with the product topology is compact by the Tikhonov theorem, and the set $\mathbf{X}$ of branches in $\mathcal{X}$ is closed in P: indeed, for each fixed $j \geqslant 0$, the set of $\mathbf{x} = (x_i)_{i \geqslant 0} \in P$ such that x_{j+1} is a child of x_j is closed in P, as it is defined by a finite number of conditions on the "coordinates" of $\mathbf{x}$, namely, that the couple $(x_j(\mathbf{x}), x_{j+1}(\mathbf{x}))$ belong to the finite set of parent-child couples at level j for the tree $\mathcal{X}$.

The set of branches $\mathbf{x}$ such that $x_i(\mathbf{x}) = x_i^*$ for some fixed node x_i^* at level i is both closed and open in $\mathbf{X}$. Hence, in the case of our tree consisting of triples $x = (V, t, r)$ with $t \in T$, the mapping $\mathbf{x} \mapsto t_i(\mathbf{x})$ is continuous from $\mathbf{X}$ to T, and the projection π

is a uniform limit of those continuous mappings, because $d(t_{i+1}(\mathbf{x}), t_i(\mathbf{x})) < r_i$ for each integer $i \geqslant 0$ and $\mathbf{x} \in \mathbf{X}$ —we know that $t_{i+1}(\mathbf{x}) \in V_i(\mathbf{x}) \subset B(t_i(\mathbf{x}), r_i)$ by $\mathbf{a}_1^*$—.

(25) If we have performed the *disjointification* of the preceding section 3.3, we can see each function $i_k(\cdot)$ as a *stopping time* relative to the fields $(\mathcal{F}_i)$.

(26) We did not give a *verbatim* statement in our Equation (44). The original result would have $b = 2$ and $3\,b^k$ replaced by $(\ln 2)\,b^k$.

(27) The notion of p-summing map can be considered for every $p > 0$, but $\pi_p(a)$ is a norm for $p \geqslant 1$ only. Otherwise it is a *quasi-norm*, just as what one has for the space L^p when $0 < p < 1$.

(28) Hence Pisier's theorem [Pis$_2$] can be rephrased to say that: *a set $\Lambda \subset \mathbb{Z}$ is a Sidon set when all integrable functions on $\mathbb{T}$ with spectrum in Λ belong to $L^{\Phi_2}(\mathbb{T}, m)$.*

(29) Let (f_n) be a sequence of symmetric Bernoulli random variables, or a sequence of $\mathrm{N}(0,1)$ Gaussian variables, independent in both cases, and let H be a Hilbert space. Introducing an orthonormal basis for H, it is fairly easy to see that

$$\Bigl(\mathrm{E}\Bigl\|\sum f_n\,a_n\Bigr\|_H^2\Bigr)^{1/2} = \bigl(\sum \|a_n\|_H^2\bigr)^{1/2}$$

for every sequence (a_n) of vectors in H. In the opposite direction, Kwapień [Kw$_1$] shows that a normed space E is *isomorphic* to a Hilbert space if and only if there is a constant $C > 0$ such that

$$C^{-1}\,\bigl(\sum \|a_n\|_E^2\bigr)^{1/2} \leqslant \Bigl(\mathrm{E}\Bigl\|\sum f_n\,a_n\Bigr\|_E^2\Bigr)^{1/2} \leqslant C\,\bigl(\sum \|a_n\|_E^2\bigr)^{1/2}$$

for every sequence (a_n) of vectors in E. Kahane's inequalities, or Shepp–Landau–Fernique's show that in the expectation above, the second moment can be replaced by any pth moment with $0 < p < \infty$ —but with a change of C—. That remark being done, we see that Kwapień's result tells us that the Hilbert space is the only normed space where the Khintchine inequalities can be rewritten by simply changing the series of squares of real coefficients into the series of squares of norms of vector coefficients.

(30) That is to say, a constant that does not depend upon anything else than p and q, perhaps $C_{p,q} = 256$ for some p and q. We restricted to $p < q$: when $p \geqslant q$, the inequality remains true but reduces to Hölder's, and then $C_{p,q} = 1$.

(31) For $q \geqslant 2$, the space E has *cotype q* when it satisfies inequalities opposite to the type-p inequalities,

$$\Bigl(\mathrm{E}\Bigl\|\sum_n \varepsilon_n\,a_n\Bigr\|_E^q\Bigr)^{1/q} \geqslant C^{-1}\,\Bigl(\sum_n \|a_n\|_E^q\Bigr)^{1/q},$$

for some $C > 0$ and all vectors a_n in E. Cotype-2 spaces have some interesting properties, for instance regarding their Euclidean sections, see [FLM]: there exists $\lambda_C > 0$ such that every n-dimensional cotype-2 space with cotype-2 constant $\leqslant C$ has subspaces F with $\dim F \geqslant \lambda_C\,n$ and Banach–Mazur distance $\leqslant 2$ (say) to a Euclidean space —a "proportional dimension", the best one can hope for—.

The space L^1 has cotype 2, hence ℓ_n^1 has large Euclidean subspaces. This was obtained by Boris Kashin [Kash], who actually proved more: there is a constant C such that for every $n \geqslant 1$, one can decompose ℓ_{2n}^1 into two n-dimensional subspaces, orthogonal in ℓ_{2n}^2 and C-isomorphic to ℓ_n^2.

The fact that for some universal $\beta > 1$ and every integer $n \geqslant 1$, the space $\ell^1_{\beta n}$ has a subspace 2-isomorphic (say) to ℓ^2_n has been generalized in [JoSc] by Johnson and Gideon Schechtman: for $0 < r < s < 2$, the space $\ell^r_{\beta n}$ has a subspace 2-isomorphic to ℓ^s_n, with $\beta = \beta(r, s)$. This was extended by Pisier [Pis_5], who proved a result involving the *stable type p constant* of a general Banach space E, let us denote it here by $C_p(E)$; for $1 \leqslant p < 2$, this theorem relates a large value of that constant to the existence of "large" ℓ^p_n subspaces in E, having dimension $n \sim C_p(E)^{1-1/p}$.

(32) The result of Pisier in [Pis_3] was stated under a certain *approximation property* assumption, that was removed by Haagerup in [Haa_2].

(33) If Tsirelson taught us that we can't expect to find c_0 or ℓ^p in every (infinite dimensional) Banach space, perhaps could we at least hope for some "regularity" in the vicinity of the Hilbert space: if E is *isomorphic* to $\ell^2(\mathbb{N})$, can we find an infinite dimensional subspace E_0 of E that is "well" isomorphic to a Hilbert space, say, such that the Banach–Mazur distance between E_0 and the Hilbert space be less than 2? This is not true, even if the bound 2 is replaced by any other bound $C > 2$ [OdS].

(34) A Banach space E is *finitely representable* in another space F when for every $\varepsilon > 0$, every finite dimensional subspace of E is $(1 + \varepsilon)$-isomorphic to a subspace of F. For example, the space $E = L^1(0, 1)$ is finitely representable into $F_0 = \ell^1$, or also into the space $F_1 = \big(\bigoplus_{n=1}^{\infty} \ell^1_n\big)_{\ell^2}$ that is the ℓ^2-sum of the spaces ℓ^1_n of increasing dimensions n. A space F is *super-reflexive* when every space E finitely representable in F is reflexive. Then, the space F_1 above is reflexive but not super-reflexive.

James showed that a space E is super-reflexive precisely when for every $\varepsilon \in (0, 1]$, there is a limit $N(\varepsilon) < \infty$ to the depth of finite linear binary trees in the unit ball of E that satisfy that the distance between any node (that is not a leaf) and its children is always $\geqslant \varepsilon$: the norm of the nodes of such "ε-trees" must grow beyond 1 (actually, and clearly, beyond any given finite limit) if one tries to extend those trees indefinitely.

(35) Enflo showed that a super-reflexive space E admits an equivalent norm that has a modulus of convexity $\delta_E(t) > 0$. Pisier shows that one can find a modulus of power type, namely: there exist $q < \infty$ and $c > 0$ such that $\delta_E(t) \geqslant c t^q$ for $t \in [0, 2]$.

The Gurarii–James theorem asserts that when E is super-reflexive (James' case of the theorem, the space E is uniformly convex for Gurarii), there exist $q < \infty$ and $c > 0$ such that for any monotone normalized basic sequence (e_n) in E, one has

$$\Big\| \sum_n u_n e_n \Big\| \geqslant c \Big(\sum_n |u_n|^q \Big)^{1/q}$$

for all scalars (u_n). Pisier applies Gurarii–James to the successive differences of E-valued martingales, that form monotone basic sequences in the spaces $L^p(\Omega, E)$, that are super-reflexive when E is and $1 < p < \infty$. Now, to each p corresponds a $q(p)$ by Gurarii–James: if one is lucky enough that $q(p) = p$, then Pisier's proof is nicely and rather easily done with; otherwise, it's quite a work to show that one can achieve $q(p) = p$.

(36) Talking once with Pełczyński about this peculiar habit of a few colleagues, in a street of Paris near *le Panthéon*, he then told me: *"I'm an atheist"*.

(37) A subset Λ of $\mathbb{Z}$ is a $\Lambda(p)$-set when for $0 < r < p$, the L^r-norms on the trigonometric polynomials with spectrum in Λ are equivalent to the L^p-norm. By the log-convexity of

the function $s \mapsto \|f\|_{L^s}$, it is enough that this would happen for *one* value $r < p$: for some r such that $0 < r < p$, there exists C such that for all coefficients (c_n), one has

$$\bigl\|\sum_{n\in\Lambda} c_n\, e_n\bigr\|_{L^p(\mathbb{T})} \leqslant C\,\bigl\|\sum_{n\in\Lambda} c_n\, e_n\bigr\|_{L^r(\mathbb{T})}, \quad e_n(t) = \mathrm{e}^{\mathrm{i}nt}, \ \ n\in\mathbb{Z}, \ \ t\in\mathbb{R}.$$

When $p > 2$ and with the choice $r = 2$, the $\Lambda(p)$-property can thus be written as

$$\bigl\|\sum_{n\in\Lambda} c_n\, e_n\bigr\|_{L^p(\mathbb{T})} \leqslant C\,\bigl(\sum_{n\in\Lambda} |c_n|^2\bigr)^{1/2}.$$

(38) Let E be a Banach space, let $(r_n)_{n\geqslant 1}$ be the sequence of Rademacher functions in $L^2 = L^2([0,1])$ and let P_R denote the orthogonal projection from L^2 onto the closed subspace spanned by the Rademacher functions. Let $L^2(E) = L^2([0,1], E)$ be the space of square integrable vector valued functions from $[0,1]$ to E. Denote by $\mathrm{Rad}(E)$ the subspace of $L^2(E)$ generated by the functions $u \mapsto r_n(u)x$, $u\in[0,1]$, $x\in E$. Then E is said to be *K-convex* if the projection $P_R \otimes \mathrm{Id}$ is bounded from $L^2(E)$ onto $\mathrm{Rad}(E)$. Pisier proved that this happens precisely when E has some non trivial type $p > 1$.

For K-convex spaces, cotype (31) and type are dual properties: if E has cotype q, then the dual space E^* has type p with $1/p + 1/q = 1$. Going from type p for E^* to cotype q for E is always true, but in the other direction, the space ℓ^1 has cotype 2 while its dual ℓ^∞ has no non-trivial type, as ℓ^∞ contains every separable Banach space as a subspace, isometrically, in particular copies of ℓ^1: if ℓ^∞ had some type $p > 1$, then ℓ^1 would also. But for the unit vector basis (e_i) of ℓ^1, the equality

$$\bigl\|\sum_{i=1}^{n} \varepsilon_i\, e_i\bigr\|_{\ell^1} = n, \qquad n \geqslant 1,$$

for all $\varepsilon_i = \pm 1$ forbids any non trivial type for ℓ^1, and hence for its "superspace" ℓ^∞.

(39) Pełczyński had a joke on the set of results obtained in Paris in the '70s: he kindly spoke about "the French Revolution in Banach Spaces". I had met him during a visit I made in Warsaw in 1974, and found a man that always tried to have a positive influence on young researchers, showing much interest, asking good questions and pushing people beyond what they had got so far. As a result, we had by the end of my visit a little paper together, devoted to (p,q)-summing operators. That the paper was not very important is in my view an evidence of what I mean to say here.

(40) For example, observe that $\ln 2 < 1 - 1/2 + 1/3 = 5/6$, hence $b > 1/\ln 2 > 6/5$; this implies that $7b(b-1) > b > 1 > \ln 2$.

References

Too many articles and books would be relevant here, especially for the "historical" part. I shall omit all those that I have never looked at (although I probably should have).

[Bour] J. Bourgain, *Bounded orthogonal systems and the $\Lambda(p)$-set problem*, Acta Math. **162** (1989), 227–245.

[Che$_1$] S. Chevet, *Mesures de Radon sur $\mathbb{R}^n$ et mesures cylindriques*, Ann. Fac. Sci. Univ. Clermont **43** (1970), 91–158.

[Che$_2$] S. Chevet, *Processus Gaussiens et volumes mixtes*, Z. für Wahrscheinlichkeitstheorie **36** (1976), 47–65.

[Dud_1] R. M. Dudley, *Sample functions of the Gaussian process*, Ann. Probab. **1** (1973), 66–103.

[Dud_2] R. M. Dudley, *V. N. Sudakov's Work on Expected Suprema of Gaussian Processes*, High Dimensional Probability VII, Progress in Probability 71 (2016), 37–43.

[Enf_1] P. Enflo, *Banach spaces which can be given an equivalent uniformly convex norm*, Israel J. Math. **13** (1972), 281–288.

[Enf_2] P. Enflo, *A counterexample to the approximation problem in Banach spaces*, Acta Math. **130** (1973), 309–317.

[Fer_1] X. Fernique, *Intégrabilité des vecteurs gaussiens*, C. R. Acad. Sc. Paris **270** (1970), 1698–1699.

[Fer_2] X. Fernique, *Minorations des fonctions aléatoires gaussiennes*, Ann. Inst. Fourier, Grenoble **24**, 2 (1974), 61–66.

[Fer_3] X. Fernique, *Régularité des trajectoires des fonctions aléatoires gaussiennes*, École d'Été de Probabilités de Saint-Flour IV – 1974. Lecture Notes in Math. 480, Springer, Berlin, 1975, p. 1–96.

[FLM] T. Figiel, J. Lindenstrauss & V. Milman, *The dimension of almost spherical sections of convex bodies*, Acta Math. **139** (1977), 53–94.

[Gord] Y. Gordon, *Some inequalities for Gaussian processes and applications*, Israel J. of Math. **50** (1985), 265–289.

[Gro_1] A. Grothendieck, *Sur les applications linéaires faiblement compactes d'espaces du type $C(K)$*, Canad. J. Math. **5** (1953), 129–173.

[Gro_2] A. Grothendieck, *Résumé de la théorie métrique des produits tensoriels topologiques*, Bol. Soc. Mat. São Paulo **8** (1953, published 1956), 1–79.

[Gro_3] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Mem. Amer. Math. Soc. **16** (1955).

[Haa_1] U. Haagerup, *The best constants in the Khintchine inequality*, Studia Math. **70** (1981), 231–283.

[Haa_2] U. Haagerup, *The Grothendieck inequality for bilinear forms on C^*-algebras*, Adv. in Math. **56** (1985), 93–116.

[Hoff] J. Hoffmann-Jørgensen, *Sums of independent Banach space valued random variables*, Studia Math. **52** (1974), 159–186.

[Jam] R. C. James, *Some self-dual properties of normed linear spaces*, Symposium on Infinite-Dimensional Topology (Louisiana State Univ., Baton Rouge, La., 1967), pp. 159–175, Ann. of Math. Stud. **69**, Princeton University Press, Princeton, NJ, 1972.

[JoLi] W. Johnson & J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in modern analysis and probability (New Haven, Conn., 1982), 189–206, Contemp. Math., **26**, Amer. Math. Soc., Providence, RI, 1984.

[JoSc] W. Johnson & G. Schechtman, *Embedding ℓ_p^m into ℓ_1^n*, Acta Math. **149** (1982), 71–85.

[Kah_1] J.-P. Kahane, *Sur les sommes vectorielles $\sum \pm u_n$*, C. R. Acad. Sci. Paris **259** (1964), 2577–2580.

[Kah_2] J.-P. Kahane, Some random series of functions, Heath and Co. Lexington Mass. 1968; Cambridge Univ. Press, Cambridge, 1985, 1993.

[Kash] B. Kashin, *The widths of certain finite-dimensional sets and classes of smooth functions* (Russian) Izv. Akad. Nauk SSSR (Ser. Mat). **41** (1977) 334–351. English translation: Math. USSR-Izv., **11** (1977), 317–333.

[Kriv] J.-L. Krivine, *Sous-espaces de dimension finie des espaces de Banach réticulés*, Ann. of Math. **104** (1976), 1–29.

[Kw_1] S. Kwapień, *Isomorphic characterizations of inner product spaces by orthogonal series with vector valued coefficients*, Studia Math. **44** (1972), 583–595.

[Kw_2] S. Kwapień, *A theorem on the Rademacher series with vector valued coefficients*, Probability in Banach spaces (Proc. First Internat. Conf., Oberwolfach, 1975), p. 157–158. Lecture Notes in Math., Vol. 526, Springer-Verlag, Berlin, 1976.

[LiQu] D. Li & H. Queffélec, *Introduction à l'étude des espaces de Banach*, Société Mathématique de France, Paris 2004. Augmented English translation: *Introduction to Banach Spaces: Analysis and Probability*, 2 vol., Cambridge Univ. Press, Cambridge 2018.

[LiPe] J. Lindenstrauss & A. Pelczynski, *Absolutely summing operators in L_p-spaces and their applications*, Studia Math. **29** (1968), 275–326.

[LiTz] J. Lindenstrauss & L. Tzafriri, Classical Banach spaces. II. Function spaces, *Ergeb. Math. Grenzgeb.*, **97**, Springer-Verlag, Berlin-New York, 1979. x+243 pp.

[Mil] V. Milman, *A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies*, (Russian) Funkcional. Anal. i Priložen. **5** (1971), 28–37.

[OdS] E. Odell & T. Schlumprecht, *The distortion problem*, Acta Math. **173** (1994), 259–281.

[Pie_1] A. Pietsch, *Absolut p-summierende Abbildungen in normierten Räumen*, Studia Math. **28** (1966/67), 333–353.

[Pie_2] A. Pietsch, Operator ideals. *North-Holland Math. Library*, **20**, North-Holland Publishing Co., Amsterdam-New York, 1980. 451 pp.

[Pis_1] G. Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math. **20** (1975), 326–350.

[Pis_2] G. Pisier, *Ensembles de Sidon et processus gaussiens*, Comptes Rendus Acad. Sci. Paris **286** (1978), 671–674.

[Pis_3] G. Pisier, *Grothendieck's theorem for noncommutative C*-algebras, with an appendix on Grothendieck's constants*, J. Functional Analysis **29** (1978), 397–415.

[Pis_4] G. Pisier, *Holomorphic semi-groups and the geometry of Banach spaces*, Annals of Math. **115** (1982), 375–392.

[Pis_5] G. Pisier, *On the dimension of the ℓ_p^n-subspaces of Banach spaces, for $1 \leqslant p < 2$*, Trans. Amer. Math. Soc. **276** (1983), 201–211.

[Ros] H. Rosenthal, *On subspaces of L^p*, Ann. of Math. **97** (1973), 344–373.

[Rud] W. Rudin, *Trigonometric series with gaps*, J. Math. Mech. **9** (1960), 203–227.

[Sch] L. Schwartz & al., *Séminaire Laurent Schwartz 1969–1970: Applications radonifiantes*, École Polytechnique, Centre de Mathématiques, Paris, 1970. iv+339 pp.

[Slep] D. Slepian, *The one-sided barrier problem for Gaussian noise*, Bell. Syst. Techn. J. **41** (1962), 463–501.

[Sud$_1$] V.N. Sudakov, *Gaussian random processes and measures of solid angles in Hilbert spaces*, Soviet Math. Doklady 12, 412–415 (1971). Russian original Doklady Akad. Nauk SSSR 197(1), 43–45.

[Sud$_2$] V. N. Sudakov, *Geometric problems in the Theory of Infinite-Dimensional Probability Distributions*, (in Russian). Trudy Mat. Inst. Steklov 141, 191 pp. (1976); English "cover to cover translation", Proc. Steklov Inst. Math. 1979 (2), American Mathematical Society.

[Szan] A. Szankowski, *$B(\mathcal{H})$ does not have the approximation property*, Acta Math. **147** (1981), 89–108.

[Szar] S. Szarek, *On the best constants in the Khinchin inequality*, Studia Math. **58** (1976), 197–208.

[Tom] N. Tomczak-Jaegermann, *The moduli of smoothness and convexity and the Rademacher averages of trace classes S^p $(1 \leqslant p < \infty)$*, Studia Math. **50** (1974), 163–182.

[Tsir] B. Tsirelson, *It is impossible to imbed ℓ_p or c_0 into an arbitrary Banach space*, (Russian) Funkcional. Anal. i Priložen. **8** (1974), 57–60. English translation: *Not every Banach space contains an imbedding of ℓ_p or c_0*, Functional Anal. Appl. **8** (1974), 138–141.