Fifty years ago,
a theorem by Xavier Fernique
opens a new way in Gaussian processes

(bm, 19-12-2024)

Contents

Links in the text are *red* or *green*, 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 Michel Talagrand's further developments on the same theme, and in the next to last section, we shall also recall a few facts from the '70s, somewhat connected to Gaussian processes. I will try my best to keep the exposition as self-contained as possible. 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). The classical result says: 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] 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 proved 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. The second section presents a proof of the Fernique theorem: we shall not exactly follow the original proof, whose probabilistic tool is essentially the Sudakov entropy bound, but use a more recent lemma that is another relatively easy consequence of the Slepian–Sudakov comparison result. The third section deals with Gaussian processes that are no longer assumed to be stationary; the results there are due to Talagrand, who first proved the existence of a *majorizing measure* [Tal_1] and introduced later the notion of generic chaining [Tal_2], see Theorem 2.

Section 4 is loosely connected to the preceding Section 3. I shall recall a few facts and results from the remote epoch of the '70s: Laurent Schwartz and the radonifying maps, Albrecht Pietsch and the factorization of p-summing operators. In that spirit, I shall apply the results of the third section to giving at 4.4 a distinctive factorization for the linear mappings that transform the Gaussian cylindrical measure into a true Radon measure.

In the last section, a few closed-form formulas are given at 5.1 for the values of the expectations $\mathrm{E}\max(g_1, g_2, \ldots, g_n)$ of the maximum of a small number $n \leqslant 5$ of independent $\mathrm{N}(0,1)$ Gaussian random variables g_i. They appear in the proof of Lemma 2 in Section 1.2. A sketch of proof for the Slepian–Sudakov result is also given.

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

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

because $\varepsilon \mapsto \mathcal{N}(T,\varepsilon)$ is non-increasing. Since $\Delta < r_0$ and $\lim r_i = 0$ we obtain

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

The latter series $\Sigma_1(T)$ is of the sort that will appear several times later. Let us now write simply $\mathcal{N}(r_{i+1})$ instead 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

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

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 be > 0 and observe that

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

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

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

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

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

For the second inequality (8), we have first $\mathrm{E}\, G_1^* = \mathrm{E}\,|g_1| = \sqrt{2/\pi} < 1 < 2\sqrt{\ln 2}$, that 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*

$$\frac{\mathrm{E}\Big(\max_{1\leqslant i\leqslant N} g_i\Big)}{\sqrt{2\ln N}} \underset{N}{\longrightarrow} 1.$$

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 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 (see section 5.1) are the facts that

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

Our last effort at 5.1 has been to establish that

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

Taking this([3]) 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

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

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

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

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 $\mathrm{N}(0,1)$ Gaussian 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 $u > 0$ rather large and $\varepsilon \in (0,1)$ small. When the integer N is large enough, say $N \geqslant N_0(u,\varepsilon) \geqslant 3$, we certainly have

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

Then

$$\sqrt{2\,\ln N} > s := \sqrt{2\ln N - \ln(2\ln N) - u} \geqslant \sqrt{(2-\varepsilon)\,\ln N} > \sqrt{\ln 3} > 1.$$

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

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

The above lower bound is valid when $N \geqslant N_0(u,\varepsilon)$. Now $\alpha = \alpha(u)$ can be made as large as we wish, and for N large enough, the inequalities

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

prove our claim. □

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 role 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.

A sketch of proof of this result is given at 5.2 in the Appendix. 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

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

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$_3$] ([4]).

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 for 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 be used to give another proof of a celebrated lemma due to William Johnson and Joram Lindenstrauss [JoLi], that was first proved using concentration of measure on the Euclidean sphere in high dimension ([5]).

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 $\varepsilon^2/2$. When $i \neq j$, we have

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

hence by Proposition 1 and Inequality (12),

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

This gives a bound on n, and thus

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

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 ([6]) *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 $\mathcal{N}_*$) in place 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 $s \in T$ and $r > 0$ 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 seen above, we may and shall assume that K is convex. This approach of Sudakov uses a suitably normalized *mixed volume* (7).

Let us consider first the simplest case where K is a segment $[0,x] \neq \{0\}$ in $\mathbb{R}^n$. Let B_n be the Euclidean unit ball in the space $\mathbb{R}^n$ and $\|x\|$ the Euclidean norm of x, let $|A|_n$ denote the n-dimensional volume of sets $A \subset \mathbb{R}^n$ and let $v_n = |B_n|_n$. We have that 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 $(B_n + uK) \setminus B_n$ consists of intervals of length $u\|x\|$ situated on the lines ℓ that are parallel to $[0,x]$ and intersect B_n. Let $x^\perp$ denote the hyperplane orthogonal to the segment $[0,x]$. The intersection points with $x^\perp$ of those lines ℓ fill the $(n-1)$-dimensional unit ball of that hyperplane $x^\perp$, 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 defining a notion that extends to embeddings in an infinite dimensional Hilbert space.

When K is a k-dimensional compact convex 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 arbitrary compact convex subsets 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

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

We shall sketch a proof of (17) in the particular case of a convex set K that is the convex hull of a finite set A of points $(a_j)_{j=1}^m$ in $\mathbb{R}^n$, with $m \geqslant 1$ and $n > 1$, namely

$$K = \mathrm{conv}(A) = \Bigl\{ \sum_{j=1}^m \lambda_j\, a_j : \sum_{j=1}^m \lambda_j = 1 \text{ and } \lambda_i \geqslant 0,\ i = 1, 2, \ldots, m \Bigr\}.$$

As the two sides of (17) are 1-homogeneous in K, we may assume that $K \subset B_n$, in other words, we shall have $\|a_i\| \leqslant 1$ for $i = 1, 2, \ldots, m$ in what follows.

Let $\mathrm{d}\sigma_{n-1}$ denote the $(n-1)$-dimensional Lebesgue measure on the unit sphere S^{n-1} of $\mathbb{R}^n$. The Lebesgue measure $\mathrm{d}x$ on $\mathbb{R}^n$ can be expressed as

$$r^{n-1}\, \mathrm{d}\sigma_{n-1}(\theta)\, \mathrm{d}r$$

where points $x \in \mathbb{R}^n$ are represented as $x = r\theta$, with $r \geqslant 0$ and $\theta \in S^{n-1}$. We suppose that $0 < u < 1$, hence $u a_j \in B_n$ for j in $\{1, \ldots, m\}$. The convex set $B_n + uK$ contains $0 = (-u a_j) + u(a_j)$, and for every direction $\theta \in S^{n-1}$ the points $x = r\theta$ with $r \geqslant 0$ that are in $B_n + uK$ form a segment from the point 0 to a point $x_0 = \rho_u(\theta)\,\theta$, where $\rho_u(\theta) \geqslant 0$. We have therefore

$$|B_n + uK|_n = \int_{S^{n-1}} \Bigl(\int_0^{\rho_u(\theta)} r^{n-1}\, \mathrm{d}r \Bigr) \mathrm{d}\sigma_{n-1}(\theta) = \int_{S^{n-1}} \frac{\rho_u(\theta)^n}{n}\, \mathrm{d}\sigma_{n-1}(\theta). \tag{18}$$

Due to the 2-smoothness of B_n, we have when $u > 0$ is small that

$$\rho_u(\theta) = 1 + u \max_{1 \leqslant j \leqslant m} (\theta \cdot a_j) + O_K(u^2). \tag{19}$$

Indeed, on one hand, the point $x_0 = \rho_u(\theta)\theta \in B_n + uK$ can be written as $x_0 = b + uy$ with $b \in B_n$ and $y \in K$, hence

$$\rho_u(\theta) = \theta \cdot x_0 \leqslant 1 + u(\theta \cdot y) \leqslant 1 + u \max_j (\theta \cdot a_j).$$

On the other hand, consider one of the points a_i, for some $i \in \{1, 2, \ldots, m\}$. If a_i belongs to the line $L_\theta = \mathbb{R}\theta$, then $x_1 = \theta + u a_i \in B_n + uK$ is also in L_θ and

$$\rho_u(\theta) = \max\{\theta \cdot x : x \in (B_n + uK) \cap L_\theta\} \geqslant \theta \cdot x_1 = 1 + u(\theta \cdot a_i).$$

Otherwise, let $\eta \in S^{n-1}$ be such that the couple (η, θ) is an orthonormal basis for the plane generated by θ and a_i, and choose η such that $\eta \cdot a_i < 0$. Consider

$$x_2 = (\sqrt{t}\,\eta + \sqrt{1-t}\,\theta) + u a_i \in B_n + uK, \quad \text{with} \quad 0 < u < 1$$

and with $t > 0$ satisfying $\sqrt{t} + u(\eta \cdot a_i) = 0 = \eta \cdot x_2$. Then x_2 is on the line L_θ and

$$\sqrt{t} = -u(\eta \cdot a_i) \leqslant u\|a_i\| \leqslant u, \quad t \leqslant u^2 < 1,$$

because we assumed that $K \subset B_n$. Since $x_2 \in (B_n + u\,K) \cap L_\theta$, we get

$$\rho_u(\theta) \geqslant \theta \cdot x_2 = \sqrt{1-t} + u\,(\theta \cdot a_i) \geqslant \sqrt{1-u^2} + u\,(\theta \cdot a_i) \geqslant \bigl(1 + u\,(\theta \cdot a_i)\bigr) - u^2,$$

and our claim (19) follows([8]). $\square$

In (18), the coefficient of the first degree term in u is therefore equal to

$$\int_{S^{n-1}} \max_j (\theta \cdot a_j)\,\mathrm{d}\sigma_{n-1}(\theta) = v_{n-1}\,h_1(K).$$

The standard Gaussian probability measure γ_n on $\mathbb{R}^n$ is expressed by

$$(2\,\pi)^{-n/2}\, r^{n-1}\, \mathrm{e}^{-r^2/2}\ \mathrm{d}\sigma_{n-1}(\theta)\,\mathrm{d}r,$$

hence, equating to 1 the integral on $\mathbb{R}^n$ of the Gaussian density, we get

$$s_{n-1} \int_0^\infty r^{n-1}\,\mathrm{e}^{-r^2/2}\ \mathrm{d}r = (2\,\pi)^{n/2}, \tag{20}$$

where s_{n-1} is the $(n{-}1)$-dimensional measure of the unit sphere S^{n-1}. The integral with respect to γ_n of the 1-homogeneous function $x \mapsto \max_j (x \cdot a_j)$ is

$$\begin{aligned}
\int_{\mathbb{R}^n} \max_j (x \cdot a_j)\,\mathrm{d}\gamma_n(x) &= (2\,\pi)^{-n/2} \int \max_j (r\theta \cdot a_j)\, r^{n-1}\,\mathrm{e}^{-r^2/2}\ \mathrm{d}\sigma_{n-1}(\theta)\,\mathrm{d}r \\
&= (2\,\pi)^{-n/2} \Bigl(\int_{S^{n-1}} \max_j (\theta \cdot a_j)\,\mathrm{d}\sigma_{n-1}(\theta) \Bigr) \,.\, \int_0^\infty r^n\,\mathrm{e}^{-r^2/2}\ \mathrm{d}r \\
&= (2\,\pi)^{-n/2}\, v_{n-1}\, h_1(K) \,.\, (n{-}1) \int_0^\infty r^{n-2}\,\mathrm{e}^{-r^2/2}\ \mathrm{d}r \\
&= (2\,\pi)^{-n/2}\, h_1(K) \,.\, s_{n-2} \int_0^\infty r^{n-2}\,\mathrm{e}^{-r^2/2}\ \mathrm{d}r = (2\,\pi)^{-1/2}\, h_1(K),
\end{aligned}$$

where we used $(n{-}1)\,v_{n-1} = s_{n-2}$ and (20) for $\mathbb{R}^{n-1}$ instead of $\mathbb{R}^n$. Furthermore, the Gaussian process associated to K consists of the Gaussian variables $X_t : x \mapsto x \cdot t$ defined for $t \in K$ on the probability space $(\mathbb{R}^n, \gamma_n)$, and

$$\sup_{t \in K} X_t(x) = \max_{1 \leqslant j \leqslant m} (x \cdot a_j), \quad \mathrm{E}\Bigl(\sup_{t \in K} X_t \Bigr) = \int_{\mathbb{R}^n} \max_{1 \leqslant j \leqslant m} (x \cdot a_j)\,\mathrm{d}\gamma_n(x) = \frac{h_1(K)}{\sqrt{2\,\pi}} \cdot \qquad \square$$

Simone Chevet [Che$_2$] has obtained results that relate higher moments of the supremum to other normalized mixed volumes of K. For the second moment, let us express it again for a set $A = \{a_1, \ldots, a_m\} \subset \mathbb{R}^n$ and for $K = \mathrm{conv}(A)$. Let $X_1, X_2, \ldots, X_m$ be the Gaussian variables on $(\mathbb{R}^n, \gamma_n)$ associated as above to the points $a_1, \ldots, a_m$, let $\gamma(i,j) = \mathrm{E}\, X_i X_j$ and let $\sigma(\omega)$ denote the smallest([9]) index in $\{1, 2, \ldots, m\}$ such that

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

If $|B_n + u\,K|_n = \sum_{k=0}^n c_k\,u^k$, then $h_2(K) = c_2 / v_{n-2}$. Chevet proves at (3.6.1) that

$$\frac{h_2(K)}{\pi} = \mathrm{E}\bigl(X_\sigma^2 - \gamma(\sigma, \sigma) \bigr) \quad \text{or else}$$

$$\frac{h_2(K)}{\pi} = \mathrm{E}\Bigl(\sup_{t \in K} X_t \Bigr)^2 - \mathrm{E}_\omega\, \mathrm{E}_{\omega'} (X_{\sigma(\omega)}(\omega'))^2 = \mathrm{E}\Bigl(\sup_{t \in K} X_t \Bigr)^2 - \sum_{i=1}^m \mathrm{P}(\sigma = i)\, \mathrm{E}\, X_i^2.$$

The quantity $\mathrm{E}\bigl(X_\sigma^2 - \gamma(\sigma, \sigma) \bigr)$ appears already([10]) in an article by Fernique [Fer$_4$].

1.4. *First applications of comparison*

We begin with a simple lemma.

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

$$\mathrm{E}\Bigl(\sup_{i,j}\bigl(X_i + Y_{i,j}\bigr)\Bigr) \geqslant \mathrm{E}\Bigl(\sup_{i\in I} X_i\Bigr) + \inf_{i\in I} \mathrm{E}\Bigl(\sup_{j\in J_i} Y_{i,j}\Bigr),$$

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

Proof. Let

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

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

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

then

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

and

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

Integrating in u concludes the proof. $\square$

The next Lemma does most of the serious job in what follows ([11]).

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 $A_1, A_2, \ldots, A_N$ 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

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

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

$$Y_s - Y_t = X_s^{(i)} - X_t^{(i)}, \quad \text{hence} \quad \mathrm{E}(Y_s - Y_t)^2 = \mathrm{E}(X_s - X_t)^2$$

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, we know that $d(s, a_i) \leqslant \rho$ and $d(t, a_j) \leqslant \rho$, 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

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

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 $S = \{s_1, s_2, \ldots, s_N\} \subset B(t_0, r)$. 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. $\square$

1.5. *Promenades*

I would like to introduce a notion of *promenade*: roughly speaking, it is a Markov chain without probability. It will place us on a longish route to Fernique's result, but fairly smooth I believe. We shall call $\mathcal{X}$ the set of states, or more specifically, the set of *places* to be successively visited step by step. For each place $x \in \mathcal{X}$ of the promenade, a set $P(x) \subset \mathcal{X}$ of places is given: this is the set of places that can be reached from the place x in one single step, we shall call them the *next-places after x*. It is possible that $P(x)$ be empty: then x is a final state where the promenade ends, an *end-point*. On the other hand we assume that $x \notin P(x)$: not moving is not considered a step.

We let a *finite path* of the promenade be a finite sequence of places

$$\xi = (x_0,\ x_1,\ \ldots,\ x_k) \ \text{ in } \ \mathcal{X},$$

where x_{i+1} is a next-place after x_i for every i such that $0 \leqslant i < k$. The *length* of the path is the number of places in ξ, namely $k + 1$. We will admit as path the trivial case when $k = 0$, a path $\xi = (x)$ of just one place, of length 1. If $y \in P(x)$, then (x, y) is a path of length 2 from x to y, but y could also be attained from x by other paths of length > 2, as in (x, z, y) for example, if $z \in P(x)$ and $y \in P(z)$.

Typically, our promenades will arise from covering with balls a compact subset T of a Hilbert space: a place would be somehow a ball B in T of some radius $r > 0$, and the *next-places* after that place will be balls B' of a 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 next-places with radius $r/3$ would increase the covering number to some $3^n r^{-n}$ balls; one can then think that each place will have about 3^n next-places after it, a constant factor depending upon the given dimension n. But we actually work in infinite dimension, so the number of next-places may increase dramatically from one place to the next in the promenade.

Given a promenade on $\mathcal{X}$, we introduce the notion of a *3-control function* N for the promenade, a function with positive values on the set $\mathcal{X}$ that will control the evolution of the cardinality $|P(x)|$. We shall make below (condition $\mathbf{c_2}$) the debatable choice to say that $N(x)$ controls the number of all places that can be attained when taking a single step of which x is one of possible destinations, rather than the number of places that could be attained *starting from* a given place: if $x \in P(x_*)$, then $|P(x_*)| \leqslant N(x)$. Perhaps paradoxically, we shall not try to keep this function N small, but on the contrary, guarantee a huge growth.

The first condition below defines a *3-growing function* N for the promenade; we shall then say that $N(x)$ is the *growth value* at the place $x \in \mathcal{X}$. Thus N is a 3-growing function for a promenade on the set $\mathcal{X}$ if it satisfies:

$\mathbf{c_1}$ — For every place $x \in \mathcal{X}$, the value of N at any next-place after x satisfies ([12])

$$y \in P(x) \Rightarrow N(y) \geqslant N(x)^3.$$

We say that N is a *3-control function* for the promenade on the set $\mathcal{X}$ if it is a 3-growing function and if in addition:

$\mathbf{c_2}$ — For every $x \in \mathcal{X}$ and $y \in P(x)$, the value of $N(y)$ is an upper bound for the number of places that can be reached from x in one step,

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

By the previous condition, we see that for a promenade on $\mathcal{X}$ with a 3-control function, the set of next-places after each place x in $\mathcal{X}$ is *finite*. The growth condition in $\mathbf{c_1}$ will be used to exert a backward hold over the number of "previous places" in the promenade: if $y \in P(x)$ is a next-place after x, we have that $N(x) \leqslant N(y)^{1/3}$. The next lemma and its corollary will not be used in the proof of Fernique's theorem. They will appear when checking afterwards that the entropy condition obtained from that proof, or the similar condition obtained later in the case of non-invariant processes, indeed suffice to control the expectation of the supremum of the Gaussian process under study.

Consider a promenade on $\mathcal{X}$. Let Ξ be a set of finite paths in $\mathcal{X}$; we assume that all the paths in Ξ have a fixed common first place $\overline{x} \in \mathcal{X}$, and that no path in Ξ can be extended and stay in Ξ; in other words, we suppose that for every path

$$\xi = (x_0, x_1, \ldots, x_j, \ldots, x_k) \quad \text{with} \quad 0 \leqslant j < k \quad \text{and} \quad x_0 = \overline{x}$$

in $\mathcal{X}$, one has that

$$\xi \notin \Xi \quad \text{or} \quad (x_0, x_1, \ldots, x_j) \notin \Xi.$$

Let us then say that the set Ξ has the *no-extension property*. We could just say that every element ξ of Ξ is *maximal in* Ξ. If

$$(x_0, x_1, \ldots, x_j) \in \Xi, \quad \text{let} \quad \lambda(\xi) = x_j$$

denote the *last place* in the finite path ξ.

Lemma 5. *Suppose that a promenade on $\mathcal{X}$ admits a 3-control function N. Let Ξ be a set of finite paths with a common first place $\overline{x}$. If the set Ξ satisfies the* no-extension property*, then*

$$\sum_{\xi \in \Xi} N(\lambda(\xi))^{-3/2} \leqslant N(\overline{x})^{-3/2}.$$

Proof. Clearly, it is enough to prove the lemma for every *finite* subset of Ξ, we thus assume from now on that Ξ itself is finite. Consider the set $\mathcal{T}_0$ of all finite paths

$$\theta = (x_0, y_1, \ldots, y_j) \quad \text{with} \quad x_0 = \overline{x}$$

of the promenade in $\mathcal{X}$. We say that the path θ has an *extension in* Ξ when there is a path $\xi = (\overline{x}, x_1, \ldots, x_k)$ in Ξ such that $k \geqslant j$ and $y_i = x_i$ for $1 \leqslant i \leqslant j$, and we let then

$$\rho_\Xi(\theta, \xi) = k - j + 1$$

be the number of places from the last place $y_j = x_j$ of θ to the last place of ξ, including x_j: if $\theta \in \Xi$, we consider that θ extends itself and we have $\rho_\Xi(\theta, \theta) = 1$. Let Ξ_θ denote the set of $\xi \in \Xi$ that extend θ in the previous sense. Letting $\theta_0 = (\overline{x})$ one sees that $\Xi_{\theta_0} = \Xi$. For each $\theta \in \mathcal{T}_0$, we introduce

$$S(\theta) = \sum_{\xi \in \Xi_\theta} N(\lambda(\xi))^{-3/2}.$$

We let $h(\theta)$ be the maximal "extension-length" $\rho_\Xi(\theta, \xi)$ of the paths ξ in Ξ_θ. It could be that no path in Ξ extends θ, in which case we have $\Xi_\theta = \emptyset$, $S(\theta) = 0$ and $h(\theta) = 0$.

If $h(\theta) = 1$, then θ is in Ξ and is the only path in Ξ_θ, thus $S(\theta) = N(\lambda(\theta))^{-3/2}$. We are going to show by induction on $h(\theta) \geqslant 1$ that

$$S(\theta) \leqslant N(\lambda(\theta))^{-3/2}, \qquad \theta \in \mathcal{T}_0.$$

Suppose that $\theta \in \mathcal{T}_0$ and $h(\theta) > 1$. Then $\theta \notin \Xi$ by the no-extension property, hence every path $\xi \in \Xi_\theta$ goes through some next-place y after the last place $x = \lambda(\theta)$ of θ. For $y \in P(x)$, let θ_y by the path obtained by extending θ by one step to the place y, so that $\lambda(\theta_y) = y$. Then, every $\xi \in \Xi_\theta$ belongs to a unique set Ξ_{θ_y}, $y \in P(x)$, and we have

$$S(\theta) = \sum_{y \in P(x)} S(\theta_y).$$

Let $y_0 \in P(x)$ be such that $N(y_0) = \min_{y \in P(x)} N(y)$. By the definition of the function h, all the places $y \in P(x)$ satisfy $h(\theta_y) < h(\theta)$, thus, by the induction hypothesis, we have for every y in $P(x)$ that

$$S(\theta_y) \leqslant N(\lambda(\theta_y))^{-3/2} = N(y)^{-3/2}, \quad \text{and} \quad N(y)^{-3/2} \leqslant N(y_0)^{-3/2}.$$

By the condition $\mathbf{c_2}$ we know that $|P(x)| \leqslant N(y_0)$, and $N(x)^3 \leqslant N(y_0)$ by $\mathbf{c_1}$, thus

$$S(\theta) = \sum_{y \in P(x)} S(\theta_y) \leqslant |P(x)| \, N(y_0)^{-3/2} \leqslant N(y_0)^{-1/2} \leqslant N(x)^{-3/2} = N(\lambda(\theta))^{-3/2}.$$

This proves our claim. Finally, we have that $\Xi_{\theta_0} = \Xi$ when $\theta_0 = (\overline{x})$ and

$$N(\overline{x})^{-3/2} = N(\lambda(\theta_0))^{-3/2} \geqslant S(\theta_0) = \sum_{\xi \in \Xi} N(\lambda(\xi))^{-3/2}. \qquad \square$$

We shall deal mainly with promenades that start from a fixed initial place $\hat{\mathrm{x}} \in \mathcal{X}$, that we call *origin* or *starting point* of the promenade. For a path $(x_i)_{0 \leqslant i \leqslant k}$ starting at the origin, hence with $x_0 = \hat{\mathrm{x}}$, we say a *path from origin*. Given an origin $\hat{\mathrm{x}}$, there is a natural subset of $\mathcal{X}$ that consists of places x that can be attained on a path from origin,

$$\xi = (x_0,\, x_1,\, \ldots,\, x_k), \quad x_0 = \hat{\mathrm{x}}, \quad x_k = x.$$

We shall denote by $\widehat{\mathcal{X}}$ the set of those *attainable places* x. Anything that will matter later on will occur in this subset $\widehat{\mathcal{X}}$ of $\mathcal{X}$. It will be however convenient ([13]) to keep the whole set $\mathcal{X}$ in our setting.

When considering *maximal paths* of a promenade with starting point $\hat{\mathrm{x}}$ we will always restrict ourselves to paths starting from that origin $\hat{\mathrm{x}}$, in other words, maximal paths in the attainable subset $\widehat{\mathcal{X}}$. These maximal paths are thus sequences $\mathbf{x} = (x_i)_{0 \leqslant i < L}$ of places in $\widehat{\mathcal{X}}$, where $L = L(\mathbf{x})$ is finite or $L = +\infty$, and:

– the place x_0 in the path $\mathbf{x}$ is the starting point $\hat{\mathrm{x}}$ of the promenade,

– the place x_{i+1} is a next-place after x_i whenever $i \in \mathbb{N}$ and $i + 1 < L$,

– when L is finite, the place x_{L-1} has no next-place, it is an *end-point* of the promenade.

We shall prefer avoiding end-points: the maximal paths of our main promenades will all be infinite, this will help us to keep a somewhat unified treatment. A maximal path $\mathbf{x}$ in $\widehat{\mathcal{X}}$ will thus look like

$$\mathbf{x} = (x_0,\, x_1,\, x_2,\, \ldots,\, x_i,\, \ldots), \quad x_{i+1} \in P(x_i) \subset \mathcal{X} \ \text{ for } \ i \geqslant 0,$$

with $x_0 = \hat{\mathrm{x}}$ being the given origin of the promenade.

A *rooted tree* provides an example of a promenade with starting point: the *places* are the nodes of the tree, the starting point of the associated promenade is the root of the tree and the next-places after a place|node are the children of that node. Conversely, given a promenade on $\mathcal{X}$ with starting point $\hat{\mathrm{x}}$, one can check that the set of paths

$$\xi = (x_0, x_1, \ldots, x_k), \quad x_i \in \mathcal{X}, \ \ x_{i+1} \in P(x_i) \ \text{ for } \ 0 \leqslant i < k, \ \ k \geqslant 0, \ \ x_0 = \hat{\mathrm{x}}$$

has a tree structure: the root is the path $(\hat{\mathrm{x}})$ and the children of ξ are the paths

$$(\hat{\mathrm{x}}, x_1, \ldots, x_k, x_{k+1}), \quad x_{k+1} \in P(x_k).$$

If a promenade has a starting point $\hat{\mathrm{x}}$ and a 3-control function N, we shall *normalize* the function N by imposing the following additional condition:

$\mathbf{c_0}$ — We have $N(\hat{\mathrm{x}}) = 1$ for the starting point $\hat{\mathrm{x}}$ of the promenade, and $N(y) \geqslant 2$ for every next-place $y \in P(\hat{\mathrm{x}})$ after $\hat{\mathrm{x}}$.

Assuming this normalization condition satisfied, let $x \neq \hat{\mathrm{x}}$ be a place in $\widehat{\mathcal{X}}$, attained by a path $\hat{\mathrm{x}}, x_1, \ldots, x_k = x$ from origin. Since $x \neq \hat{\mathrm{x}}$, we have that $k > 0$. Because $N(x_1) \geqslant 2$ by condition $\mathbf{c_0}$ and $N(x_{j+1}) \geqslant N(x_j)^3$ by $\mathbf{c_1}$, we see that $N(x_{j+1}) \geqslant N(x_j)^3 > N(x_j)$

for $j \geqslant 1$. It follows that $N(x) \geqslant 2$ for any *attainable* place $x \neq \hat{\mathrm{x}}$; also, the values of N along a path from origin are increasing, so that such a path can never come back to a place where it has been before ([14]).

Let $(x_j)_{0 \leqslant j \leqslant k}$ be a finite path from origin in $\mathcal{X}$. Since $N(x_j) \geqslant 2$ for any place x_j with $j \geqslant 1$, and $N(x_{j+1}) \geqslant N(x_j)^3$, we see easily by induction that ([15])

for every j such that $0 \leqslant j \leqslant k$, we have that $N(x_j) \geqslant 2^j$.

When $j = 2$, one has actually $N(x_j) \geqslant 2^3 = 8 > 2^2$, so that

for every j such that $1 \leqslant j \leqslant k$, we have that $N(x_j) > 2^j$.

Suppose now that all the maximal paths in $\widehat{\mathcal{X}}$ are infinite, that is to say: for every attainable place $x \in \widehat{\mathcal{X}}$, the set $P(x)$ of next-places is never empty. It follows that for any maximal path $(x_i)_{i \geqslant 0}$ in $\widehat{\mathcal{X}}$, we have

$$N(x_i) \geqslant 2^i \ \text{ for } i \geqslant 0, \quad \sum_{j=1}^{\infty} N(x_j)^{-1} < 1 \ \text{ thus } \ \sum_{j=0}^{\infty} N(x_j)^{-1} < 2. \tag{21}$$

Suppose that a promenade on $\mathcal{X}$ admits a starting point and a normalized 3-control function N; let M be a number $\geqslant 1$ and consider the subset of $\mathcal{X}$ defined by

$$\mathcal{X}_M = \{x \in \mathcal{X} : N(x) \leqslant M\}. \tag{22}$$

Because $N(x) < N(y)$ when $y \in P(x)$, all "previous places" in $\mathcal{X}$ of a place $y \in \mathcal{X}_M$ belong to $\mathcal{X}_M$. The starting point $\hat{\mathrm{x}}$ of the promenade in $\mathcal{X}$ belongs to the set $\mathcal{X}_M$, because $M \geqslant 1$ and $N(\hat{\mathrm{x}}) = 1$ according to $\mathbf{c_0}$. It follows from the first inequality in (21) that all paths in $\mathcal{X}_M$ that start at $\hat{\mathrm{x}}$ are finite, and since the set $P(x)$ is finite for every place $x \in \mathcal{X}$ by $\mathbf{c_2}$, we know that the set of points in $\mathcal{X}_M$ that are attainable from $\hat{\mathrm{x}}$ is finite ([16]).

Let us introduce the set Ξ of paths that are *maximal* among paths from the origin $\hat{\mathrm{x}}$ that remain in $\mathcal{X}_M$ at all stages. Maximality implies the no-extension property for Ξ, therefore the previous lemma gives

$$\sum_{\xi \in \Xi} N(\lambda(\xi))^{-3/2} \leqslant N(\hat{\mathrm{x}})^{-3/2} = 1.$$

By the definition of Ξ we know that $\lambda(\xi) \in \mathcal{X}_M$ for every $\xi \in \Xi$, hence $N(\lambda(\xi)) \leqslant M$ and

$$|\Xi|\, M^{-3/2} \leqslant \sum_{\xi \in \Xi} N(\lambda(\xi))^{-3/2} \leqslant 1.$$

This ends the proof for the next corollary.

Corollary 2. *Let a promenade with origin $\hat{\mathrm{x}}$ on the set $\mathcal{X}$ admit a normalized 3-control function N, and suppose that the subset $\mathcal{X}_M$ is defined by* (22)*. The number of paths of the promenade that are* maximal *among the paths that start from $\hat{\mathrm{x}}$ and stay in $\mathcal{X}_M$ is bounded above by $M^{3/2}$.*

If $\mathcal{X}$ is a rooted tree with root $\hat{\mathrm{x}}$, the associated promenade on $\mathcal{X}$, with starting point $\hat{\mathrm{x}}$, is defined by letting the set $P(x)$ of next-places after a node $x \in \mathcal{X}$ be the set of children of x in the tree. Giving a node x in this rooted tree is equivalent to giving the unique path ξ_x starting at the root $\hat{\mathrm{x}}$ and having x as last place. Let us say that a subset D of the tree $\mathcal{X}$ is *disconnected* if whenever $x \neq y$ are two elements of D, there is no path from x to y, nor from y to x; in other words, the nodes x and y are not comparable for the tree order. For example, the set L of leaves of a finite tree is disconnected. Clearly, the set Ξ_D of paths ξ_x associated to the elements x of a disconnected set D has the no-extension property, and we therefore get immediately from Lemma 5 the next corollary.

Corollary 3. *Let $\mathcal{X}$ be a rooted tree and D a* disconnected *subset of $\mathcal{X}$. If the associated promenade on $\mathcal{X}$ admits a normalized 3-control function N, then*

$$\sum_{x \in D} N(x)^{-3/2} \leqslant 1.$$

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

$$h.X_g = X_{g+h}, \quad h \in G.$$

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

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \quad \text{and} \quad (X_{g.t_1}, X_{g.t_2}, \ldots, X_{g.t_k}) \tag{23}$$

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

$$\varphi(t_0, r) = \mathrm{E}\Bigl(\sup_{s \in B(t_0, r)} X_s\Bigr) = \mathrm{E}\Bigl(\sup_{s \in B(t_1, r)} X_s\Bigr) = \varphi(t_1, r). \tag{24}$$

In this invariant setting [17], 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

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

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 (24) and (25). 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

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

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 4. *Let $(X_t)_{t\in T}$ be a centered Gaussian process that satisfies the invariance condition (24). 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*

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

The idea of associating a tree to the process can be traced in Chap. 7 of Fernique's work [Fer$_3$]; we shall not need the more restrictive tree structure in what follows, we shall merely consider a promenade. The maximal paths of this promenade will all be infinite, at the cost of doing things in a way [18] 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 also very inexact description of the procedure: we will construct a set $\mathcal{X}$ whose places correspond to smaller and smaller balls $B(t, r)$ in T. The "richness" of $\mathcal{X}$ will reflect the entropy of T, as the next-places after the place corresponding to $B(t, r)$ will arise from covering $B(t, r)$ by balls with a smaller radius. We shall apply Corollary 4 repeatedly when passing from a place to its next-places. Given a ball $B(t, r)$ associated to a place 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 4 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 be associated to the next-places after the place corresponding to $B(t, r)$. Next, we would find for these next-places 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 consecutive differences $\varphi(r_i) - \varphi(r_{i+1})$ will finally 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. *Defining a promenade*

We shall introduce a set of places $\mathcal{X}$, a promenade on $\mathcal{X}$ and a function N that will be a 3-control function for this promenade. A place $x \in \mathcal{X}$ will have the form (19)

$$x = (t, r, n) \ \text{ with } \ t \in T, \ \ r > 0, \ \ n \geqslant 1.$$

We define three functions on $\mathcal{X}$ that are the coordinates of the places $x = (t, r, n)$, namely

$$\theta(x) = t, \ \ \rho(x) = r, \ \ N(x) = n.$$

We say that $\theta(x)$ is the *position* in T of the place x. The *radius function* ρ will have an extremely simple behaviour with respect to the promenade: when moving from a place x to a next-place $y \in P(x)$, the radius will simply be divided by 3,

$$\rho(y) = \rho(x)/3 \ \text{ whenever } \ y \in P(x).$$

The function N will turn out to be a 3-control function for the promenade to be defined. As it was explained briefly above, the expectation of the supremum of the values X_s of the process, for s in the ball $B(t, r) \subset T$, will be examined and used. Recall that under our invariance assumptions, we have

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

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 starting point of the promenade is $\hat{\mathrm{x}} = (t_0, r_0, 1)$, we have $\rho(\hat{\mathrm{x}}) = r_0$, the radius associated to $\hat{\mathrm{x}}$, and the initial value of N is $N(\hat{\mathrm{x}}) = 1$. The description of the promenade and of the function N goes as follows:

Suppose that a place $x_* = (t_*, r_*, n_*) \in \mathcal{X}$ is given. 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 role in what follows: the case where $|S| > n_*^3$ corresponds to when $B(t_*, r_*)$

is "fairly large" and we shall then call the step from x_* to its next-places a "rich step", the case $|S| \leqslant n_*^3$ being of course the "poor" one. In the rich step case, we let

$$n = |S|, \quad \text{and in the poor case} \quad n = n_*^3, \quad \text{so that} \quad n = \max(n_*^3, |S|).$$

For each $s \in S$, we introduce a *next-place x after x_** that has the form

$$x = (s, r, n), \quad \text{therefore} \quad P(x_*) = \{(s, r, n) : s \in S\}.$$

We thus have the same value $N(x) = n$ for all next-places x after x_*. There is at least one next-place after the place x_* —because the set S is not empty—, and the number of next-places after x_* is at most n, since $|P(x_*)| = |S| \leqslant n$.

A few comments are in order.

— The way in which the radius function ρ varies along our promenade induces a hierarchy on the places in $\mathcal{X}$ that are *attainable* from the origin $\hat{\mathrm{x}}$: a place $x = (t, r, n)$ that appears at the ith stage $x_i = x$ of a path

$$\xi = (x_0, x_1, \ldots, x_i) \quad \text{starting at} \quad x_0 = \hat{\mathrm{x}}$$

has necessarily $r = 3^{-i} r_0$, and thus, a place $x = (t, r, n)$ for which $r = 3^{-i} r_0$ can only appear at the ith stage of a path starting at $\hat{\mathrm{x}}$. I found nicer to deal above with the whole family of places (t, r, n), but only the places that are attainable from $\hat{\mathrm{x}}$ will play a role later.

— In the poor case, the value $n = n_*^3$ is merely an *upper bound* for the number of next-places after the place x_* ([20]).

— The function N has a special property that was not asked in general for growing functions or control functions: the value $N(x)$ is the same for all the next-places x after a given place x_*. We say then that the function is *previsible*,

$$x, x' \in P(x_*) \ \Rightarrow \ N(x) = N(x').$$

Let us check that the function N is a normalized previsible 3-control function for our promenade on the set $\mathcal{X}$: for the starting place $\hat{\mathrm{x}} = (t_0, r_0, 1)$, we have $N(\hat{\mathrm{x}}) = 1$; also, we have decided that $\Delta < r_0 < 4\Delta/3$, with $\Delta = \mathrm{diam}(T)$. It yields first that $B(t_0, r_0) = T$; next, it implies that $B(t, r_0/3) \neq T$ for any point $t \in T$, because

$$\mathrm{diam}(B(t, r_0/3)) \leqslant 2(r_0/3) < 8\Delta/9 < \Delta.$$

When we defined the set $P(\hat{\mathrm{x}})$, we have introduced a set S_1 that is a $(r_0/3)$-p-net for the ball $B(t_0, r_0) = T$: the balls of radius $r_0/3$ centered at the points of S_1 cover T, we have thus $|S_1| \geqslant 2$. Then, we see that $|S_1| > 1 = N(\hat{\mathrm{x}})^3$, this is a rich step, it follows from our construction that $N(y) = |S_1| \geqslant 2$ when $y \in P(\hat{\mathrm{x}})$ and the condition $\mathbf{c_0}$ for a control function is satisfied. For every place x_* we made sure that $N(x) \geqslant N(x_*)^3$ when $x \in P(x_*)$, thus N satisfies $\mathbf{c_1}$, and we have $|P(x_*)| \leqslant N(x)$, condition $\mathbf{c_2}$. In addition, the function N is *previsible*, since it was defined in a way that it has the same value $N(x)$ for all the next-places x after any given place x_*.

2.1.2. *Gaussian estimates*

Let us fix a place $x_* = (t_*, r_*, n_*)$ in $\mathcal{X}$, from where we shall take a single step of the promenade. Passing from x_* to all the next-places $x = (t, r, n)$ after x_*, we get from Corollary 4 a relation between the expectations of the suprema of the Gaussian process $(X_t)_{t \in T}$ over a ball centered at t_* on one hand, and over balls centered at the various positions t on the other. The radii of these balls will be fixed multiples of r_*, the (common) multiple for the next-places being smaller than that for the original place x_*. The set of points t appearing in the next-places x after x_* form the set S that was chosen to be an r-p-net for $B(t_*, r_*)$: this set S is thus r-separated and

$$S \subset B(t_*, r_*) \subset \bigcup_{s \in S} B(s, r).$$

Recall that $r = r_*/3$ and that the control value $N(x) = n$, when $x \in P(x_*)$, only depends upon the fixed place x_* (the function N is previsible). There are two cases to consider:

— suppose first that we are in the "rich step" case: then we know that $n = |S|$. We let $2\delta = r$, so that the set $S \subset B(t_*, r_*)$ is 2δ-separated. Notice that $\delta/2 = r/4 = r_*/12$; applying Corollary 4 we get

$$\varphi(r_* + r_*/12) \geqslant (r_*/24)\sqrt{\ln n} + \varphi(r_*/12),$$

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

$$\varphi(2r_*) \geqslant (r_*/24)\sqrt{\ln n} + \varphi(r_*/12); \tag{26}$$

— otherwise, in the "poor case" for the step from x_*, we know that $n = n_*^3$. We essentially do nothing in this case: we just find convenient to introduce $r_{**} = 3r_*$ —it could be the radius r_{**} of a place x_{**} after which x_* is a next-place— and we observe that

$$r_*\sqrt{\ln n} = \sqrt{3}\, r_*\sqrt{\ln n_*} = \frac{\sqrt{3}}{3}\, r_{**}\sqrt{\ln n_*} \leqslant \frac{2}{3} r_{**}\sqrt{\ln n_*} \tag{27}$$

because $\sqrt{3} < 2$, hence

$$(r_*/24)\sqrt{\ln n} \leqslant \frac{2}{3}(r_{**}/24)\sqrt{\ln n_*}.$$

Obviously we have that

$$\varphi(r_*/12) \leqslant \varphi(2r_*),$$

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

$$\varphi(2r_*) + \frac{2}{3}\kappa r_{**}\sqrt{\ln n_*} \geqslant \kappa r_*\sqrt{\ln n} + \varphi(r_*/12). \tag{28}$$

We now observe that this inequality is clearly valid also in the "rich" case, simply because we have then (26) and $\kappa r_{**}\sqrt{\ln n_*} \geqslant 0$.

Let $\mathbf{x} = (x_j)_{j \geqslant 0}$ denote a maximal path of the promenade on $\mathcal{X}$: the place x_0 in the path $\mathbf{x}$ is the starting point $\hat{\mathrm{x}}$ in $\mathcal{X}$, and x_{i+1} is a next-place after x_i for every $i \geqslant 0$. Let us write

$$x_i = (t_i, r_i, n_i) \text{ for } i \geqslant 0, \quad \text{with} \quad N(x_i) = n_i.$$

It follows from the behaviour of the radius function ρ along the promenade that

$$r_i = \rho(x_i) = 3^{-i}\rho(\hat{\mathrm{x}}) = 3^{-i} r_0.$$

It will be convenient to set $r_{-1} = 3r_0$. If we want to apply (28) to the step from the place $x_* = (t_*, r_*, n_*) = x_i$ to the place $x = (t, r, n) = x_{i+1}$, where $i \geqslant 0$, we have to set

$$r_* = r_i, \quad n_* = n_i, \quad n = n_{i+1} \quad \text{and} \quad r_{**} = 3r_* = r_{i-1}.$$

Then (28) becomes

$$\varphi(2r_i) + \frac{2}{3}\kappa r_{i-1}\sqrt{\ln n_i} \geqslant \kappa r_i\sqrt{\ln n_{i+1}} + \varphi(r_i/12). \tag{29}$$

We may consider that the elements of the places in the path $\mathbf{x}$ are functions of that path and write $x_i = x_i(\mathbf{x}) = (t_i(\mathbf{x}), r_i, N_i(\mathbf{x}))$ for the place x_i at stage i in the path $\mathbf{x}$, with the control value $N_i(\mathbf{x}) = N(x_i(\mathbf{x})) = n_i$. Now, for any maximal path $\mathbf{x}$ and every integer $i \geqslant 0$, Equation (29) gives

$$\varphi(2r_i) + \frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i(\mathbf{x})} \geqslant \kappa r_i \sqrt{\ln N_{i+1}(\mathbf{x})} + \varphi(r_i/12). \tag{30}$$

Summing (30) from $i = 0$ to an arbitrary $k \geqslant 0$, using $r_{-1}\sqrt{\ln N_0(\mathbf{x})} = 0$ —because we have that $N_0(\mathbf{x}) = n_0 = 1$— 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 maximal path $\mathbf{x}$. Observe that

$$\varphi(r_i/12) \geqslant \varphi(2r_{i+3}) \tag{31}$$

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 maximal path $\mathbf{x}$ of the promenade on $\mathcal{X}$ we have that

$$\sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{9}{\kappa}\,\varphi(2r_0) = \frac{9}{\kappa}\, E^* = 216\ \mathrm{E}\Big(\sup_{t \in T} X_t\Big). \tag{32}$$

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 maximal paths $\mathbf{x}$, that is thus 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 set $\mathcal{X}$, with its promenade and control function N. For every place x in $\mathcal{X}$ with $x = (t, r, n) = (\theta(x), \rho(x), N(x))$, 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 $P(x)$ denotes the set of next-places after x.

$\mathbf{a_0}$ — The starting point of the promenade is $\hat{x} = (t_0, r_0, 1)$, where $t_0 = \theta(\hat{x})$ is an arbitrary point in T. The radius $r_0 = \rho(\hat{x})$ is chosen such that $\Delta < r_0 < 4\Delta/3$, where Δ is the diameter of T; it implies that $\beta(\hat{x}) = T$.

$\mathbf{a_1}$ — For every place $x \in \mathcal{X}$, the ball $\beta(x) = B(\theta(x), \rho(x))$ associated to x is covered by the balls associated to the next-places $y = (\theta(y), \rho(y), N(y))$ after x,

$$\beta(x) \subset \bigcup_{y \in P(x)} \beta(y), \ \text{ and } \ \theta(y) \in B(\theta(x), \rho(x)), \ \ \rho(y) = \rho(x)/3 \ \text{ for}^{(21)} \text{ any } \ y \in P(x).$$

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

$$N(\hat{x}) = 1, \quad y \in P(\hat{x}) \Rightarrow N(y) \geqslant 2;$$
$$y \in P(x) \ \Rightarrow \ \big(N(y) \geqslant N(x)^3 \ \text{ and } \ |P(x)| \leqslant N(y)\big)$$

and

$$y, y' \in P(x) \;\Rightarrow\; N(y) = N(y').$$

We did not yet make use of the invariance condition (25). When we have defined the next-places after a place $x_* = (t_*, r_*, n_*)$, we have set $r = r_*/3$ and we have then introduced a r-p-net of n points for the ball $B(t_*, r_*)$, where $n = n(x_*)$ could depend upon the place x_*, without using the fact that according to the condition (25), 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 n of points for every other ball $B(t, r_*)$ of that same radius r_*. Doing this from the start $\hat{x}$ on, we may replace the functions of maximal paths $\mathbf{x} \mapsto N_i(\mathbf{x})$ by *constant values* N_i, for each $i \geqslant 0$, and write the conclusion of the promenade estimate as

$$\sigma_1 = \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} < +\infty. \tag{33}$$

2.2. *The entropy-like condition is sufficient*

Let us consider a promenade with a starting point $\hat{x}$, on a set $\mathcal{X}$ of places that have the form $x = (t, r, n)$ as in the previous sections, with t in the index set T of an invariant centered Gaussian process $(X_t)_{t \in T}$, and with $r > 0$, $n \geqslant 1$. As we have done before, we set $r = \rho(x)$ and $n = N(x)$, and we assume that the promenade on $\mathcal{X}$, the functions N and ρ satisfy the three properties $\mathbf{a_0}$, $\mathbf{a_1}$, $\mathbf{a_2}$. In addition, we assume here that given any maximal path $\mathbf{x} = (x_i)_{i \geqslant 0}$ in $\mathcal{X}$, the values $N_i = N(x_i)$ do not depend on that particular path $\mathbf{x}$, but only on the stage i reached in the promenade. We shall explain that the condition $\sigma_1 < \infty$ obtained at (33) allows one, conversely, to bound the expectation of the supremum of the process $(X_t)_{t \in T}$.

We start from the initial place $\hat{x} = (t_0, r_0, 1)$, 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 of the promenade, that will form a maximal path $\mathbf{x}(\tau)$. We thus begin with setting $x_0(\tau) = \hat{x}$, we let $t_0(\tau) = t_0$ and we have that $\tau \in T = B(t_0(\tau), r_0)$. Next, we know that the point τ is contained in at least one of the balls $B(t_1, r_1)$ associated to the positions $t_1 = \theta(x_1)$ of the next-places x_1 after $\hat{x}$, that cover $B(t_0, r_0)$ (this is condition $\mathbf{a_1}$); we let $x_1(\tau) = (t_1(\tau), r_1, n_1)$ be a next-place after $x_0(\tau) = \hat{x}$ for which we have that $\tau \in B(t_1(\tau), r_1)$. Then, using $\mathbf{a_1}$ repeatedly, we go on choosing successive places $x_{i+1}(\tau) = (t_{i+1}(\tau), r_{i+1}, n_{i+1})$ such that $x_{i+1}(\tau)$ is a next-place after $x_i(\tau)$ 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 [22] 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 move in T from $t_i(\tau)$ to $t_{i+1}(\tau)$ has size $< r_i$, we shall use it below.

Here, the control function N takes the same value N_i on all the places at the ith stage of the promenade $\mathcal{X}$. If we fix $k \geqslant 0$ and consider in $\mathcal{X}$ the subset

$$\mathcal{X}_k = \{x \in \mathcal{X} : N(x) \leqslant N_k\},$$

we know by Corollary 2 that there are at most $N_k^{3/2}$ maximal paths from $\hat{x}$ for the promenade restricted to $\mathcal{X}_k$, and clearly, the last places of those maximal paths are precisely the places at the kth stage of the promenade on $\mathcal{X}$. Hence, when τ varies in T

and when the integer $i \geqslant 0$ is fixed, the global number M_{i+1} of paths from $\hat{\mathrm{x}}$ to all the places $x_{i+1}(\tau)$ mentioned above admits([23]) the bound

$$M_{i+1} \leqslant N_{i+1}^{3/2} < N_{i+1}^2 \tag{34}$$

(recall that $N_{i+1} > 1$ because $i + 1 > 0$). In particular, when τ varies, the number of all possible steps from $x_i(\tau)$ to $x_{i+1}(\tau)$ is bounded by $M_{i+1} < N_{i+1}^2$.

Because $\mathrm{Var}(X_{t_{i+1}(\tau)} - X_{t_i(\tau)}) = d(t_{i+1}(\tau), t_i(\tau))^2 \leqslant r_i^2$ 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 the proof 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 (34) that there are $M_i \leqslant N_i^2$ paths from the start $\hat{\mathrm{x}}$ to the places at the ith stage of the promenade; in particular, the set T_i of points $t_i \in T$ appearing in the places (t_i, r_i, n_i) of that ith stage 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 (32) 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

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

If we give up invariance for the process $(X_t)_{t\in T}$, we lose Equality (24). Then, for each point $s \in T$ and $r > 0$, we can do nothing but consider

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

Without invariance and applying Corollary 1, it only remains from (26), when $i = 0$ for example, that

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

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 ingredient in the construction of the promenade. We have to make divisions of T into "zones" where the function φ will be controlled in a suitable way. Now, the collection of next-places after a place x_* 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 (24) the rapidly growing numbers $n = N(x)$ to this end, locating values of φ in small intervals of size $\mathrm{O}(n^{-1})$.

The basic homogeneous zones from which other homogeneous zones will be built are defined as follows: fix a radius $r > 0$ and an integer $m \geqslant 1$; for every integer α such that $0 \leqslant \alpha < m$, consider the (possibly empty) subset of T defined by

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

where

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

Notice that

$$T = \bigcup_{0\leqslant\alpha<m} W_\alpha(r,m), \tag{35}$$

since the segments $Z(\alpha,m)$ cover $[0,E^*]$ and $0 \leqslant \varphi \leqslant E^*$. We shall thus divide the successive sets obtained in the invariant case —there, they were the balls $B(t,r)$— into "homogeneous zones".

The new set $\mathcal{X}$ of places for the promenade will now consist of the 4-tuples

$$x = (V, t, r, n) \ \text{ where } \ t \in T, \ \ r > 0, \ \ n \geqslant 1 \ \text{ as before,}$$

and where

$$V \subset B(t,r) \ \text{ is non-empty.} \tag{36}$$

We say that V is the *region* associated to x. It will be useful to define the function R that indicates the region $V = R(x)$, and to set again

$$\theta(x) = t, \ \ \rho(x) = r, \ \ N(x) = n.$$

We could then (mostly uselessly) write $x = (R(x), \theta(x), \rho(x), N(x))$. In the construction of next-places, we shall make sure that V be a “homogeneous zone” where a certain variation of φ will be controlled.

The new control function will be defined in two phases: starting from a place x_*, we first find a bound n on the entropy of $V_* = R(x_*)$, and then a bound in n^2 for the number of places x where one can go in one step from x_*, so that the actual control value will be n^2. The function N will be called the *growing function* for the promenade, it will satisfy the conditions $\mathbf{c_0}$ and $\mathbf{c_1}$, and the function N^2 will be the *3-control function*, still satisfying $\mathbf{c_0}$ and $\mathbf{c_1}$, clearly, and satisfying $\mathbf{c_2}$ in addition.

3.1. *A new promenade*

The construction of the new promenade on $\mathcal{X}$ 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 a radius 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 starting place of our promenade is $\hat{\mathrm{x}} = (V_0, t_0, r_0, 1)$ with $V_0 = T$. So far the description of the start is exactly as before —save for the addition of $V_0 = T$ to the triple $(t_0, r_0, 1)$—. It is still convenient to let $r_{-1} = 3r_0$. Let us describe the modified construction step.

Suppose that a place $x_* = (V_*, t_*, r_*, n_*)$ is considered in the new set $\mathcal{X}$. 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 phases, 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

$$V_* \subset \bigcup_{s \in S} B(s, r). \tag{37}$$

We let $n = \max(n_*^3, |S|)$, this will be the growth value for all the next-places after x_*. Again, the *rich case* is when $n = |S| > n_*^3$, and otherwise we have the *poor case* $n = n_*^3$.

Here comes the difference: in the invariant case, the next-places after x_* corresponded directly to the points s in S and thus to the balls $B(s, r)$; now, for each $s \in S$, we make a further splitting of $B(s, r)$, arising from the splitting of T into sets $W_\alpha(r, n)$, where r and n are defined above: the first component V of the next-places after x_* will be contained in one of the intersections $B(s, r) \cap W_\alpha(r, n)$. Precisely, the next-places after $x_* = (V_*, t_*, r_*, n_*)$ will be the places $x = (V, s, r, n)$, where we decided already that

$$r = r_*/3, \quad n = \max(n_*^3, |S|), \quad \text{and where in addition}$$
$$V = V_* \cap B(s, r) \cap W_\alpha(r, n) \neq \emptyset, \ \ s \in S, \ \ 0 \leqslant \alpha < n.$$

For every next-place $x = (V, s, r, n)$ after x_*, we let $N(x) = n$ —a value that depends only upon x_*— and we have by construction that

$$V \subset V_*, \quad s \in V_*, \quad V \subset B(s, r). \tag{38}$$

The number of next-places after x_* is less than or equal to $|S|n \leqslant n^2$, and we shall thus define $N^2(x) = n^2$, it will be the control value for all those next-places $x \in P(x_*)$. Intersecting the “coverings” in (37) and (35), we see that

$$V_* \subset \bigcup \bigl\{ B(s, r) \cap W_\alpha(r, n) : s \in S, \ 0 \leqslant \alpha < n \bigr\}, \tag{39}$$

which implies that the regions V corresponding to all next-places $x = (V, s, r, n)$ after x_* cover V_*. By construction, when v runs in the set V, the values of $\varphi(v, r/12)$ stay in one of the segments $Z(\alpha, n)$ of length E^*/n, where $0 \leqslant \alpha < n$.

For each place $x = (V, t, r, n)$ in $\mathcal{X}$ let

$$z(x) = \inf \{\varphi(v, r/12) : v \in V\} \geqslant 0. \tag{40}$$

From what we have just said, we know when $x = (V, t, r, n)$ is a next-place after some place x_* that

$$v \in V \;\; \Rightarrow \;\; z(x) \leqslant \varphi(v, r/12) \leqslant z(x) + E^*/n,$$

and for the origin $\hat{\mathrm{x}}$, we also have that

$$v \in V_0 \;\; \Rightarrow \;\; z(\hat{\mathrm{x}}) \leqslant \varphi(v, r_0/12) \leqslant z(\hat{\mathrm{x}}) + E^*/n_0,$$

because $n_0 = 1$ and $0 \leqslant \varphi \leqslant E^*$.

We check that the function N is a normalized 3-growing function for the promenade on the set $\mathcal{X}$: for the start $\hat{\mathrm{x}} = (V_0, t_0, r_0, 1)$, we have $V_0 = T$, $N(\hat{\mathrm{x}}) = 1$ and we have assumed that $\Delta < r_0 < 4\Delta/3$. 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 satisfies that $N(x) \geqslant N(x_*)^3$ for all next-places x after any place x_*, thus N satisfies the condition $\mathbf{c_1}$, and it 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 $|P(x_*)| \leqslant N^2(x)$ when $x \in P(x_*)$, the condition $\mathbf{c_2}$ for N^2 to be a control function is satisfied. Finally, the function N takes on the same value $N(x)$ for all next-places x after a place x_*, it is thus *previsible*.

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

3.1.1. *Estimates*

Suppose that x_* is a given place in the new set $\mathcal{X}$, with $x_* = (V_*, t_*, r_*, n_*)$ and growth value $N(x_*) = n_*$. Recall that a r-p-net S for V_* with $r = r_*/3$ has been introduced, and that $n = \max(n_*^3, |S|)$ is the growth value $N(x)$ for all the next-places x after x_*. There are as before two possibilities:

— in the "rich case", we have $n = |S| > n_*^3$ and we know that S is r-separated, contained in $B(t_*, r_*)$ because $S \subset V_* \subset B(t_*, r_*)$ by (36) and (38); therefore, by Corollary 1 applied with the value $2\delta = r$ —and thus with $\delta/2 = r/4 = r_*/12$— we obtain

$$\begin{aligned}
\varphi(t_*, r_* {+} r_*/12) &= \mathrm{E}\Big(\sup_{t \in B(t_*, r_* + r_*/12)} X_t \Big) \\
&\geqslant (r_*/24)\sqrt{\ln n} + \min_{s \in S} \mathrm{E}\Big(\sup_{t \in B(s, r_*/12)} X_t \Big) \\
&\geqslant \kappa r_* \sqrt{\ln n} + \inf_{v \in V_*} \varphi(v, r_*/12) = \kappa r_* \sqrt{\ln n} + z(x_*),
\end{aligned}$$

by the definition (40) of $z(x_*)$; we just lazily write and keep in mind that

$$\varphi(t_i, 2r_*) \geqslant \kappa r_* \sqrt{\ln n} + z(x_*)\,;$$

— otherwise, we are in the poor case, thus $n = n_*^3$. On one hand, letting $r_{**} = 3r_*$, we use as before $\frac{2}{3} \kappa r_{**} \sqrt{\ln n_*} \geqslant \kappa r_* \sqrt{\ln n}$ —see Equation (27)—; on the other hand,

given $v \in V_*$, we have
$$z(x_*) \leqslant \varphi(v, r_*/12) \leqslant \varphi(t_*, r_* + r_*/12) \leqslant \varphi(t_*, 2r_*)$$
because $B(v, r_*/12) \subset B(t_*, r_*{+}r_*/12)$ since $v \in V_* \subset B(t_*, r_*)$.

In both cases, we conclude that

$$\varphi(t_*, 2r_*) + \frac{2}{3}\kappa r_{**}\sqrt{\ln n_*} \geqslant \kappa r_*\sqrt{\ln n} + z(x_*). \tag{41}$$

This replaces Inequality (28) obtained in the invariant case.

3.1.2. *Adding up*

Let us consider a maximal path $\mathbf{x} = (x_i)_{i\geqslant 0}$ from $\hat{x}$ of the promenade on the new set $\mathcal{X}$, and write $x_i = x_i(\mathbf{x}) = (V_i(\mathbf{x}), t_i(\mathbf{x}), r_i, N_i(\mathbf{x}))$ 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 path $\mathbf{x}$ that remains fixed in the following lines. Applying (41) to $x_* = x_i$, we conclude that

$$\varphi(t_i, 2r_i) + \frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i} \geqslant \kappa r_i\sqrt{\ln N_{i+1}} + z(x_i). \tag{42}$$

Adding (42) from $i = 0$ to an arbitrary $k \geqslant 0$ and reorganizing as before the root-of-log terms we get

$$\sum_{i=0}^{k} \varphi(t_i, 2r_i) \geqslant \frac{\kappa}{3}\sum_{i=0}^{k} r_i\sqrt{\ln N_{i+1}} + \sum_{i=0}^{k} z(x_i). \tag{43}$$

Inequality (31) has to be revised, we argue now as follows: let $x_i = (V_i, t_i, r_i, n_i)$ be the place of $\mathbf{x}$ at stage $i \in \mathbb{N}$; 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

$$\varphi(t_{i+3}, 2r_{i+3}) \leqslant z(x_i) + E^*/N_i\,; \tag{44}$$

indeed, we see again that $2r_{i+3} = 2r_i/27 < r_i/12$, and we also know by Equation (38) that $t_{i+3} \in V_{i+2} \subset V_i$ ([25]), 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 (44) 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 (43) 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 (21) 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

$$\sigma_1(\mathbf{x}) := \sum_{i=0}^{\infty} r_i\sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{15}{\kappa}\,E^* = 360\,\mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr) \tag{45}$$

for every maximal path $\mathbf{x}$ of the promenade on the set $\mathcal{X}$.

3.1.3. *Summing up*

For every place $x = (R(x), \theta(x), \rho(x), N(x))$ in $\mathcal{X}$, we have said that $R(x)$ is the *region associated to the place x*. Let us review the properties of the promenade on $\mathcal{X}$ and of the functions ρ and N.

$\mathbf{a_0^*}$ — (compare to $\mathbf{a_0}$) The starting point of the promenade is $\hat{\mathrm{x}} = (V_0, t_0, r_0, 1)$, where $V_0 = T$ and where $t_0 = \theta(\hat{\mathrm{x}})$ is a point in T. The radius $r_0 = \rho(\hat{\mathrm{x}})$ 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 $\mathbf{a_1}$ and (38)) For every place x in the set $\mathcal{X}$, the region $R(x)$ is the union of the regions associated to the next-places $y = (R(y), \theta(y), \rho(y), N(y))$ after x:

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

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

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

$$N(\hat{\mathrm{x}}) = 1, \quad y \in P(\hat{\mathrm{x}}) \Rightarrow N(y) \geqslant 2\,;$$
$$y, y' \in P(x) \ \Rightarrow \ N(y) = N(y')\,; \qquad y \in P(x) \Rightarrow N(y) \geqslant N(x)^3\,;$$
$$y \in P(x) \Rightarrow |P(x)| \leqslant N^2(y).$$

For every maximal path $\mathbf{x} = (x_i(\mathbf{x}))_{i\in\mathbb{N}}$ and $i \geqslant 0$, one has that $N(x_i(\mathbf{x})) \geqslant 2^i$ (see (21)).

Remark 1. *Suppose that a promenade on $\mathcal{X}$ satisfies the properties $\mathbf{a_0^*}$ and $\mathbf{a_1^*}$. For every point τ in T, we can find a maximal path $\mathbf{x} = (x_i)_{i\geqslant 0}$ of the promenade such that writing*

$$x_i = x_i(\mathbf{x}) = (V_i(\mathbf{x}), t_i(\mathbf{x}), r_i, n_i), \qquad i \geqslant 0,$$

we have that $t_i(\mathbf{x})$ tends to τ, and more precisely

$$\tau \in V_i(\mathbf{x}) \subset B(t_i(\mathbf{x}), r_i), \quad t_{i+1}(\mathbf{x}) \in B(t_i(\mathbf{x}), r_i) \ \text{ for each } \ i \geqslant 0. \tag{46}$$

The reason is the same as before. We let $x_0(\mathbf{x}) = \hat{\mathrm{x}}$ be the starting point of the promenade; then we have $\tau \in V_0 = T$ by condition $\mathbf{a_0^*}$. Assuming that $x_i = x_i(\mathbf{x})$ has been selected such that $\tau \in V_i(\mathbf{x}) = R(x_i)$, we can find a next-place x_{i+1} after x_i such that $\tau \in R(x_{i+1})$, because by $\mathbf{a_1^*}$ these regions $R(x_{i+1})$ cover $R(x_i)$ when x_{i+1} varies in $P(x_i)$. We then let $x_{i+1}(\mathbf{x}) = x_{i+1}$ and we go on with the construction of the path $\mathbf{x}$. By $\mathbf{a_1^*}$ again, we know that (46) is satisfied. □

If the metric space (T, d) is *dense it itself*, that is to say without isolated point, we may radically —though rather artificially [26]— simplify the setting presented in the conditions $\mathbf{a_0^*}$ to $\mathbf{a_2^*}$. Let $\mathcal{X}_0$ denote the set of places in $\mathcal{X}$ that can be reached by a path starting from the origin $\hat{\mathrm{x}}$ (this includes $\hat{\mathrm{x}}$ itself). The set $\mathcal{X}_0$ is countable, it is therefore possible [27] to define a one-to-one mapping $x \mapsto t_x$ from $\mathcal{X}_0$ to a countable subset T_0 of T, in such a way that $t_{\hat{\mathrm{x}}} = t_0 = \theta(\hat{\mathrm{x}})$ and

$$x = (V, t, r, n) \in \mathcal{X}_0 \quad \Rightarrow \quad d(t_x, t) < r/2 = \rho(x)/2. \tag{47}$$

We then transfer the promenade on $\mathcal{X}$ to a promenade on T_0: the starting point is t_0 and we set

$$t_y \in P(t_x) \quad \Leftrightarrow \quad y \in P(x).$$

We define $N_0(t_x) = N^2(x)$ (we won't need considering both N and N^2 anymore). The function N_0 is a previsible 3-control function for that promenade on T_0.

Proposition 2. *Suppose that $(X_t)_{t\in T}$ is a centered Gaussian process with a finite expectation E^* for its supremum, and that (T,d) is dense in itself. There exists then a promenade on T and a* previsible *3-control function N_0 for that promenade satisfying: for every $\tau \in T$, there is a maximal path $\mathbf{t} = (t_i)_{i\geqslant 0}$ of the promenade that tends to τ and such that*

$$\sum_{i\geqslant 0} d(t_{i+1}, t_i)\sqrt{\ln N_0(t_{i+1})} \leqslant C\, E^*,$$

with C a universal constant. One may ask in addition that

$$d(t_{i+1}, t_i) < 2.3^{-i} r_0, \quad i \geqslant 0,$$

where $\Delta < r_0 < 4\Delta/3$, with $\Delta > 0$ being the diameter of T.

Proof. We first consider the set $\mathcal{X}$, the promenade on $\mathcal{X}$ and the control function N^2 that led us to (45); the properties $\mathbf{a_0^*}$ to $\mathbf{a_2^*}$ are satisfied. We define the promenade on T and the 3-control function N_0 by transferring, via the mapping $x \mapsto t_x$ from $\mathcal{X}_0$ to $T_0 \subset T$, the corresponding elements that were defined on $\mathcal{X}$, letting

$$P(t_x) = \{t_y : y \in P(x)\}, \quad N_0(t_x) = N^2(x).$$

We complete the definition by setting $P(t) = \emptyset$ and $N_0(t) = 1$ when $t \in T\setminus T_0$, for example. Let $\tau \in T$; applying Remark 1, we can find a maximal path $\mathbf{x}$ in $\mathcal{X}_0$ satisfying (46), with

$$\mathbf{x} = (x_i(\mathbf{x})_{i\geqslant 0}, \quad x_i(\mathbf{x}) = (V_i, t_i(\mathbf{x}), r_i, n_i).$$

Let

$$s_0 = t_0 = t_{\hat{x}}, \quad s_i = t_{x_i(\mathbf{x})} \text{ for each } i \geqslant 1.$$

The mapping $x \mapsto t_x$ satisfies (47): we see therefore that

$$d(\tau, s_i) \leqslant d(\tau, t_i(\mathbf{x})) + d(t_i(\mathbf{x}), s_i) < r_i + r_i/2 < 2r_i,$$

and by (47) and $\mathbf{a_1^*}$ we get

$$d(s_{i+1}, s_i) \leqslant r_{i+1}/2 + d(t_{i+1}(\mathbf{x}), t_i(\mathbf{x})) + r_i/2 < 2r_i.$$

We have a maximal path $\mathbf{s} = (s_i)_{i\geqslant 0}$ of the promenade on T such that s_i tends to τ and such that $d(s_{i+1}, s_i) < 2r_i$ for $i \geqslant 0$. We know by (45) that

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

Now $d(s_{i+1}, s_i) < 2r_i$ and $N_0(s_{i+1}) = N^2_{i+1}(\mathbf{x})$, the result follows with $C \leqslant 3 \times 360$. $\square$

3.2. *Suppose we have that nice promenade*

Let $\mathcal{X}$ be a set of places $x = (V,t,r,n)$ 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 $B(t,r)$. Suppose that the promenade 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 maximal path of the promenade 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 place x_i at stage $i \geqslant 0$, we shall also write $x_i = (V_i, t_i, r_i, n_i)$ and $n_i = N(x_i)$. Assume that for some constant K_1 and for every maximal path $\mathbf{x}$ in the set $\mathcal{X}$, 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 $(X_t)_{t\in T}$.

Let us fix a stage $i \in \mathbb{N}$ in the promenade. Consider a place $x_i = (V_i, t_i, r_i, n_i)$ at the ith stage. Let us first estimate the probability of a (relatively) *large jump* when passing from x_i to a specific fixed next-place $x_{i+1} = (V_{i+1}, t_{i+1}, r_{i+1}, n_{i+1})$ after 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 at t_i. Consider $u > 0$, consider the path $\xi_i = (x_i, x_{i+1})$ of length 2 and let the size of the jump be

$$s(u, \xi_i) = \bigl(u + \sqrt{8 \ln n_{i+1}}\,\bigr)\, d(t_{i+1}, t_i),$$

where $n_{i+1} = N(x_{i+1})$ depends on the previous place x_i only, by condition $\mathbf{a_2^*}$. We shall recall it by writing $n_{i+1} = N_+[x_i]$. By the definition of the metric d on T, we have

$$\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > s(u, \xi_i)\bigr) = \mathrm{P}\bigl(|g| > u + \sqrt{8 \ln n_{i+1}}\,\bigr)$$

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

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

Note that $n_{i+1} \geqslant 2^{i+1}$ by condition $\mathbf{a_2^*}$. Letting $c_i(u) = 2^{-i-1}\,\mathrm{e}^{-u^2/2}$, we rewrite the last two lines as

$$\begin{aligned}\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > s(u, \xi_i)\bigr) &\leqslant \mathrm{e}^{-u^2/2}\, n_{i+1}^{-1}\, n_{i+1}^{-3}\\ \leqslant \mathrm{e}^{-u^2/2}\, 2^{-i-1}\, n_{i+1}^{-3} &= c_i(u) n_{i+1}^{-3} = c_i(u) N_+[x_i]^{-3}.\end{aligned} \tag{48}$$

Let Ξ_i be the set of paths of length $(i+2)$ from the start $\hat{\mathrm{x}}$ to any place x_{i+1} at the $(i+1)$ stage of the promenade. Clearly, this set has the no-extension property. The control function here being N^2, it follows from Lemma 5 that

$$\sum_{\xi\in\Xi_i} N^2(\lambda(\xi))^{-3/2} = \sum_{\xi\in\Xi_i} N(\lambda(\xi))^{-3} \leqslant 1,$$

where $\lambda(\xi)$ is the last place in ξ, that is to say, the place x_{i+1} at stage $(i+1)$ in the present case. This sum ([28]) is larger than or equal to the sum limited to the family F_i of all paths $\xi = (x_i, x_{i+1})$ from a place x_i at the ith stage to a place $x_{i+1} \in P(x_i)$ at the $(i+1)$th stage of the promenade. It follows that

$$\sum_{\xi\in F_i} N(\lambda(\xi))^{-3} = \sum_{(x_i, x_{i+1})\in F_i} N(x_{i+1})^{-3} = \sum_{(x_i, x_{i+1})\in F_i} N_+[x_i]^{-3} \leqslant 1.$$

We obtain therefore from (48) that

$$c_i(u) \sum_{(x_i, x_{i+1})\in F_i} N_+[x_i]^{-3} \leqslant c_i(u)$$

is an upper bound for the total probability that a jump of size $s(u, \xi_i)$ occur between the ith stage and the next for *some* path of length $(i+2)$ going from the start to any place at the $(i+1)$th stage. We want to apply this to different integers $i \geqslant 0$: starting from the initial place $\hat{\mathrm{x}}$ of the promenade, the first jump to consider is between $\hat{\mathrm{x}}$ and its next-places $x_1 \in P(\hat{\mathrm{x}})$, a jump with a size $s(u, \xi_0)$ and probability $\leqslant c_0(u)$. So, the probability $p(u)$ of finding for some $i \geqslant 0$ a jump of size $s(u, \xi_i)$ between the stages i and $(i+1)$ of the promenade is bounded above by

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

Therefore, except for a set $S(u) \subset \Omega$ that has probability $P(S(u)) \leqslant p(u)$, we obtain that when $\omega \notin S(u)$, *all* moves

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

along *any* maximal path $\mathbf{x}$ are less than $s(u, (x_i(\mathbf{x}), x_{i+1}(\mathbf{x})))$, so that outside $S(u)$, we have

$$\sum_{i=0}^{\infty} |X_{t_{i+1}(\mathbf{x})} - X_{t_i(\mathbf{x})}| \leqslant \sum_{i=0}^{\infty} \bigl(u + \sqrt{8 \ln N_{i+1}(\mathbf{x})}\,\bigr) \operatorname{dist}(t_{i+1}(\mathbf{x}), t_i(\mathbf{x})).$$

By $\mathbf{a}_1^*$, we know that $t_{i+1}(\mathbf{x}) \in V_i(\mathbf{x}) \subset B(t_i(\mathbf{x}), r_i)$, hence $d(t_{i+1}(\mathbf{x}), t_i(\mathbf{x})) < r_i$. Then, outside $S(u)$, we have

$$\begin{aligned}(49) \qquad \sum_{i=0}^{\infty} |X_{t_{i+1}(\mathbf{x})} - X_{t_i(\mathbf{x})}| &\leqslant \Bigl(\sum_{i=0}^{\infty} r_i\Bigr) u + \sum_{i=0}^{\infty} r_i \sqrt{8 \ln N_{i+1}(\mathbf{x})} \\ &< (3/2) r_0\, u + 3K_1 \leqslant 2\Delta u + 3K_1.\end{aligned}$$

We know by Remark 1 that every $\tau \in T$ is the limit of the sequence $(t_i(\mathbf{x}))_{i \geqslant 0}$ for a certain maximal path $\mathbf{x}$, 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 (49) 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

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

Remark 2. It is not difficult to adapt the above proof and get:

Suppose that $(X_t)_{t \in T}$ is a centered Gaussian process, and that (T, d) is dense in itself. The expectation E^* of the supremum of the process is finite if and only if there exists a constant K_0 and a promenade on T with a starting point $t_0 \in T$ and with a previsible *3-control function N, satisfying that: for every $\tau \in T$, there is a maximal path $\mathbf{t} = (t_i)_{i \geqslant 0}$ of the promenade, starting at t_0, that tends to τ and such that*

$$\sum_{i \geqslant 0} d(t_{i+1}, t_i) \sqrt{\ln N(t_{i+1})} \leqslant K_0.$$

The first direction is given by Proposition 2. When trying conversely to bound E^* using the inequality above, one begins by considering for each $i \geqslant 0$ a path $\xi_i = (t_i, t_{i+1})$ of the promenade on T, and by setting the new bound

$$s(u, \xi_i) = \bigl(u + \sqrt{5 \ln N(t_{i+1})}\bigr)\, d(t_{i+1}, t_i), \quad u > 0.$$

Observing again that $N_+[t_i] := N(t_{i+1}) \geqslant 2^{i+1}$ if t_{i+1} is a place at stage $(i{+}1)$ on a path starting from t_0, we obtain now

$$\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > s(u, \xi_i)\bigr) \leqslant \mathrm{e}^{-u^2/2}\, N_+[t_i]^{-5/2} \leqslant c_i(u)\, N_+[t_i]^{-3/2},$$

with $c_i(u)$ as before. One applies again Lemma 5. At the end, note that for every maximal path $\mathbf{t} = (t_i)_{i \geqslant 0}$ of the promenade on T, one has

$$\sqrt{\ln 2} \sum_{i \geqslant 0} d(t_{i+1}, t_i) \leqslant \sum_{i \geqslant 0} d(t_{i+1}, t_i) \sqrt{\ln N(t_{i+1})} \leqslant K_0$$

and therefore

$$\sum_{i \geqslant 0} s(u, \xi_i) \leqslant \Bigl(\frac{u}{\sqrt{\ln 2}} + \sqrt{5}\Bigr) K_0.$$

One concludes as before that E^* is finite, bounded by a universal multiple of K_0. □

3.2.1. *Only the rich really count, and other remarks*

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

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

in $\mathbf{x}$ consists only of *poor steps* from x_i to x_{i+1}, for every integer i between $j+1$ and k. By Equation (27) we have

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

thus

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

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

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

— We could have tried to relate the construction of the "new promenade" in section 3 to the successive balls introduced in the invariant case. Indeed, we can construct the promenade on $\mathcal{X}$ 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 promenade on $\mathcal{X}^*$ 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 promenade on $\mathcal{X}^*$ the value N_i^* refers to a single set V_{i-1}.

— The construction of the successive regions V_i associated to the places 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 have to make the sets V_i *disjoint* by a slight modification of the procedure. Let us come back to the construction step of the next-places after a given place $x_* = (V_*, t_*, r_*, n_*)$: first, we have set $r = r_*/3$ and we have introduced a r-p-net S for the set V_*; then, the regions V for the next-places $x = (V, t, r, n)$ after 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 $x = (V, s, r, n)$, and that the various sets V corresponding to *all* the next-places after 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 have considered 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

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

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 $\mathrm{dist}(t, S) < r$ 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 $Z(\alpha, m)$, that we define now for every $m \geqslant 1$ by

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

a collection of $(m-1)$ half-open intervals; then the new sets $Z(\alpha, m)$ are disjoint, for each fixed $m \geqslant 1$. Letting $n_* = N(x_*)$ be the growth value for the previous place x_*, we introduce again the integer $n = \max(n_*^3, |S|)$ for the growth value at each next-place x after x_*; finally, the next-places $x = (V, t, r, n)$ after x_* are constructed from the elements

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

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. (29)

Each set V_i at the ith stage is an atom of a 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* in the probabilistic sense: its value depends only upon the preceding field $\mathcal{F}_i$—. Because the sets V_i of a same stage i are disjoint, we know that each point τ in T determines a *unique* path $\mathbf{x} = \mathbf{x}(\tau)$ with the property that $\tau \in V_i(\mathbf{x}) \subset B(t_i(\mathbf{x}), r_i)$ for every $i \geqslant 0$.

This construction of fields on T could go along with defining a probability measure on the space $(T, \mathcal{F})$, where $\mathcal{F}$ is the σ-field generated by the finite fields: this is what will be done in the next section, but without relying on the disjointification procedure presented above.

3.3. *Majorizing measure*

We will assume here that T is a compact space, and we are going to define an interesting probability measure on T: 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 [Fer$_3$]). The results in the next Section 3.4 imply that this measure exists, however, the form given there in Equation (52) 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 we have

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

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 $\tau \in T$.

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

$$x_i = (V_i, t_i, r_i, n_i)$$

with $r_i = 3^{-i} r_0$. 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 we assume that $E^* < \infty$), and for every path $\mathbf{x}$ in $\mathcal{X}$, where we have $x_i(\mathbf{x}) = (V_i(\mathbf{x}), t_i(\mathbf{x}), r_i, N_i(\mathbf{x}))$ 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 maximal paths $\mathbf{x}$ in $\mathcal{X}$ is a compact space for its natural tree-topology ([30]), and it projects on T via the continuous ([30]) mapping π that associates a limit point in T to each maximal path $\mathbf{x}$, by letting

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

Knowing all what we know now, the most sensible thing to do seems to consider the "natural" probability measure ν on $\mathbf{X}$ associated to the promenade ([31]) on $\mathcal{X}$ —or to the filtration, if we did perform the disjointification—. Let $i \geqslant 0$ and let

$$\xi_i = (x_0, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_i)$$

be a path of the promenade, with $x_0 = \hat{\mathrm{x}}$ the starting point; we associate to ξ_i the closed-open subset ξ_i^* of $\mathbf{X}$ consisting of maximal paths that begin with ξ_i, namely

$$\xi_i^* = \{\mathbf{x} \in \mathbf{X} : x_j(\mathbf{x}) = \overline{x}_j,\ 1 \leqslant j \leqslant i,\ \ x_0(\mathbf{x}) = \hat{\mathrm{x}}\}.$$

Letting $\xi_0 = (\hat{\mathrm{x}})$ be the trivial path, we have that $\xi_0^* = \mathbf{X}$ and since we are looking for a probability measure on $\mathbf{X}$ we set first

$$\nu(\xi_0^*) = \nu(\mathbf{X}) = 1.$$

If $\xi_{i+1} = (\hat{\mathrm{x}}, \overline{x_1}, \overline{x_2}, \ldots, \overline{x_i}, \overline{x_{i+1}})$ is an extension of the path ξ_i by one more step going from the last place $\overline{x_i}$ of ξ_i to a new place $\overline{x_{i+1}} \in P(\overline{x_i})$, and knowing that $|P(\overline{x_i})|$ is the number of the places $\overline{x_{i+1}}$ that can be reached from $\overline{x_i}$ (a number $\geqslant 1$), we let

$$\nu(\xi_{i+1}^*) = \frac{\nu(\xi_i^*)}{|P(\overline{x_i})|}.$$

Clearly, the set $\xi_i^* \subset \mathbf{X}$ is the disjoint union of the sets ξ_{i+1}^* corresponding to the extensions of ξ_i by a next-place $\overline{x_{i+1}}$ after $\overline{x_i}$. It follows that $\nu(\xi_i^*)$ is equal to the sum of all those values $\nu(\xi_{i+1}^*)$, so that these coherent values do define a probability measure ν on the space $\mathbf{X}$ of maximal paths. For each finite path $\xi_i = (\hat{\mathrm{x}}, x_1, \ldots, x_i)$ in $\mathcal{X}$, it follows from the construction of the promenade that $|P(x_j)| \in \{1, \ldots, N^2(x_{j+1})\}$ when $0 \leqslant j < i$, and we get for the measure of ξ_i^* that

$$\nu(\xi_i^*) = \prod_{j=0}^{i-1} \frac{1}{|P(x_j)|} \geqslant \prod_{j=1}^{i} \frac{1}{N^2(x_j)} =: \frac{1}{M(x_i)}.$$

We see as before in Corollary 2 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 ξ is a finite path and if $x = (V, t, r, n) \in \mathcal{X}$ is the last place visited by ξ, the image by π of $\xi^* \subset \mathbf{X}$ is a compact subset of T that is contained in the closure of V in T: indeed, if $\mathbf{x} \in \xi^*$ and $V = V_i(\mathbf{x})$, we have that $t_{j+1}(\mathbf{x}) \in V_j(\mathbf{x}) \subset V_i(\mathbf{x}) = V$ for every integer $j \geqslant i$, and $t_{j+1}(\mathbf{x}) \in V$ tends to $\pi(\mathbf{x})$. If $r < r'$ and since $V \subset B(t, r)$, we get that

$$\pi(\xi^*) \subset \{s \in T : d(t, s) \leqslant r\} \subset B(t, r')$$

hence

$$\nu(\xi^*) \leqslant \mu(\pi(\xi^*)) \leqslant \mu(B(t, r')).$$

Let τ be an arbitrary point in T and let $\mathbf{x}$ be a path such that $\tau \in V_i(\mathbf{x})$ for each $i \geqslant 0$. Let us write $x_i(\mathbf{x}) = x_i = (V_i, t_i, r_i, n_i)$ and $\xi_i = (\hat{\mathrm{x}}, x_1, \ldots, x_i)$. 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

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

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 promenade on $\mathcal{X}$, 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 maximal path $\mathbf{x}$ in the set $\mathcal{X}$; for every integer $k \geqslant 0$, let $i_k(\mathbf{x})$ be the smallest integer $i \geqslant 0$ for which ([32]) 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 path $\mathbf{x}$ a "variable analog" $\sigma_2(\mathbf{x})$ of $\Sigma_2(T)$ by setting

$$\sigma_2(\mathbf{x}) = \sum_{k=0}^{\infty} r_{i_k(\mathbf{x})}\, b^{k/2}. \tag{50}$$

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

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

therefore

$$\sigma_2(\mathbf{x}) \leqslant \frac{b}{b^{1/2} - 1}\, \sigma_1(\mathbf{x}). \tag{51}$$

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 $X_k \subset \mathcal{X}$ of places $x_{i_k(\mathbf{x})}(\mathbf{x})$, where $\mathbf{x}$ varies in the set of maximal paths of $\mathcal{X}$. The places $x \in X_k$ are characterized by

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

The places in X_k are the end-points of the maximal paths of the promenade restricted to the subset $\mathcal{X}_M$ from Equation (22), 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 Corollary 2, 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 path $\mathbf{x}$, hence $x_{i_0(\mathbf{x})}(\mathbf{x}) = \hat{\mathrm{x}}$: the subset $\mathcal{X}_{M_0}$ is reduced to the start $\hat{\mathrm{x}} \in \mathcal{X}$, and therefore $X_0 = \{\hat{\mathrm{x}}\}$.

For every index $k \geqslant 0$, let T_k denote the set of *positions* $\theta(x) \in T$ corresponding to the places $x \in X_k$, 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 $X_0 = \{\hat{\mathrm{x}}\} = \{(V_0, t_0, r_0, 1)\}$. If the expectation E^* of the supremum of the process $(X_t)_{t \in T}$ is finite, we know by (45) 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 (51) 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 path $\mathbf{x} = \mathbf{x}(\tau)$ such that τ belongs to the region $R(x)$ for each place x in $\mathbf{x}$; if $k \geqslant 0$ is given, consider the index $i = i_k(\mathbf{x})$ associated to that path. Then $x_i(\mathbf{x}) = x_{i_k(\mathbf{x})}(\mathbf{x})$ is an element of X_k. This means that if we write $x_i(\mathbf{x}) = (V_i, t_i, r_i, n_i)$, 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

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

We finally abandon the set $\mathcal{X}$, the promenade and control function. Here is the final word: the sets (T_k) are finite subsets of T such that

$$\left\{ \begin{aligned} &|T_0| = 1\,; \qquad \ln |T_k| < 3\, b^k \quad \text{and} \\ &\text{for every } \tau \in T, \quad \sum_{k=0}^{\infty} \mathrm{dist}(\tau, T_k)\, b^{k/2} \leqslant K_2. \end{aligned} \right. \tag{52}$$

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 for 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 (52) and such that

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

We can at last state Talagrand's theorem ([33]).

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 (52).*

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

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

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 Euclidean space $\mathbb{R}^n$ to a normed linear space F. Consider the standard Gaussian probability measure γ_n on $\mathbb{R}^n$, defined by

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

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

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

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 on it, 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 by $a : \mathbb{R}^n \to F$—, 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 have said that each linear functional ξ can be viewed as a (centered) Gaussian random variable defined on $(\mathbb{R}^n, \gamma_n)$ by

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

Then the family of random variables $\xi = t \in T$ is a centered Gaussian process $(X_t)_{t\in T}$, where $X_t = t = \xi$, 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 with 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 of course related to the notion of a *Radon measure*, a measure on a topological space M that is supported by a countable union of compact subsets of M. It is by attending the 1969 seminar [Sch] on this radonifying 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 $r B_H$ 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 $r B_H$ 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 $\Gamma(r B_H) = 0$, 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 Γ-almost everywhere 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. In the same way we can 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 family 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 set $K_u \subset \mathbb{R}^{\mathbb{N}}$ 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

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

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 $\ell^\infty(\mathbb{N})$, a *Radon measure* on the topological space $\ell^\infty(\mathbb{N})$ equipped with the weak* topology $\sigma(\ell^\infty(\mathbb{N}), \ell^1(\mathbb{N}))$.

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

$$\int_F \|y\|_F^p \, \mathrm{d}a(\mu)(y) = \int_E \|a\,x\|_F^p \, \mathrm{d}\mu(x) \leqslant C^p \sup_{x^* \in B(E^*)} \int_E \bigl|\langle x^*, x\rangle\bigr|^p \, \mathrm{d}\mu(x). \tag{53}$$

The p-summing norm ([34]) $\pi_p(a)$ of a is defined to be the smallest possible C in the above inequality. Restricting the definition to the Dirac probability measures on E, we can see that $\|a\| \leqslant \pi_p(a)$.

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 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

$$\|a\,x\|_F^p \leqslant C^p \int_{B_{E^*}} \bigl|\langle x^*, x\rangle\bigr|^p \, \mathrm{d}\mathrm{P}_a(x^*), \quad x \in E. \tag{54}$$

Deducing Inequality (53) from this is a simple matter of applying the Fubini theorem, followed by an easy $L^1(\mathrm{P}_a)$–$L^\infty(\mathrm{P}_a)$ bound,

$$\begin{aligned}
\int_F \|y\|_F^p \, \mathrm{d}a(\mu)(y) &= \int_E \|a\,x\|_F^p \, \mathrm{d}\mu(x) \\
&\leqslant C^p \int_{B_{E^*}} \Bigl(\int_E \bigl|\langle x^*, x\rangle\bigr|^p \, \mathrm{d}\mu(x) \Bigr) \, \mathrm{d}\mathrm{P}_a(x^*) \\
&\leqslant C^p \sup_{x^* \in B_{E^*}} \Bigl(\int_E \bigl|\langle x^*, x\rangle\bigr|^p \, \mathrm{d}\mu(x) \Bigr).
\end{aligned} \tag{55}$$

Proving the existence of the Pietsch measure requires a fairly 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 involving a sort of "standard model" for p-summing maps. Let K_E denote the compact space B_{E^*} equipped with the weak* topology. Classically, one can think of E as being the subspace of $C(K_E)$ (the space of continuous functions on K_E) consisting of the functions

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

with the Hahn–Banach theorem implying that E is *isometrically* embedded in $C(K_E)$ in that way. The preceding chain of inequalities (55) 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 $L^p(K_E, \mathrm{P}_a)$. 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 (54) 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 the space $L^p(K_E, \mathrm{P}_a)$. 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 Aleksander 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

$$\sum_{j=1}^{n} \|a\,x_j\|_F^p \leqslant C^p \sup_{x^* \in B_{E^*}} \sum_{j=1}^{n} \bigl|\langle x^*, x_j\rangle\bigr|^p, \qquad (x_j)_{j=1}^n \subset E. \tag{56}$$

Replacing each vector x_j with $\lambda_j^{1/p} y_j$ for some real $\lambda_j > 0$, we see that (56) is directly equivalent to restricting our definition given at (53) to the finitely supported nonnegative measures $\mu = \sum_{j=1}^n \lambda_j\,\delta_{y_j}$ on E. Passing by density from finitely supported to general finite measures, we see that (53) is equivalent to (56).

Suppose now that $p \geqslant 1$ and that a linear mapping $a : E \to F$ satisfies

$$\int_E \|a\,x\|_F \,\mathrm{d}\mu(x) \leqslant C \sup_{x^* \in B(E^*)} \Bigl(\int_E \bigl|\langle x^*, x\rangle\bigr|^p \,\mathrm{d}\mu(x)\Bigr)^{1/p}, \tag{57}$$

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 (57), this implies that *when the mapping a is p-summing, it is also q-summing for all $q \geqslant p$.*

We show now that (57) implies (56). Let $(x_j)_{j=1}^n$ be given and let us try to prove (56). 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

$$\lambda_j = \|a x_j\|_F^p > 0, \quad y_j = \frac{1}{\|a x_j\|_F} x_j \in E, \quad \mu_j = \frac{\lambda_j}{\sum_{i \in R} \lambda_i}.$$

We observe that $\|a y_j\|_F = 1$, and (57) applied to $\mu = \sum_{j \in R} \mu_j \, \delta_{y_j}$ gives the result,

$$\begin{aligned} 1 = \sum_{j \in R} \mu_j = \sum_{j \in R} \mu_j \, \|a y_j\|_F &= \int_E \|a y\|_F \, \mathrm{d}\mu(y) \\ &\underset{(57)}{\leqslant} C \sup_{x^* \in B_{E^*}} \Bigl(\sum_{j \in R} \mu_j \, |\langle x^*, y_j \rangle|^p \Bigr)^{1/p} \\ &= C \sup_{x^* \in B_{E^*}} \Bigl(\frac{\sum_{j \in R} |\langle x^*, x_j \rangle|^p}{\sum_{j \in R} \|a x_j\|_F^p} \Bigr)^{1/p}. \qquad \square \end{aligned}$$

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 $\Phi(0) = 0$, and we extend Φ to $\mathbb{R}$ by letting $\Phi(t) = \Phi(|t|)$ when $t < 0$. If (S, ν) is a measure space, one says that a measurable function f on S belongs to the space $L^\Phi(\nu)$ when there is a real $\delta > 0$ such that the integral of $\Phi(\delta f)$ with respect to ν is finite; hence, Φ and $u \mapsto \Phi(cu)$ define the same space of functions, for every $c > 0$.

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 [Fer3]. 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,

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

and conversely, this bound on the tail yields for example that $\mathrm{E}\, \Phi_2(\frac{1}{2} X)$ is finite. It is rather 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 (35)

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

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

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

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 (57) for example: for some C and for every *probability* measure μ on E, one has

$$\sup_{\xi \in B(E^*)} \bigl\| x \mapsto \langle \xi, x \rangle \bigr\|_{L^\Phi(\mu)} \leqslant 1 \quad \Rightarrow \quad \int_E \|a x\|_F \,\mathrm{d}\mu(x) \leqslant C. \tag{58}$$

Suppose that a linear map $a \neq 0$ satisfies (58), and consider the set Σ of functions f on the unit ball B_{E^*} of the dual E^* of E that have the form

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

where $n \geqslant 1$, $\lambda_i \geqslant 0$ and $\sum_j \lambda_j = 1$, and $\|a y_i\|_F > C$ for every $i = 1, 2, \ldots, n$. This set Σ is a convex set of functions that are continuous on the compact $K_E = B_{E^*}$ equipped with the weak* topology. One sees from (58) applied to $\mu = \sum_i \lambda_i \,\delta_{y_i}$ that Σ 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 $f \in \Sigma$, in particular,

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

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

$$\int_{K_E} \Phi\Bigl(C \,\frac{\langle \xi, x \rangle}{\|a x\|_F} \Bigr) \,\mathrm{d}\mathrm{P}_a(\xi) \geqslant 1 \ \text{ when } \ a x \neq 0, \ \text{ or } \ \|a x\|_F \leqslant C \bigl\| \xi \mapsto \langle \xi, x \rangle \bigr\|_{L^\Phi(\mathrm{P}_a)} \tag{59}$$

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 (59), with the Pietsch measure, back to the definition (58) 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 with $t = \|a x\|_F^{-1} < 1$ that

$$\|a x\|_F \,\Phi\Bigl(\frac{\langle \xi, x \rangle}{\|a x\|_F} \Bigr) = \Phi\Bigl(\frac{\langle \xi, x \rangle}{\|a x\|_F} \Bigr) \Big/ \Bigl(\frac{1}{\|a x\|_F} \Bigr) \leqslant \Phi\bigl(\langle \xi, x \rangle \bigr).$$

Using (59), where we suppose that $C = 1$ for simplicity, we have

$$1 \leqslant \int_{K_E} \Phi\Bigl(\frac{\langle \xi, x \rangle}{\|a x\|_F} \Bigr) \,\mathrm{d}\mathrm{P}_a(\xi),$$

then, when $\|a x\|_F \geqslant 1$ we get

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

Suppose that μ is a probability measure on E such that

$$\sup_{\xi \in B(E^*)} \bigl\| x \mapsto \langle \xi, x \rangle \bigr\|_{L^\Phi(\mu)} \leqslant 1, \quad \text{or} \quad \sup_{\xi \in K_E} \int_E \Phi\bigl(\langle \xi, x \rangle \bigr) \,\mathrm{d}\mu(x) \leqslant 1.$$

It follows that

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

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 (58) with $C = 2$, which is somewhat unsatisfactory, but it 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 (57) and (58), that being Φ_2-summing is a weaker property 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

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

seen previously in Section 4.1. Indeed, let μ be a probability measure on $E = \ell^\infty(\mathbb{N})$ that satisfies a uniform Φ_2-scalar estimate; let us suppose for instance that for every coordinate function $x \in \ell^\infty(\mathbb{N}) \mapsto x_i$ with $i \geqslant 0$, we have

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

By Markov again, it implies for each coordinate function that

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

Then

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

This yields of course that the norm function $x \in \ell^\infty(\mathbb{N}) \mapsto \|\beta x\|_\infty$ belongs to $L^{\Phi_2}(\mu)$.

From this property of β follows that every linear mapping $a : E \to F$ that factors through a "sub-object" of β satisfies

$$\sup_{\xi\in B(E^*)} \big\|x \mapsto \langle \xi, x\rangle\big\|_{L^{\Phi_2}(\mu)} \leqslant 1 \quad \Rightarrow \quad \Big\|x \mapsto \|ax\|_F\Big\|_{L^{\Phi_2}(\mu)} \leqslant C \tag{60}$$

for some C and every probability measure μ on E. By such a "sub-object factorization", we mean that there is a mapping $\alpha : E \to \ell^\infty(\mathbb{N})$, linear and bounded, such that

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

In particular, any diagonal map $\delta : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ such that $|\delta_i| \leqslant c\beta_i$ for some c and every $i \geqslant 0$ can be written $\delta = \beta\circ\alpha$, with α the bounded diagonal map on $\ell^\infty(\mathbb{N})$ that is given by $\alpha_i = \delta_i/\beta_i$ for $i \geqslant 0$. These maps δ are thus Φ_2-summing.

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

$$\rho^{-1} = \sum_{i=0}^{\infty} \frac{1}{(i+3)^2} < \int_2^\infty \frac{\mathrm{d}t}{t^2} = \frac{1}{2}, \quad \text{thus} \quad \rho > 2.$$

The probability measure P_β on $\ell^1(\mathbb{N})$ is in particular a measure on the unit ball of the dual of $\ell^\infty(\mathbb{N})$ —this dual "contains" $\ell^1(\mathbb{N})$—, and we shall see that

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

Indeed, 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. *Closing the path*

Among the young people in the (very) little group of researchers interested in Functional Analysis and Banach Spaces in the '70s at the "Centre de Mathématiques" directed by Laurent Schwartz at École Polytechnique, Paris, there was a strong belief in an affirmative answer to the following question (but not the slightest clue for a solution): *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 on the Hilbert space will be a true measure on the range space; in the other direction, it is almost evident 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 (52) 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: by definition, the image of the cylindrical measure γ_H on H is a Gaussian Radon measure ν on F. We then have, for example by [Fer$_1$], that

$$\int_F \|y\|_F\,\mathrm{d}\nu(y) < \infty.$$

We can describe the measure ν as 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 $i \geqslant n$. Now

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

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

$$\begin{aligned} &\lim_n \int_{H_n} \|a\,x\|_F\,\mathrm{d}\gamma_n(x) \\ (61) \qquad &= \lim_n \int_{H_n} \sup_{y^*\in B_{F^*}} \langle a^*y^*, \omega\rangle\,\mathrm{d}\gamma_n(\omega) \\ &= \int_{\mathbb{R}^\mathbb{N}} \sup_{y^*\in B_{F^*}} \langle a^*y^*, \omega\rangle\,\mathrm{d}\Gamma(\omega) = \int_F \|y\|_F\,\mathrm{d}\nu(y) < \infty. \end{aligned}$$

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 $\sum_{i \in \mathbb{N}} \xi_i^2 = \|\xi\|^2_{\ell^2(\mathbb{N})}$.

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 (61). Note that

$$d(s,t) := \|X_s - X_t\|_{L^2(\Gamma)} = \|\omega \mapsto \langle s - t, \omega \rangle\|_{L^2(\Gamma)} = \|s - t\|_{\ell^2(\mathbb{N})}.$$

We shall use (52) 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 (52) 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

$$|T_0| = 1, \quad \ln |T_k| < 3\, b^k, \quad \sum_{k \geqslant 0} \mathrm{dist}(t, T_k)\, b^{k/2} \leqslant K_2, \quad t \in T, \tag{62}$$

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

$$\xi(s,t) = \frac{1}{d(t,s)}\, (t - s) \in H^* \tag{63}$$

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

$$\ln |\Xi_k| \leqslant 1 + 3\, b^k + 3\, b^{k+1} < (1 + 3 + 3b)\, b^k < 7b.b^k \tag{64}$$

because $b > 1$ and $\ln(1+u) < 1 + \ln u$ when $u \geqslant 1$. 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 $(d_j)_{j \geqslant 0}$ as subsequence. We define a first linear mapping f from $\ell^1(\mathbb{N})$ to H^* by letting

$$f(e_n) = x_n^*, \tag{65}$$

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

$$f^*(x) = \bigl(\langle x_n^*, x \rangle\bigr)_{n \geqslant 0}, \quad \text{and} \quad \|f^*(x)\|_\infty = \sup_n \bigl|\langle x_n^*, x \rangle\bigr| \geqslant \sup_j \bigl|\langle d_j, x \rangle\bigr| = \|x\|_H \tag{66}$$

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 B_{H^*}. The mapping f^* is therefore an isometry from H into $\ell^\infty(\mathbb{N})$.

Next, we need find a diagonal map $\delta : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$, given by

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

with coefficients $(\delta_n)_{n\geqslant 0}$ satisfying $|\delta_n| \leqslant c\beta_n$ for some c and all integers $n \geqslant 0$, *i.e.*, such that the sequence $(\sqrt{\ln(n+3)}\,\delta_n)$ 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

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

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

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

When $n \in I_k$, we have $\delta^{(1)}(e_n) = b^{-k/2}\,e_n$, or

$$e_n = b^{k/2}\,\delta^{(1)}(e_n). \tag{67}$$

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 (62) 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 $t_0 = t_0(\tau) = 0$). Recall from (63) that

$$t_{k+1}(\tau) - t_k(\tau) = d\big(t_k(\tau), t_{k+1}(\tau)\big)\,\xi\big(t_k(\tau), t_{k+1}(\tau)\big).$$

For a certain $n_k(\tau) \in I_k \subset \mathbb{N}$ we have

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

Now by (67) and the definition (65) of f we have

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

and therefore

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

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

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

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

$$\|a\, x\|_F = \sup_{\tau \in T} \langle \tau, x \rangle \leqslant 2\, K_2\, \|\delta\, f^* x\|_\infty, \tag{68}$$

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

$$\begin{array}{ccccccc} & & \ell^\infty & \xrightarrow[\delta]{} & \ell^\infty & & \\ & & \cup & & \cup & & \\ H & \xrightarrow[f^*]{} & E_\infty & \xrightarrow[\delta_1]{} & F_\infty & \xrightarrow[a_1]{} & F \end{array} \tag{69}$$

where f^* is the isometric embedding of H into $\ell^\infty(\mathbb{N})$ given at (66), 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 (68) 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

$$\delta_n = b^{-k/2} \leqslant c\, \beta_n = \frac{c}{\sqrt{\ln(n+3)}}, \quad \text{or} \quad \ln(n+3) \leqslant c^2\, b^k, \quad k \geqslant 0. \tag{70}$$

Knowing that $n \in I_k$ and letting $\eta = 7\, b$, we have by (64) that

$$n \leqslant \sum_{s=0}^{k} |I_s| = \sum_{s=0}^{k} |\Xi_s| \leqslant \sum_{s=0}^{k} \exp(\eta\, b^s).$$

Because $b > 1/(\ln 2) > 1$ we can check([36]) that $\eta\,(b-1) > \ln 2$, and when $s \geqslant 1$,

$$b^s - b^{s-1} \geqslant b - 1, \quad \exp(\eta\, b^{s-1}) \leqslant \mathrm{e}^{\eta(1-b)} \exp(\eta\, b^s) < (1/2) \exp(\eta\, b^s).$$

It follows that

$$n \leqslant \Bigl(\sum_{n=0}^{k} 2^{-n} \Bigr) \exp(\eta\, b^k) < 2 \exp(7b\,.\,b^k)$$

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 (70) with $c = \sqrt{11\, b}$. $\square$

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 (68) in the form

$$\|a x\|_F \leqslant 2 K_2 \|\delta f^* x\|_\infty = 2 K_2 \|\alpha \beta f^* x\|_\infty \leqslant 2 c K_2 \|\beta f^* x\|_\infty,$$

then modify the diagram (69) and factor a through a "sub-object" β_1 of our canonical diagonal mapping β,

$$\begin{array}{ccccccc} & & \ell^\infty & \xrightarrow[\beta]{} & \ell^\infty & & \\ & & \bigcup & & \bigcup & & \\ H & \xrightarrow[f^*]{} & E_\infty & \xrightarrow[\beta_1]{} & G_\infty & \xrightarrow[a_2]{} & F \end{array}$$

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 $\|a_2\| \leqslant 2 c K_2$.

5. Appendix

5.1. *Expectation of the maximum of a few independent Gaussian variables*

It is quite possible that *Mathematica* or another software of the kind can immediately give the formulas obtained in this section. We will nevertheless go on with our own calculations. Let $g_1, g_2, \ldots, g_n$ denote independent $\mathrm{N}(0,1)$ Gaussian random variables, and set

$$g_n^* = \max_{1 \leqslant i \leqslant n} g_i.$$

We shall obtain expressions for $\mathrm{E}\, g_n^*$ when $n = 2$, $n = 3$, $n = 4$ and $n = 5$.

Consider the Gaussian distribution function Φ, defined for $x \in \mathbb{R}$ by

$$\Phi(x) = \int_{-\infty}^{x} \mathrm{e}^{-u^2/2}\, \rho\, \mathrm{d}u = \mathrm{P}(g_1 < x), \quad \text{where} \quad \rho = \frac{1}{\sqrt{2\pi}}.$$

By independence, we know that

$$\mathrm{P}(g_n^* < x) = \Phi(x)^n,$$

so that the density of the distribution of g_n^* is given by

$$n \Phi(x)^{n-1} \rho\, \mathrm{e}^{-x^2/2}, \quad \text{and we have} \quad \int_{\mathbb{R}} \Phi(x)^{n-1}\, \mathrm{e}^{-x^2/2}\, \mathrm{d}x = \frac{1}{\rho n}. \tag{71}$$

For θ such that $0 < \theta < \pi/2$ and $k \geqslant 0$ an integer, consider

$$A_k(\theta) = \int_{\mathbb{R}} \Phi(x)^k\, \mathrm{e}^{-(\tan^2\theta) x^2/2}\, \mathrm{d}x,$$

$$B_k(\theta) = \int_{\mathbb{R}} \Phi(x)^k\, \mathrm{e}^{-(\tan^2\theta) x^2/2}\, x\, \mathrm{d}x, \quad C_k(\theta) = \int_{\mathbb{R}} \Phi(x)^k\, \mathrm{e}^{-(\tan^2\theta) x^2/2}\, x^2\, \mathrm{d}x.$$

We immediately see that

$$A_0(\theta) = \int_{\mathbb{R}} \mathrm{e}^{-(\tan^2\theta) x^2/2}\, \mathrm{d}x = \frac{1}{\rho \tan\theta}, \tag{72}$$

$$B_0(\theta) = \int_{\mathbb{R}} \mathrm{e}^{-(\tan^2\theta) x^2/2}\, x\, \mathrm{d}x = 0, \quad C_0(\theta) = \int_{\mathbb{R}} \mathrm{e}^{-(\tan^2\theta) x^2/2}\, x^2\, \mathrm{d}x = \frac{1}{\rho \tan^3\theta}.$$

Let $\theta_0 = \pi/4$, so that $\tan\theta_0 = 1$. We have already seen at (71) that

$$A_k(\theta_0) = \int_{\mathbb{R}} \Phi(x)^k \,\mathrm{e}^{-x^2/2}\,\mathrm{d}x = \frac{1}{\rho(k+1)}.$$

Note that

$$\mathrm{E}\, g_n^* = \int_{\mathbb{R}} n\,\Phi(x)^{n-1}\rho\,\mathrm{e}^{-x^2/2}\,x\,\mathrm{d}x = n\,\rho\,B_{n-1}(\theta_0).$$

Using integration by parts and letting $T = \tan\theta > 0$, we obtain

$$\begin{aligned} B_k(\theta) &= \frac{1}{T^2}\int_{\mathbb{R}} \Phi(x)^k \,\mathrm{e}^{-T^2x^2/2}\,T^2\,x\,\mathrm{d}x \\ &= \frac{1}{T^2}\Big(\Big[-\Phi(x)^k\,\mathrm{e}^{-T^2x^2/2}\Big]_{-\infty}^{\infty} + \int_{\mathbb{R}} k\,\Phi(x)^{k-1}\rho\,\mathrm{e}^{-x^2/2}\,\mathrm{e}^{-T^2x^2/2}\,\mathrm{d}x\Big) \\ &= \frac{k\,\rho}{\tan^2\theta}\int_{\mathbb{R}} \Phi(x)^{k-1}\,\mathrm{e}^{-(1+\tan^2\theta)x^2/2}\,\mathrm{d}x. \end{aligned}$$

Let us define $\theta^{(1)}, \theta^{(2)} \in (0, \pi/2)$ by $\tan^2(\theta^{(j)}) = j + \tan^2\theta$, $j = 1, 2$. We see that

$$B_k(\theta) = \frac{k\,\rho}{\tan^2\theta}\,A_{k-1}(\theta^{(1)}).$$

When $k = 1$, we have

$$B_1(\theta) = \frac{\rho}{\tan^2\theta}\,A_0(\theta^{(1)}) = \frac{\rho}{\tan^2\theta}\,\frac{1}{\rho\tan\theta^{(1)}} = \frac{1}{(\tan^2\theta)\sqrt{1+\tan^2\theta}}.$$

Letting $\theta_1 = (\theta_0)^{(1)}$, so that $\tan\theta_1 = \sqrt{2}$, we get

$$\mathrm{E}\, g_n^* = n\,\rho\,B_{n-1}(\theta_0) = n\,\rho\,\frac{(n-1)\,\rho}{\tan^2\theta_0}\,A_{n-2}(\theta_1),$$

and we shall keep it for later as

$$(G_n) \qquad \mathrm{E}\, g_n^* = n(n-1)\,\rho^2\,A_{n-2}(\theta_1).$$

Using (72) we obtain immediately

$$\mathrm{E}\, g_2^* = 2\,\rho^2\,A_0(\theta_1) = 2\,\rho^2\,\frac{1}{\rho\tan\theta_1} = \frac{2\,\rho}{\sqrt{1+\tan^2\theta_0}} = \frac{2\,\rho}{\sqrt{2}},$$

and thus

$$(G_2) \qquad \mathrm{E}\, g_2^* = \frac{1}{\sqrt{\pi}}.$$

For C_k, with $T = \tan\theta > 0$ again, we have

$$\begin{aligned} C_k(\theta) &= \frac{1}{T^2}\int_{\mathbb{R}} \Phi(x)^k\,x\,.\,\mathrm{e}^{-T^2x^2/2}\,T^2\,x\,\mathrm{d}x \\ &= \frac{1}{T^2}\Big(\Big[-\Phi(x)^k\,x\,.\;\mathrm{e}^{-T^2x^2/2}\Big]_{-\infty}^{\infty} + \int_{\mathbb{R}} \big(\Phi(x)^k\,x\big)'\,\mathrm{e}^{-T^2x^2/2}\,\mathrm{d}x\Big) \\ &= \frac{1}{T^2}\Big(\int_{\mathbb{R}} k\Phi(x)^{k-1}\rho\,x\,\mathrm{e}^{-(1+\tan^2\theta)x^2/2}\,\mathrm{d}x + \int_{\mathbb{R}} \Phi(x)^k\,\mathrm{e}^{-(\tan^2\theta)x^2/2}\,\mathrm{d}x\Big). \end{aligned}$$

It follows that

$$C_k(\theta) = \frac{k\,\rho}{\tan^2\theta}\,B_{k-1}(\theta^{(1)}) + \frac{1}{\tan^2\theta}\,A_k(\theta),$$

therefore

$$C_1(\theta) = \frac{1}{\tan^2\theta}\,A_1(\theta).$$

When $k > 1$, we rewrite

$$C_k(\theta) = \frac{k\,\rho}{\tan^2\theta}\,\frac{(k-1)\,\rho}{\tan^2\theta^{(1)}}\,A_{k-2}(\theta^{(2)}) + \frac{1}{\tan^2\theta}\,A_k(\theta),$$

thus

$$C_k(\theta) = \frac{k\,(k-1)\,\rho^2}{(\tan^2\theta)\,\tan^2(\theta^{(1)})}\,A_{k-2}(\theta^{(2)}) + \frac{1}{\tan^2\theta}\,A_k(\theta).$$

We shall try to identify $A_k(\theta)$ from its derivative and its value at θ_0,

$$A_k'(\theta) = -\int_{\mathbb{R}} \Phi(x)^k\,\frac{\tan\theta}{\cos^2\theta}\,x^2\;\mathrm{e}^{-(\tan^2\theta)x^2/2}\;\mathrm{d}x = -\frac{\tan\theta}{\cos^2\theta}\,C_k(\theta).$$

When $k = 1$ we see that

$$A_1'(\theta) = -\frac{\tan\theta}{\cos^2\theta}\,C_1(\theta) = -\frac{\tan\theta}{\cos^2\theta}\,\frac{1}{\tan^2\theta}\,A_1(\theta) = -\frac{1}{\sin\theta\cos\theta}\,A_1(\theta).$$

We have first to solve

$$f'(\theta) = -\frac{1}{\sin\theta\cos\theta}\,f(\theta),$$

for which we choose

$$f(\theta) = \frac{1}{\tan\theta} = \frac{\cos\theta}{\sin\theta}, \qquad f(\theta_0) = 1.$$

We thus obtain that

$$A_1(\theta) = A_1(\theta_0)\,f(\theta) = A_1(\theta_0)\,\frac{\cos\theta}{\sin\theta} = \frac{1}{2\rho}\,\frac{\cos\theta}{\sin\theta} = \frac{1}{2\rho\,\tan\theta}.$$

With this we get

$$\mathrm{E}\,g_3^* = 6\rho^2\,A_1(\theta_1) = \frac{3\rho}{\tan(\theta_1)}$$

and we have $\tan\theta_1 = \sqrt{2}$, thus

$$\mathrm{E}\,g_3^* = \frac{3\rho}{\sqrt{2}} = \frac{3}{2\sqrt{\pi}}. \tag{G_3}$$

When $k > 1$ let us come back to

$$\begin{aligned} A_k'(\theta) &= -\frac{\tan\theta}{\cos^2\theta}\,C_k(\theta) \\ &= -\frac{\tan\theta}{\cos^2\theta}\left(\frac{k(k-1)\rho^2}{\tan^2\theta\,\tan^2(\theta^{(1)})}\,A_{k-2}(\theta^{(2)}) + \frac{1}{\tan^2\theta}\,A_k(\theta)\right) \end{aligned}$$

thus

$$A_k'(\theta) = -\frac{k\,(k-1)\,\rho^2}{\sin\theta\cos\theta\tan^2(\theta^{(1)})}\,A_{k-2}(\theta^{(2)}) - \frac{1}{\sin\theta\cos\theta}\,A_k(\theta).$$

The solution for A_k will take the form $A_k(\theta) = u_k(\theta)\,f(\theta)$, where we must have

$$u_k'(\theta)f(\theta) + u_k(\theta)f'(\theta) = -\frac{k\,(k-1)\,\rho^2}{\sin\theta\cos\theta\tan^2(\theta^{(1)})}\,A_{k-2}(\theta^{(2)}) - \frac{1}{\sin\theta\cos\theta}\,u_k(\theta)f(\theta),$$

therefore

$$\begin{aligned} u_k'(\theta) &= -\frac{\sin\theta}{\cos\theta}\,\frac{k\,(k-1)\,\rho^2}{\sin\theta\cos\theta\tan^2(\theta^{(1)})}\,A_{k-2}(\theta^{(2)}) \\ &= -\frac{k\,(k-1)\,\rho^2\,(1+\tan^2\theta)}{\tan^2(\theta^{(1)})}\,A_{k-2}(\theta^{(2)}) = -k\,(k-1)\,\rho^2\,A_{k-2}(\theta^{(2)}). \end{aligned}$$

When $k = 2$,

$$u_2'(\theta) = -2\rho^2\, A_0(\theta^{(2)}) = -2\rho\, \frac{1}{\tan\theta^{(2)}}$$

or

$$u_2'(\theta) = -2\rho\, \frac{1}{\sqrt{2+\tan^2\theta}} = -2\rho\, \frac{\cos\theta}{\sqrt{2-\sin^2\theta}}$$

hence

$$u_2(\theta) = \kappa_2 - 2\rho\, \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big).$$

We have $\sin\theta_0 = \sin(\pi/4) = 1/\sqrt{2}$ and since $A_2 = u_2 f$ and $f(\theta_0) = 1$, we get

$$\frac{1}{3\rho} = A_2(\theta_0) = u_2(\theta_0) = \kappa_2 - 2\rho\, \arcsin\Big(\frac{1}{2}\Big) = \kappa_2 - \frac{2\rho\,\pi}{6} = \kappa_2 - \frac{1}{6\rho}$$

(notice that $2\pi\rho^2 = 1$) hence $\kappa_2 = 1/(3\rho) + 1/(6\rho) = 1/(2\rho)$. Finally,

$$\begin{aligned} u_2(\theta) &= \frac{1}{2\rho} - 2\rho\, \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big) = 2\rho\,\Big(\frac{1}{4\rho^2} - \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big)\Big) \\ &= 2\rho\,\Big(\frac{\pi}{2} - \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big)\Big) = 2\rho\, \arccos\Big(\frac{\sin\theta}{\sqrt{2}}\Big), \end{aligned}$$

and

$$A_2(\theta) = f(\theta)\, u_2(\theta) = \frac{\cos\theta}{\sin\theta}\,.2\rho\, \arccos\Big(\frac{\sin\theta}{\sqrt{2}}\Big) = \frac{2\rho}{\tan\theta}\, \arccos\Big(\frac{\sin\theta}{\sqrt{2}}\Big).$$

From $\sin^2(\theta^{(1)})/\cos^2(\theta^{(1)}) = 2$ we deduce $\sin(\theta^{(1)}) = \sqrt{2/3}$. With this we obtain now

$$\begin{aligned} \mathrm{E}\, g_4^* &= 12\,\rho^2\, A_2(\theta_1) = 12\,\rho^2\, \frac{2\rho}{\tan\theta_1}\, \arccos\Big(\frac{\sin\theta_1}{\sqrt{2}}\Big) \\ &= \frac{24\,\rho^3}{\sqrt{2}}\, \arccos\Big(\frac{1}{\sqrt{3}}\Big) = \frac{6}{\pi^{3/2}}\, \arccos\Big(\frac{1}{\sqrt{3}}\Big), \end{aligned}$$

or

$$(G_4) \qquad \mathrm{E}\, g_4^* = \frac{6}{\pi^{3/2}}\, \arctan\sqrt{2}.$$

We move now to u_3. We know that

$$\begin{aligned} u_3'(\theta) &= -3.2\rho^2\, A_1(\theta^{(2)}) = -6\,\rho^2\, \frac{1}{2\rho}\, \frac{1}{\tan(\theta^{(2)})} \\ &= -3\,\rho\, \frac{1}{\tan(\theta^{(2)})} = -3\,\rho\, \frac{1}{\sqrt{2+\tan^2\theta}}. \end{aligned}$$

We have seen this before. We get

$$u_3(\theta) = \kappa_3 - 3\,\rho\, \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big).$$

Now

$$u_3(\theta_0) = A_3(\theta_0) = \frac{1}{4\rho} = \kappa_3 - 3\rho\, \arcsin\Big(\frac{1}{2}\Big) = \kappa_3 - \frac{3\rho\pi}{6} = \kappa_3 - \frac{3\rho}{12\rho^2} = \kappa_3 - \frac{1}{4\rho}$$

thus $\kappa_3 = 1/(2\rho)$ and

$$\begin{aligned} A_3(\theta) &= u_3(\theta)\, f(\theta) = \Big(\frac{1}{2\rho} - 3\,\rho\, \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big)\Big)\, \frac{\cos\theta}{\sin\theta} \\ &= 3\,\rho\, \frac{\cos\theta}{\sin\theta}\, \Big(\frac{1}{6\,\rho^2} - \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big)\Big) = 3\,\rho\, \frac{\cos\theta}{\sin\theta}\, \Big(\frac{\pi}{3} - \arcsin\Big(\frac{\sin\theta}{\sqrt{2}}\Big)\Big). \end{aligned}$$

With this we obtain now

$$\begin{aligned}\mathrm{E}\, g_5^* &= 20\,\rho^2\, A_3(\theta_1) = 20\,\rho^2\, \frac{3\,\rho}{\tan\theta_1}\,\Bigl(\frac{\pi}{3} - \arcsin\Bigl(\frac{\sin\theta_1}{\sqrt{2}}\Bigr)\Bigr)\\ &= \frac{60\,\rho^3}{\sqrt{2}}\,\Bigl(\frac{\pi}{3} - \arcsin\Bigl(\frac{1}{\sqrt{3}}\Bigr)\Bigr) = \frac{60}{\sqrt{2}\,(2\pi)^{3/2}}\,\Bigl(\frac{\pi}{3} - \arcsin\Bigl(\frac{1}{\sqrt{3}}\Bigr)\Bigr),\end{aligned}$$

thus

$$(G_5) \qquad \mathrm{E}\, g_5^* = \frac{15}{\pi^{3/2}}\,\Bigl(\frac{\pi}{3} - \arcsin\Bigl(\frac{1}{\sqrt{3}}\Bigr)\Bigr) > 1.162.$$

The next step would be to solve

$$u_4'(\theta) = -12\,\rho^2\, A_2(\theta^{(2)}) = -12\,\rho^2\, \frac{2\rho}{\tan(\theta^{(2)})}\, \arccos\Bigl(\frac{\sin(\theta^{(2)})}{\sqrt{2}}\Bigr).$$

We see that $\cos(\theta^{(2)}) = 1/\sqrt{3+\tan^2\theta}$, and

$$\begin{aligned}u_4'(\theta) &= -\frac{24\,\rho^3}{\sqrt{2+\tan^2\theta}}\, \arccos\Bigl(\sqrt{\frac{2+\tan^2\theta}{6+2\tan^2\theta}}\Bigr)\\ &= -\frac{24\,\rho^3}{\sqrt{2+\tan^2\theta}}\, \arctan\Bigl(\sqrt{\frac{4+\tan^2\theta}{2+\tan^2\theta}}\Bigr).\end{aligned}$$

That seems to me pretty much intractable.

5.2. *Sketch of proof for the comparison result*

We shall sketch a proof of the Slepian–Sudakov comparison result, essentially following Li and Queffélec [LiQu], Chap. 8. For doing it, we can limit ourselves to Gaussian processes with a *finite* index set T, say $|T| = n$; in other words, we restrict the discussion to two centered Gaussian random vectors $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$. We define the function φ on $\mathbb{R}^n$ by

$$\varphi(u_1, u_2, \ldots, u_n) = \max(u_1, u_2, \ldots, u_n), \qquad u_i \in \mathbb{R},\ i = 1, 2, \ldots, n.$$

We have to prove that $\mathrm{E}\,\varphi(Y) \leqslant \mathrm{E}\,\varphi(X)$ under the Slepian–Sudakov hypothesis

$$(H) \qquad \mathrm{E}(Y_i - Y_j)^2 \leqslant \mathrm{E}(X_i - X_j)^2, \quad 1 \leqslant i, j \leqslant n.$$

We shall use some special properties of φ, stated in the next lemma.

Lemma 6. *Let $(e_i)_{i=1}^n$ be the canonical basis of $\mathbb{R}^n$. When $i \neq j$ we have*

$$(*) \qquad \varphi(x + te_i + t\,e_j) - \varphi(x + te_j) - \varphi(x + t\,e_i) + \varphi(x) \leqslant 0, \quad x \in \mathbb{R}^n,\ t \in \mathbb{R}.$$

Also,

$$(**) \qquad \varphi(x + t\textstyle\sum_{i=1}^n e_i) = \varphi(x_1{+}t, x_2{+}t, \ldots, x_n{+}t) = \varphi(x) + t, \quad x \in \mathbb{R}^n,\ t \in \mathbb{R}.$$

If ψ is a non-negative C^∞ function defined on $\mathbb{R}^n$, with a compact support, then the convolution $\Phi = \varphi * \psi$ satisfies everywhere that

$$\frac{\partial^2\Phi}{\partial x_i \partial x_j} \leqslant 0 \ \text{ when } \ i \neq j, \ \text{ and } \ \sum_{j=1}^n \frac{\partial^2\Phi}{\partial x_i \partial x_j} = 0 \ \text{ for } \ i = 1, \ldots, n.$$

Proof. The second property $(**)$ of φ is clear. For the first property $(*)$, we select $i \neq j$, we may suppose that $t > 0$ and $x_i \leqslant x_j$; then

$$\max(x_i + t, x_j + t) = \max(x_i, x_j + t), \quad \text{thus} \quad \varphi(x + t\,e_i + t\,e_j) = \varphi(x + t\,e_j),$$

the first difference in $(*)$ is therefore equal to zero, and $-\varphi(x+t\,e_i)+\varphi(x) \leqslant 0$ because φ is non-decreasing with respect to each of its n variables.

It is easy to see that convolutions with non-negative functions ψ will preserve the property $(*)$, while the property $(**)$ becomes

$$\Phi(x_1{+}t,\, x_2{+}t,\, \ldots,\, x_n{+}t) = \Phi(x) + t \int_{\mathbb{R}^n} \psi(y)\,\mathrm{d}y.$$

Then $(*)$ for Φ implies the first result for Φ'' when t tends to 0, the second is obtained by differentiating twice the above equality for Φ, first with respect to one of the variables, say x_i, killing the t term on the right-hand side, then with respect to t. $\square$

We have to prove that $\mathrm{E}\,\varphi(Y) \leqslant \mathrm{E}\,\varphi(X)$ and for this, it suffices to establish the corresponding result for the functions $\Phi = \varphi * \psi$. Indeed, using a sequence (ψ_n) of non-negative functions with integral 1 and supports tending to $\{0\}$, we will obtain the result for φ by approximation (37).

The convex and Lipschitz function φ admits partial right derivatives, everywhere equal to 0 or 1. If we denote them by $\partial_i\varphi$, $i = 1, 2, \ldots, n$, we can see for every $x \in \mathbb{R}^n$ (with Lebesgue's dominated convergence theorem for instance) that

$$\partial_i\Phi(x) = \int_{\mathbb{R}^n} \partial_i\varphi(x-y)\,\psi(y)\,\mathrm{d}y = \int_{\mathbb{R}^n} \partial_i\varphi(y)\,\psi(x-y)\,\mathrm{d}y.$$

Then, letting $\partial^\alpha = \partial_{j_1}\partial_{j_2}\ldots\partial_{j_k}$, the next partial derivatives of Φ can be written as

$$\partial^\alpha\partial_i\Phi(x) = \int_{\mathbb{R}^n} \partial_i\varphi(y)\,\partial^\alpha\psi(x-y)\,\mathrm{d}y, \quad x \in \mathbb{R}^n,$$

and they are therefore bounded on $\mathbb{R}^n$, because

$$|\partial^\alpha\partial_i\Phi(x)| \leqslant \int_{\mathbb{R}^n} \bigl|\partial^\alpha\psi(x-y)\bigr|\,\mathrm{d}y =: K_\alpha(\psi), \quad x \in \mathbb{R}^n. \tag{73}$$

This will be used below for bounding the remainder in the Taylor formula for Φ.

The obvious verification of the following lemma is left to the reader.

Lemma 7. *Let $A = (a_{i,j})$ be a symmetric $n \times n$ matrix such that $\sum_{j=1}^n a_{i,j} = 0$ for each $i = 1, \ldots, n$. One has for any $u = (u_i)_{i=1}^n$ in $\mathbb{R}^n$ that*

$$A(u,u) := \sum_{i,j} a_{i,j}\,u_iu_j = -\frac{1}{2}\sum_{i,j} a_{i,j}(u_i-u_j)^2 = -\frac{1}{2}\sum_{i\neq j} a_{i,j}(u_i-u_j)^2.$$

Letting X and Y be independent, we shall study the evolution of the function

$$t \;\mapsto\; \mathrm{E}\,\Phi\bigl(\sqrt{1-t}\,Y + \sqrt{t}\,X\bigr) =: f(t)$$

for t varying from 0 to 1, and use it for comparing $f(0) = \mathrm{E}\,\Phi(Y)$ and $f(1) = \mathrm{E}\,\Phi(X)$. Let $b^2 + \varepsilon^2 + a^2 = 1$, where $\varepsilon > 0$ is "small", a, b are $\geqslant 0$, and consider the rather small step from $t = s_0 = a$ to $t = s_1 = \sqrt{\varepsilon^2 + a^2}$, that we decide to describe in the following way: let $Y^{(1)}, Y^{(2)}$ be two "copies" of Y and $X^{(1)}, X^{(2)}$ two copies of X, with $Y^{(1)}, Y^{(2)}, X^{(1)}, X^{(2)}$ independent. Let us introduce

$$V_0 = b\,Y^{(1)} + \varepsilon\,Y^{(2)} + a\,X^{(2)}, \qquad V_1 = b\,Y^{(1)} + \varepsilon\,X^{(1)} + a\,X^{(2)}.$$

The distribution of V_0 is that of $\sqrt{1-s_0}\,Y + \sqrt{s_0}\,X$, while the distribution of V_1 is that of $\sqrt{1-s_1}\,Y + \sqrt{s_1}\,X$, hence $f(s_0) = \mathrm{E}\,\Phi(V_0)$ and $f(s_1) = \mathrm{E}\,\Phi(V_1)$. It will be convenient to set $U = b\,Y^{(1)} + a\,X^{(2)}$, so that $V_0 = U + \varepsilon\, Y^{(2)}$ and $V_1 = U + \varepsilon\, X^{(1)}$. With Taylor's formula for Φ, applied to the random elements U and $Y^{(2)}$ of $\mathbb{R}^n$, we can write

$$(74) \qquad \Phi(V_0) = \Phi(U + \varepsilon\, Y^{(2)}) = \Phi(U) + \varepsilon\, \Phi'_U \cdot Y^{(2)} + \frac{\varepsilon^2}{2}\, \Phi''_U(Y^{(2)}, Y^{(2)}) + R_0\,;$$

the (random) remainder R_0 written in integral form admits a bound that involves the third derivative $\Phi'''_{U+s\,Y^{(2)}}$ at the points of the segment $[U, U+\varepsilon\, Y^{(2)}]$. By (73) one has

$$|R_0| \leqslant K(\psi)\,.\,\varepsilon^3 \|Y^{(2)}\|^3,$$

where the norm for $Y^{(2)}$ is (for example) the Euclidean norm in $\mathbb{R}^n$. Using the last assertion in Lemma 6 we may apply Lemma 7 to the matrix $\bigl((\Phi''_U)_{i,j}\bigr)$ of Φ'' at the point U, and rewrite

$$\Phi(V_0) = \Phi(U) + \varepsilon\, \Phi'_U \cdot Y^{(2)} - \frac{\varepsilon^2}{4} \sum_{i \neq j} (\Phi''_U)_{i,j}\, (Y^{(2)}_i - Y^{(2)}_j)^2 + R_0.$$

Using independence and the finiteness of Gaussian moments we obtain

$$\mathrm{E}\, \Phi(V_0) = \mathrm{E}\, \Phi(U) - \frac{\varepsilon^2}{4} \sum_{i \neq j} \mathrm{E}(\Phi''_U)_{i,j}\; \mathrm{E}(Y_i - Y_j)^2 + C_0\, \varepsilon^3$$

and in the same way for $\Phi(V_1) = \Phi(U + \varepsilon\, X^{(1)})$, we have

$$\mathrm{E}\, \Phi(V_1) = \mathrm{E}\, \Phi(U) - \frac{\varepsilon^2}{4} \sum_{i \neq j} \mathrm{E}(\Phi''_U)_{i,j}\; \mathrm{E}(X_i - X_j)^2 + C_1\, \varepsilon^3,$$

with $|C_0|, |C_1| \leqslant C = C(\psi, X, Y)$, depending on the function ψ and on the distributions of X, Y only. From Lemma 6 we have that $(\Phi''_U)_{i,j} \leqslant 0$ when $i \neq j$, and from the hypothesis (H) of Slepian–Sudakov we conclude that

$$f(s_0) = \mathrm{E}\, \Phi(V_0) \leqslant \mathrm{E}\, \Phi(V_1) + (C_0 - C_1)\, \varepsilon^3 \leqslant f(s_1) + 2\, C\, \varepsilon^3.$$

At last, we can go from $t = t_0 = 0$ to $t = t_N = 1$ in N steps $t_i \to t_{i+1}$, in such a way that $t_{i+1}^2 - t_i^2 = \varepsilon^2 = 1/N$ for $0 \leqslant i < N$; treating each step as we did above when going from s_0 to s_1 (we had $s_1^2 - s_0^2 = \varepsilon^2$), we know that $f(t_i) - f(t_{i+1}) \leqslant 2\, C\, \varepsilon^3$, thus

$$\mathrm{E}\, \Phi(Y) = f(0) \leqslant f(1) + 2\, N\, C \varepsilon^3 = \mathrm{E}\, \Phi(X) + 2\, C/\sqrt{N},$$

with N that can be arbitrarily large. The Slepian–Sudakov result follows. $\square$

Remark 3. If we also want to prove the Slepian comparison lemma, we need to replace the function φ by a function k_s defined by

$$k_s(x) = \mathbf{1}_{\{\varphi(x) > s\}}, \quad x \in \mathbb{R}^n,$$

where s is an arbitrarily given real number, and to compare $\mathrm{E}\, k_s(Y)$ with $\mathrm{E}\, k_s(X)$. The function k_s has the form $k = h \circ \varphi$ where h is a non-decreasing function on the real line (in the case of k_s, it is $h_s = \mathbf{1}_{(s,\infty)}$). It is easy to see from the proof of Lemma 6 that any such function k satisfies $(*)$. If we regularize k as $K = k * \psi$ with a non-negative smooth function ψ, we still have therefore that

$$\frac{\partial^2 K}{\partial x_i \partial x_j} \leqslant 0 \quad \text{when} \quad i \neq j\,;$$

we can complete Lemma 7 by observing that for *any* symmetric matrix $(a_{i,j})$, one has

$$\sum_{i,j} a_{i,j}\, u_i u_j = -\frac{1}{2} \sum_{i \neq j} a_{i,j} (u_i - u_j)^2 + \sum_{i=1}^n \Bigl(\sum_{j=1}^n a_{i,j} \Bigr) u_i^2.$$

If we try to repeat the above proof with K instead of Φ, there will be no change for the treatment of the terms $a_{i,j} = \frac{\partial^2 K}{\partial x_i \partial x_j}$ with $i \neq j$, but now we have no control on the sign of the coefficient $\sum_{j=1}^n a_{i,j}$ of u_i^2: we have to have an equality here when we replace Y by X, that is to say, we need to assume that

$$(H') \qquad \mathrm{E}\, Y_i^2 = \mathrm{E}\, X_i^2, \quad i = 1, 2, \ldots, n.$$

One more comment: having both (H) and (H') amounts to assuming that

$$\mathrm{E}\, X_i X_j \leqslant \mathrm{E}\, Y_i Y_j, \quad i, j = 1, 2, \ldots, n \ \text{ and } \ \mathrm{E}\, Y_i^2 = \mathrm{E}\, X_i^2, \quad i = 1, 2, \ldots, n; \quad \text{so,}$$

if we just care about Slepian's lemma, we would start the same proof, but in the study of Taylor's expansion (74) we won't go through differences and Lemma 7, we will directly compare the expectations of $\sum_{i,j} \frac{\partial^2 K}{\partial x_i \partial x_j} Y_i Y_j$ and $\sum_{i,j} \frac{\partial^2 K}{\partial x_i \partial x_j} X_i X_j$.

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 metric 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) 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. 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 $n \geqslant 2$.

(4) 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:

$$4 \frac{d}{d\alpha} \bigl[\{\sup_{t \in T} Z_\alpha(t)\}\bigr] = \sum_{\substack{s,t \in T \times T \\ s \neq t}} \frac{d}{d\alpha} [\Delta_{Z_\alpha}(s,t)] \int \frac{dx}{dx_s \, dx_t} \int g_\alpha(x) \, du,$$

only adding 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$_3$].

(5) 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

$$A_\varepsilon = \{x \in S^{n-1} : \mathrm{dist}(x, A) < \varepsilon\}$$

is taken with respect to the geodesic metric 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.

(6) The *packing problem* for solid balls of a given radius consists in 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, considering *open* covering balls and δ-separation defined by $d(s, t) \geqslant \delta$.

(7) 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 $\mathbb{R}^n$.

(8) We are considering the Gaussian random variables X_t that are defined on $(\mathbb{R}^n, \gamma_n)$ by $X_t(\omega) = \omega \cdot t$, for ω and t in $\mathbb{R}^n$. Since $\max_{1 \leqslant j \leqslant m} (\theta \cdot a_j) = \max_{t \in K} (\theta \cdot t)$ because we have $K = \operatorname{conv}(a_1, \ldots, a_m)$, we did actually check at (19) that

$$\rho_u(\theta) = 1 + u \max_{t \in K} X_t(\theta) + O(u^2),$$

so it won't be too surprising that $h_1(K)$, that is proportional to the first order in u of the volume of $B_n + uK$, will relate to $\mathrm{E}\big(\max_{t \in K} X_t\big)$.

(9) We say *smallest* in order to have σ defined everywhere. Actually, if $X_1, X_2, \ldots, X_m$ are merely *distinct*, the "maximal index" σ is almost surely unique, and more precisely, one has $X_i \neq X_j$ almost surely when $i \neq j$, simply because $X_i - X_j$ is then a non-zero Gaussian random variable.

(10) If we translate K by some b, replacing each a_i by $a_i + b$, hence each Gaussian variable X_i by $X_i + Y$ with $Y(x) = x \cdot b$, then $h_2(K + b) = h_2(K)$ will not change but

$$\mathrm{E}\Big(\sup_i (X_i + Y)\Big)^2 = \mathrm{E}\Big(Y + \sup_i X_i\Big)^2 = \mathrm{E}\Big(\sup_i X_i\Big)^2 + 2\,\mathrm{E}\Big(Y \sup_i X_i\Big) + \mathrm{E}\, Y^2$$

will change in general. Fernique in [Fer$_4$] pointed out that the quantity $\mathrm{E}\big(X_\sigma^2 - \gamma(\sigma, \sigma)\big)$ is invariant under translation, and he gave geometrical interpretations of that quantity when $\dim K = 2$ or 3.

(11) 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 [Tal$_2$], 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 4 are set in place, the proof of the Fernique theorem reduces to an almost mechanical tree-manipulation.

(12) There is nothing magical about this cube N^3. We could rewrite everything with another power N^α, as long as $\alpha > 1$. For example, the exponent $3/2$ in Lemma 5 would be then replaced by the exponent $\alpha' = \alpha/(\alpha - 1)$ conjugate to α.

(13) We have a good reason for not restricting the discussion to the attainable subset $\widehat{\mathcal{X}}$. While it will be very easy to describe the set of states $\mathcal{X}$ and the origin $\hat{\mathrm{x}}$ that we use later below, it would be very difficult if not impossible to describe explicitely the subset $\widehat{\mathcal{X}}$ (and it would be useless, on top of that).

(14) In the set $\mathcal{X} \setminus \widehat{\mathcal{X}}$ of non-attainable places, we could have a *loop* $(x_0, x_1, \ldots, x_k)$ with $k > 0$, $N(x_j) = 1$ for $0 \leqslant j \leqslant k$ and $P(x_j) = \{x_{j+1}\}$ for $0 \leqslant j < k$, $P(x_k) = \{x_0\}$.

(15) 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 $\exp(c 3^i)$ when $i \geqslant 1$, where we can set $c = (\ln 2)/3$, according to the fact that $N_1 \geqslant 2$.

(16) This follows easily from the well-known *König's infinity lemma*, 1927, an easy but logically interesting result.

(17) 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 generates a group, but we will ignore it completely.

(18) If T has no isolated point, all maximal paths 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.

(19) One could think of defining places x simply as couples $x = (B, n)$ where B is an open ball $B(t, r)$ in T, but it would not be satisfactory when t is an isolated point in T, for which $B(t, r) = \{t\}$ for small $r > 0$. Then, knowing only the ball $B = B(t, r)$, we would lose the information about r, that we need in our arguments.

(20) We may encounter degenerate cases $x = (t, r, n)$ 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 role in the further discussion.

(21) Remember that we decide to set $r_{-1} = 3r_0$. In the construction of the promenade 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.

(22) When T is finite, we have of course $t_i(\tau) = \tau$ when $i \geqslant i_0$. Then the successive places in a path have the form

$$x_j = (\tau, r_j, n_j), \quad \text{when} \quad j \geqslant i_0,$$

and are distinguished by the values of the radii r_j (also by the values of the n_j). Our "unnatural" treatment was meant to avoid considering the finite case separately.

(23) 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 at the ith stage is comparable to the logarithm of the number N_i of next-places after a *single place* at the $(i-1)$th stage.

(24) Here, any fixed sequence of integers m_i with $\sum m_i^{-1} < \infty$ could be used in place of the values $m_i = N(x_i)$, that depend upon maximal paths $(x_i)_{i \geqslant 0}$ of the promenade.

(25) In the invariant case, we did not insist that the ball $B(t_{i+1}, r_{i+1})$ associated to a next-place x_{i+1} after $x_i = (t_i, r_i, n_i)$ be *contained* in the ball $B(t_i, r_i)$ associated to the previous place x_i. But here, we need to know that for $k > 1$, the points in the regions V_{i+k} at the next stages 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 integer $i \geqslant 0$.

Also, we can see in the next equation that what we really need is not so much the fact that $\varphi(t_j, r_i/12)$ remains *almost the same* when $j > i$ than the fact that it will not grow larger.

(26) What is artificial is that we are *coding* a node $x \in \mathcal{X}$, with several properties attached to it such as the region V, the radius r and more importantly the growth value n, by using the one-to-one choice of a *coding point* $t_x \in T$ meant to represent all those values at once.

([27]) Suppose that f is a mapping from a countable set X_0 to a metric space (T, d) with no isolated point, and $\rho > 0$ a positive function on X_0. It is easy to define a one-to-one mapping f_1 from X_0 to T such that $d(f_1(x), f(x)) < \rho(x)$ for every $x \in X_0$; indeed, if $(x_n)_{n \geqslant 0}$ is a list of the elements of X_0, we may select inductively the values of f_1 as follows: we let $f_1(x_0) = f(x_0)$; next, suppose that $f_1(x_j)$ was chosen for $0 \leqslant j \leqslant n$; since the set T has no isolated point, we know that the ball $B(f(x_{n+1}), \rho(x_{n+1}))$ is infinite: we can thus select a point $f_1(x_{n+1})$ in that ball, distinct from the preceding values $f_1(x_j)$, for all j such that $0 \leqslant j \leqslant n$.

([28]) For a promenade, it is not impossible that two or more different paths $\xi \in \Xi_i$ end with the same "sub-path" $(x_i, x_{i+1}) \in F_i$.

([29]) Now that the sets V at a given stage in the promenade are disjoint, we could define the places of the promenade simply as couples

$$(V, t), \quad \text{with} \quad t \in T,$$

that appear in the disjointification procedure. Indeed, if $\xi = (V_i, t_i)$ is such a "new place" at stage $i > 0$, there would be *unique* place (V_{i-1}, t_{i-1}) at stage $(i-1)$ such that V_{i-1} contains V_i, and we might retrieve without ambiguity the rule of the promenade.

([30]) For each $i \geqslant 0$, let X_i be the finite set of places at stage i in the promenade on $\mathcal{X}$, 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 maximal paths 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 next-place after 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 (place, next-place)-couples at stage j for the promenade on $\mathcal{X}$.

The set of paths $\mathbf{x}$ such that $x_i(\mathbf{x}) = \overline{x}_i$ for some fixed place $\overline{x}_i$ at stage i is both closed and open in $\mathbf{X}$. Hence, in the case of our set consisting of 4-tuples $x = (V, t, r, n)$ 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 the 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 $\mathbf{a_1^*}$—.

([31]) We have said that a promenade is more or less a Markov chain without probability. The simplest way to make a Markov chain out of a promenade is of course to assign the same probability $1/n$ to the n possible moves that can be achieved from a place x to its next-places $y \in P(x)$, where $n = |P(x)| \geqslant 1$.

([32]) 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 $(\mathcal{F}_i)$.

([33]) We did not give a *verbatim* statement in our Equation (52). The original result would have $b = 2$, and the bound $3\,b^k$ replaced by $(\ln 2)\,b^k$, so that $|T_k| < 2^{2^k}$.

([34]) 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 $0 < p < 1$.

([35]) Hence Pisier's theorem [Pis] 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 $L^{\Phi_2}(\mathbb{T}, m)$.*

As for the behaviour of L^p-norms, observe that the maximal value of $u^p\,\mathrm{e}^{-u}$ for $u \geqslant 0$ is achieved when $u = p$, hence $u^p \leqslant (p/\mathrm{e})^p\,\mathrm{e}^u$. Thus for any $c > 0$ and $p \geqslant 1/2$ we have

$$\int_S \bigl(c\,f(s)^2\bigr)^p\,\mathrm{d}\nu(s) \leqslant \Bigl(\frac{p}{\mathrm{e}}\Bigr)^p \int_S \exp\bigl(c\,f(s)^2\bigr)\,\mathrm{d}\nu(s).$$

If the integral on the right-hand side is $\leqslant 2$, we obtain that $\|f\|_{2p} \leqslant 2\,(p/[c\,\mathrm{e}])^{1/2}$. The opposite direction is obtained from the Taylor's series of the exponential function,

$$\int_S \exp\bigl(c\,f(s)^2\bigr)\,\mathrm{d}\nu(s) = 1 + \sum_{n=1}^{\infty} \frac{c^n\,\|f\|_{2n}^{2n}}{n!} \leqslant 1 + \sum_{n=1}^{\infty} \Bigl(2\,c\,\mathrm{e}\,\frac{\|f\|_{2n}^2}{2\,n}\Bigr)^n\,;$$

the series is convergent if $\|f\|_p \leqslant \sqrt{\theta/(2\,c\,\mathrm{e})}\,\sqrt{p}$ for some $\theta \in (0,1)$ and every $p \geqslant 2$, and has sum $\leqslant 1$ if $\theta \leqslant 1/2$. We used the easy fact that

$$n! = \int_0^{\infty} u^n\,\mathrm{e}^{-u}\,\mathrm{d}u \geqslant \int_n^{\infty} u^n\,\mathrm{e}^{-u}\,\mathrm{d}u \geqslant n^n \int_n^{\infty} \mathrm{e}^{-u}\,\mathrm{d}u = n^n\,\mathrm{e}^{-n}\,.$$

(36) For example, observe that $\ln 2 < 1 - 1/2 + 1/3 = 5/6$, hence $b > 1/\ln 2 > 6/5$; this implies that $7(b-1) > 1$ and $7b(b-1) > b > 1 > \ln 2$.

(37) Approximation by convolution is useful for dealing with more general examples, but for our specific function φ we might use the explicit C^∞ approximation

$$\Phi_\lambda(x) = \frac{1}{\lambda}\,\ln\Bigl(\sum_{i=1}^n \mathrm{e}^{\lambda x_i}\Bigr) \underset{\lambda \to +\infty}{\longrightarrow} \max(x_1, x_2, \ldots, x_n) = \varphi(x), \quad x \in \mathbb{R}^n.$$

One sees readily that Φ_λ satisfies $(**)$; when $\lambda > 0$ and $i \neq j$ we have

$$\frac{\partial^2 \Phi_\lambda}{\partial x_i \partial x_j}(x) = -\lambda\,\frac{\mathrm{e}^{\lambda x_i}\,\mathrm{e}^{\lambda x_j}}{\Bigl(\sum_{i=k}^n \mathrm{e}^{\lambda x_k}\Bigr)^2} < 0, \quad x \in \mathbb{R}^n.$$

Also, one can check that the partial derivatives of Φ_λ are bounded on $\mathbb{R}^n$.

More generally, any function $\Phi(x) = g\bigl(\sum_{i=1}^n f(x_i)\bigr)$ satisfies when $i \neq j$ that

$$\frac{\partial^2 \Phi}{\partial x_i \partial x_j}(x) = g''\Bigl(\sum_{i=1}^n f(x_i)\Bigr) f'(x_i) f'(x_j),$$

so that a Slepian-like comparison result can be expected between $\mathrm{E}\,\Phi(X)$ and $\mathrm{E}\,\Phi(Y)$ when f is monotone on the real line and g concave on the image interval $n\,f(\mathbb{R})$, and under the Slepian's assumptions for the centered Gaussian processes X and Y.

References

[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.

[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.

[Fer$_4$] X. Fernique, *Corps convexes et processus Gaussiens de petit rang*, Z. für Wahrscheinlichkeitstheorie **35** (1976), 349–353.

[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.

[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.

[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. Pełczyński, *Absolutely summing operators in L_p-spaces and their applications*, Studia Math. **29** (1968), 275–326.

[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.

[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] G. Pisier, *Ensembles de Sidon et processus gaussiens*, Comptes Rendus Acad. Sci. Paris **286** (1978), 671–674.

[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.

[Tal$_1$] M. Talagrand, *Regularity of Gaussian processes*, Acta Math. **159** (1987), 99–149.

[Tal$_2$] M. Talagrand, The generic chaining. Upper and lower bounds of stochastic processes, Springer Monogr. Math. Springer-Verlag, Berlin, 2005. viii+2222 pp.