Supremum of a Process in Terms of Trees

Olivier Guédon[1] and Artem Zvavitch[2]

[1] Equipe d'Analyse, Université Paris 6, 4 Place Jussieu, Case 186, 75005 Paris, France *guedon@ccr.jussieu.fr*

[2] Mathematics Department, Mathematical Sciences Bldg, University of Missouri, Columbia, MO 65211, USA *zvavitch@math.missouri.edu*

Summary. In this paper we study the quantity $\mathbb{E}\sup_{t\in T} X_t$, where X_t is some random process. In the case of the Gaussian process, there is a natural sub-metric d defined on T. We find an upper bound in terms of labelled-covering trees of (T,d) and a lower bound in terms of packing trees (this uses the knowledge of packing numbers of subsets of T). The two quantities are proved to be equivalent via a general result concerning packing trees and labelled-covering trees of a metric space. Instead of using the majorizing measure theory, all the results involve the language of entropy numbers. Part of the results can be extended to some more general processes which satisfy some concentration inequality.

1 Introduction

Let (T,d) be a compact metric space and for all $t \in T$, X_t be a collection of random variables such that $\mathbb{E}X_t = 0$. The aim of this paper is to present a different approach to the theory of majorizing measures. To avoid the problem of measurability of $\sup_{t\in T} X_t$, we take, as usual, the following definition:

$$\mathbb{E}\sup_{t\in T} X_t = \sup\left\{\mathbb{E}\sup_{t\in T_f} X_t,\ T_f \text{ finite subset of } T\right\}. \qquad (*)$$

It allows us to assume without loss of generality that (T,d) is in fact a finite metric space, which will make the presentation of the statements clearer. It means that in a general compact metric space (T,d), we take a very fine net on the set T to approach the quantity $\mathbb{E}\sup_{t\in T} X_t$. We want to present a new way to provide an estimate of this quantity where $(X_t)_{t\in T}$ is in particular a Gaussian process. In this case, there is a natural sub-metric d defined on T by $d(s,t)^2 = \mathbb{E}|X_s - X_t|^2$ and of course, by taking a quotient, we can assume that d is a metric on T. We recall a result of Talagrand in terms of the majorizing measure

Theorem [T1]. *If T is a finite set, $(X_t)_{t\in T}$ is a Gaussian process with the natural sub-metric d associated, then, up to universal constants, $\mathbb{E}\sup_{t\in T} X_t$ is similar to the quantity*

$$\inf \sup_{t\in T} \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t,\varepsilon))}}\, d\varepsilon,$$

where the infimum is taken over all probability measures on T, and $B(t,\varepsilon) = \{s \in T, d(s,t) \le \varepsilon\}$.

One of the biggest problems is to provide a uniform approach for construction of a "good" measure.

The main tools of this paper are "packing" and "labelled-covering" trees. The idea to use these objects comes from works of Talagrand [T3], [T5]. In [T3] he defined the notion of an s-tree and he shows, using the majorizing measures technique, that an s-tree provides an estimate for $\mathbb{E}\sup_{t\in T} X_t$. This point of view has been very fruitful in the study of embeddings of subspace of L_p into ℓ_p^n for $0 < p < 1$ [Z]. Here we would like to present a geometrical method for providing bounds for the supremum of a process which satisfies a concentration type inequality, where instead of measures, we will consider special families of sets of our metric space T. The main idea is to present a straightforward technique which like the theorems of Dudley and Sudakov involves the language of entropy numbers.

There are two different sections in this paper. First, we present the notions of packing and labelled-covering trees and define how to measure the size of such trees. The main result of this part is a general comparison of these two quantities. The second part is devoted to the study of upper and lower bounds of (∗) when the process satisfies a concentration type inequality. We obtain an improvement of Dudley's result which gives directly, iterating this result, an upper bound in function of the size of labelled-covering trees of the compact metric space (T,d). For the lower bound, an additional hypothesis is a Sudakov type minoration of $\mathbb{E}\sup(X_{t_1},\ldots,X_{t_N})$ for well separated points in T. In this part, we consider for simplicity a particular case of the Gaussian process but the spirit of this idea allows generalization when the process satisfies other types of concentrations and other Sudakov type minorations [L], [T2]. We obtain an expression in terms of the size of packing trees and combining this with the result of the first part, it shows that in the Gaussian (or Euclidean) case, all these quantities are similar up to universal constants.

2 Trees of Sets

Consider a finite metric space (T,d).

Recall that a tree of subsets of T is a finite collection $\mathcal{F}$ of subsets of T with the property that for all $A,B \in \mathcal{F}$, either $A \cap B = \emptyset$, or $A \subset B$, or $B \subset A$. We say that B is a *son* of A if $B \subset A$, $B \ne A$ and

$$C \in \mathcal{F}, \quad B \subset C \subset A \quad \Longrightarrow \quad C = B \text{ or } C = A.$$

We assume that $\mathcal{A}_1$ consists of one single set (this is the root of $\mathcal{F}$) and that for each $k \in \mathbb{N}^\star$, $\mathcal{A}_{k+1}$ is a finite collection of subsets of T such that each of them is a son of a set in $\mathcal{A}_k$. A *branch* of $\mathcal{F}$ is a sequence $A_1 \supset A_2 \supset \ldots$

such that A_{k+1} is a son of A_k. A branch is *maximal* if it is not contained in a longer branch. To each $A \in \mathcal{F}$ we denote by $N(A)$ the number of sons of A.

Let $B_1, \dots, B_{N(A)}$ be sons of A. We denote by ℓ_A a one-to-one map

$$\ell_A : \{B_1, \dots, B_{N(A)}\} \to \{1, \dots, N(A)\}.$$

Consider some fixed number $r \geq 120$. A tree $\mathcal{F}$ is called a *packing tree* if to each $A \in \mathcal{F}$, we can associate an integer $n(A) \in \mathbb{Z}$ such that

1) for all sons B of A, $\operatorname{diam}(B) \leq 2r^{-n(A)}$,
2) if B and B' are two distinct sons of A then $d(B, B') \geq 30r^{-n(A)}$.

We define the *size* $\gamma_p(\mathcal{F}, d)$ *of a packing tree* $\mathcal{F}$ to be the infimum over all possible maximal branches of

$$\sum_{k \geq 1} r^{-n(A_k)} \sqrt{\log\left(N(A_k)\right)}.$$

A tree $\mathcal{F}$ is called a *labelled-covering tree* if

1) for any $t \in T$ there is a maximal branch $A_1 \supset A_2 \supset \dots$ such that $t = \bigcap A_k$,
2) to each $A \in \mathcal{F}$ is associated a labelled function ℓ_A (which numerates each son of A) and an integer $n(A) \in \mathbb{Z}$ such that $\operatorname{radius}(A) \leq r^{-n(A)}$ (we allow $n(A) = +\infty$ when the set A is a single point).

Finally we define the *size* $\gamma_c(\mathcal{F}, d)$ *of a labelled-covering tree* $\mathcal{F}$ as the supremum over all possible maximal branches of

$$\sum_{k \geq 1} r^{-n(A_k)} \sqrt{\log\left(e\ell_{A_k}(A_{k+1})\right)}.$$

We denote by $Cov(T, d)$ (respectively, $Pac(T, d)$) the set of all labelled-covering (respectively, packing) trees in T. The first theorem shows a connection between the definitions of size of packing trees and of labelled-covering trees.

Theorem 1. *There exists a constant $C > 1$ such that for any finite metric space (T, d)*

$$\inf_{\mathcal{F} \in Cov(T,d)} \gamma_c(\mathcal{F}, d) \leq C \sup_{\mathcal{F} \in Pac(T,d)} \gamma_p(\mathcal{F}, d).$$

To prove it, we will use the following theorem due to Talagrand.

Theorem [T5]. *Consider a finite metric space (T, d) and the largest $i \in \mathbb{Z}$ such that* $\operatorname{radius}(T) \leq r^{-i}$. *Assume that for $j \geq i$ there are functions $\phi_j : T \to \mathbb{R}^+$ with the following property:*

For any point s of T, any integer $j \geq i$ and $N \geq 1$, if $t_1, \dots t_N$ are N points in $B(s, r^{-j})$ such that

$$d(t_{l'}, t_l) \geq r^{-j-1}, \text{ for any } l, l' \leq N, l \neq l',$$

then we have

$$\phi_j(s) \geq \alpha r^{-j} \sqrt{\log N} + \min_{l \leq N} \phi_{j+2}(t_l). \tag{1}$$

Assume also that $(\phi_j)_{j \geq i}$ is a decreasing sequence of functions. Then

$$\inf_{\mathcal{F} \in Cov(T,d)} \gamma_c(\mathcal{F}, d) \leq \frac{5}{\alpha} \sup_{t \in T} \phi_i(t).$$

For completeness of the paper, we reproduce here a proof of this result which is almost the proof of Proposition 4.3 of [T5].

Proof. Our goal is to construct a labelled-covering tree $\mathcal{F}$ such that

$$\sum_{k \geq 1} r^{-n(A_k)} \sqrt{\log \left(e\ell_{A_k}(A_{k+1}) \right)} \leq C \sup_{t \in T} \phi_i(t),$$

for any branch $\{A_1 \supset \dots \supset A_k \supset \dots\}$ in $\mathcal{F}$.

We will inductively construct our covering tree.

First step: $k = 1$.
The first step consists of taking $A_1 = T$, $n(A_1) = n(T) = i$ and we define $a_1(A_1) \in A_1$ such that

$$A_1 \subset B\big(a_1(A_1), r^{-i}\big).$$

Iterative step: from k to $k+1$.
Assume that we have constructed the k^{th} level $\mathcal{A}_k$ of the tree $\mathcal{F}$ (which is a covering of the set T) such that

1) $T = A_k^1 \cup \dots \cup A_k^d$,
2) for each set A_k of this covering, either A_k is a single point or there exists $a_k(A_k) \in A_k$ such that $A_k \subset B(a_k(A_k), r^{-n(A_k)})$ with the biggest possible integer $n(A_k)$.

If all the sets of this covering consist of single points then the construction is finished (and this situation will appear because T is a finite set). Now we show how to partition any given element A_k of this covering. If A_k is a single point then $n(A_k) = +\infty$ and $A_1 \supset \dots \supset A_k$ is a maximal branch so we have nothing to do. Assume now that A_k is not a single point.
We pick $t_1 \in A_k$ such that

$$\phi_{n(A_k)+2}(t_1) = \max \big\{ \phi_{n(A_k)+2}(t); t \in A_k \big\}.$$

Then the first son of A_k is

$$B_1 = A_k \cap B(t_1, r^{-n(A_k)-1})$$

and $a_{k+1}(B_1) = t_1$. We define $n(B_1)$ as the biggest integer such that $B_1 \subset B(t_1, r^{-n(B_1)})$. To construct B_2 we repeat this procedure, replacing A_k by $A_k \setminus B_1$. This set is not empty because $r > 2$, A_k is not a single point and by the maximum condition on $n(A_k)$.

Finally we have constructed points $t_1, \ldots, t_N$ ($N \geq 2$) and sons $B_1, \ldots, B_N$ such that for any $m \in \{1, \ldots, N\}$,

$$t_m \in A_k \setminus \bigcup_{l<m} B(t_l, r^{-n(A_k)-1})$$

and

$$\phi_{n(A_k)+2}(t_m) = \max \Big\{ \phi_{n(A_k)+2}(t); t \in A_k \setminus \bigcup_{l<m} B\big(t_l, r^{-n(A_k)-1}\big) \Big\}.$$

It is clear (by construction) that $B_1, \ldots, B_N$ are sons of A_k, form a covering of A_k and that $n(B_m) \geq n(A_k) + 1$. Also by construction $d(t_l, t_{l'}) \geq r^{-n(A_k)-1}$, $d(a_k(A_k), t_m) \leq r^{-n(A_k)}$ and taking $j = n(A_k)$, we obtain by definition of our functions ϕ_j that for any $m \in \{1, \ldots, N\}$,

$$\phi_{n(A_k)}\big(a_k(A)\big) \geq \alpha r^{-n(A_k)} \sqrt{\log m} + \min_{l \leq m} \phi_{n(A_k)+2}(t_l).$$

We labelled the sons by setting $\ell_A(B_m) = m$ so

$$\phi_{n(A_k)}\big(a_k(A)\big) \geq \alpha r^{-n(A_k)} \sqrt{\log \ell_A(B_m)} + \min_{l \leq m} \phi_{n(A_k)+2}(t_l).$$

By construction of the points $\{t_l\}$, if $l < l'$,

$$\phi_{n(A_k)+2}(t_l) \geq \phi_{n(A_k)+2}(t_{l'}),$$

so we get

$$\min_{l \leq m} \phi_{n(A_k)+2}(t_l) \geq \phi_{n(A_k)+2}(t_m).$$

At this stage, for each set A_k of our starting covering of T, we have constructed a labelled function ℓ_{A_k}, sons who form a covering of A_k such that for all sons A_{k+1} of A_k, $n(A_{k+1}) \geq n(A_k) + 1$, and point $a_{k+1}(A_{k+1})$ such that

$$\phi_{n(A_k)}\big(a_k(A_k)\big) \geq \alpha r^{-n(A_k)} \sqrt{\log \ell_{A_k}(A_{k+1})} + \phi_{n(A_k)+2}\big(a_{k+1}(A_{k+1})\big).$$

Next we observe that of course, for all sons A_{k+2} of A_{k+1}, $a_{k+2}(A_{k+2}) \in A_{k+1}$ so by construction of $a_{k+1}(A_{k+1})$,

$$\phi_{n(A_k)+2}(a_{k+1}(A_{k+1})) \geq \phi_{n(A_k)+2}\big(a_{k+2}(A_{k+2})\big).$$

But $n(A_{k+2}) \geq n(A_{k+1}) + 1 \geq n(A_k) + 2$ (by construction) and as $(\phi_j)_{j\geq i}$ is a decreasing sequence of functions,

$$\phi_{n(A_k)+2}\big(a_{k+2}(A_{k+2})\big) \geq \phi_{n(A_{k+2})}\big(a_{k+2}(A_{k+2})\big),$$

and finally, for all branches $A_k \supset A_{k+1} \supset A_{k+2}$,

$$\phi_{n(A_k)}\big(a_k(A_k)\big) \geq \alpha r^{-n(A_k)}\sqrt{\log \ell_{A_k}(A_{k+1})} + \phi_{n(A_{k+2})}\big(a_{k+2}(A_{k+2})\big).$$

Conclusion.
If we sum up the last inequality for $k \geq 1$, we get

$$\phi_{n(A_1)}\big(a_1(A_1)\big) + \phi_{n(A_2)}\big(a_2(A_2)\big) \geq \alpha \sum_{k\geq 1} r^{-n(A_k)}\sqrt{\log \ell_{A_k}(A_{k+1})}$$

which gives (because the sequence $(\phi_j)_{j\geq i}$ is decreasing), for all branches $A_1 \supset \ldots \supset A_k \supset \ldots$ of the labelled-covering tree $\mathcal{F}$

$$\alpha \sum_{k\geq 1} r^{-n(A_k)}\sqrt{\log \ell_{A_k}(A_{k+1})} \leq 2 \sup_{t\in T} \phi_i(t).$$

Now call

$$S_1 = \sup_{\text{maximal branch}} \sum_{k\geq 1} r^{-n(A_k)}\sqrt{\log \ell_{A_k}(A_{k+1})}$$

and

$$S_2 = \sup_{\text{maximal branch}} \sum_{k\geq 1} r^{-n(A_k)}\sqrt{\log e\ell_{A_k}(A_{k+1})}.$$

It is clear that $S_1 \geq r^{-n(A_1)}\sqrt{\log 2}$. By construction, for all sons A_{k+1} of A_k, $n(A_{k+1}) \geq n(A_k) + 1$ then for all maximal branch $A_1 \supset \ldots \supset A_k \supset \ldots$ of the labelled-covering tree $\mathcal{F}$,

$$\begin{aligned}\sum_{k\geq 1} r^{-n(A_k)}\sqrt{\log e\ell_{A_k}(A_{k+1})} &\leq \sum_{k\geq 1} r^{-n(A_k)}\Big(1 + \sqrt{\log \ell_{A_k}(A_{k+1})}\Big) \\ &\leq S_1 + \frac{r-1}{r} r^{-n(A_1)} \leq \frac{5}{2} S_1\end{aligned}$$

because r is large enough. It proves that for this tree $\mathcal{F}$,

$$S_2 \leq 5S_1/2 \leq 5 \sup_{t\in T} \phi_i(t)/\alpha.$$

□

Proof (of Theorem 1). Let i be the largest integer such that radius$(T) \leq r^{-i}$. For a set $A \subset T$, let $\gamma_p(A) = \sup_{\mathcal{F}\in Pac(A,d)} \gamma_p(\mathcal{F}, d)$, and for all integers $j \geq i$, define the function $\phi_j : T \to \mathbb{R}^+$ by

$$\forall s \in T,\ \phi_j(s) = \gamma_p\big(B(s, 2r^{-j})\big).$$

The sequence $(\phi_j)_{j\geq i}$ is decreasing and, by definition of $i \in \mathbb{Z}$,

$$\sup_{t\in T} \phi_i(t) = \gamma_p(T).$$

To prove Theorem 1, we need to check assumption (1) of the previous theorem. Fix some $j \geq i$ and $s \in T$. Let $t_1, \dots, t_N$ be points in $B(s, r^{-j})$ with $d(t_l, t_{l'}) \geq r^{-j-1}$, then

$$\phi_{j+2}(t_l) = \gamma_p\big(B(t_l, 2r^{-j-2})\big) \qquad \text{and} \qquad B(t_l, 2r^{-j-2}) \subset B(s, 2r^{-j})$$

and

$$d\big(B(t_l, 2r^{-j-2}), B(t_{l'}, 2r^{-j-2})\big) \geq r^{-j-1} - 4r^{-j-2} \geq \frac{1}{4} r^{-j-1}.$$

Consider in $B(s, 2r^{-j})$ a two level packing tree whose first level is $B(s, 2r^{-j})$ and whose second level consists of

$$\big\{B_l = B(t_l, 2r^{-j-2})\big\}_{l\leq N}.$$

Take $n(B(s, 2r^{-j})) = j+2$ then for each son B, B' of $B(s, 2r^{-j})$, $\mathrm{diam}(B) \leq 2r^{-n(B(s,2r^{-j}))}$ and

$$d(B, B') \geq \frac{1}{4} r^{-j-1} \geq 30\, r^{-n(B(s,2r^{-j}))} = \frac{30}{r}\, r^{-j-1}$$

because r is large enough ($r = 120$). By definition of the size of packing trees,

$$\gamma_p\big(B(s, 2r^{-j})\big) \geq r^{-j-2}\sqrt{\log N} + \min_{l\leq N} \gamma_p\big(B(t_l, 2r^{-j-2})\big),$$

or

$$\phi_j(s) \geq \frac{1}{r^2}\, r^{-j}\sqrt{\log N} + \min_{l\leq N} \phi_{j+2}(t_l),$$

so we can apply the previous theorem with $\alpha = 1/r^2$. $\square$

3 Application to Random Processes

Let (T, d) be a finite metric space, and for all $t \in T$, X_t be a collection of random variables such that $\mathbb{E}X_t = 0$. In this part, we show how the quantities defined in the above sections are related to the study of $\mathbb{E}\sup_{t\in T} X_t$.

We will say that the process $(X_t)_{t\in T}$ satisfies a concentration inequality (H) if there exists $c > 0$ such that

$$\begin{cases} \text{for all subsets } A \subset T, \text{ for all } t_0 \in T, \\ \text{if } Y_{A,t_0} = \sup_{t\in A}(X_t - X_{t_0}) \text{ and } \sigma = \sup_{t\in A} d(t, t_0) \text{ then} \\ \forall u \geq 0, \mathbb{P}\big(|Y_{A,t_0} - \mathbb{E}Y_{A,t_0}| \geq u\big) \leq 2\exp\Big(-c\Big(\dfrac{u}{\sigma}\Big)^2\Big). \end{cases}$$

Remark. This hypothesis (H) implies a deviation inequality: for all $(s,t) \in T$,

$$\mathbb{P}(|X_s - X_t| \geq u) \leq 2\exp\left(-c\Big(\frac{u}{d(s,t)}\Big)^2\right).$$

Indeed, choose $s = t_0$ and $A = \{t\}$ then $\sigma = d(s,t)$ and $Y_{A,t_0} = X_t - X_{t_0}$ which gives the result. Maurey and Pisier ([P] Theorem 4.7) have proved that (H) is satisfied for the Gaussian process (with $c = 2/\pi^2$) and Talagrand [T4] proved it for the Bernoulli process. We don't know if it is true for a general subgaussian process, i.e. a process which satisfies only a deviation inequality as above.

3.1 Relation with the Size of Covering Trees

When the process $(X_t)_{t\in T}$ satisfies such a concentration inequality, we obtain an upper bound of $\mathbb{E}\sup_{t\in T} X_t$ in terms of the size of labelled-covering trees of T with respect to the metric d. The next result is an improvement of Lemma 3.4.4 in [Fe] which was the usual Dudley's upper bound.

Theorem 2. *If the process $(X_t)_{t\in T}$ satisfies a concentration inequality (H), there exists a constant $C_1 > 0$ (depending only on the constant c in (H)) such that for all $N \in \mathbb{N}^\star$, for all subsets $A_1, \dots, A_N$ of T, and $A = A_1 \cup \dots \cup A_N$, we have*

$$\mathbb{E}\sup_{t\in A} X_t \leq \sup_{1\leq \ell\leq N}\Big(C_1 \mathrm{diam} A\sqrt{\log e\ell} + \mathbb{E}\sup_{t\in A_\ell} X_t\Big).$$

Proof. Let $t_0 \in A$ then $\mathbb{E}\sup\limits_{t\in A} X_t = \mathbb{E}\sup\limits_{t\in A}(X_t - X_{t_0})$. For all $\ell \in \{1,\dots,N\}$, let $Y_\ell = \sup\limits_{t\in A_\ell}(X_t - X_{t_0})$ then

$$\mathbb{E}\sup_{t\in A} X_t = \mathbb{E}\sup_{1\leq\ell\leq N} Y_\ell.$$

Let S be defined by

$$S = \sup_{1\leq\ell\leq N}\Big(c_1 \mathrm{diam}(A)\sqrt{\log e\ell} + \mathbb{E}\sup_{t\in A_\ell} X_t\Big),$$

where c_1 will be defined later in accordance with the constant $c > 0$ in the hypothesis (H).
As $\sup_{1\leq\ell\leq N} Y_\ell$ is a non-negative random variable,

$$\begin{aligned}\mathbb{E}\sup_{1\leq\ell\leq N} Y_\ell &= \int_0^{+\infty} \mathbb{P}\big(\exists\, \ell \in \{1\dots,N\}, Y_\ell > u\big)du \\ &\leq K + \int_K^{+\infty} \mathbb{P}\big(\exists\, \ell \in \{1\dots,N\}, Y_\ell > u\big)du.\end{aligned}$$

By definition of S, for all $\ell \in \{1,\dots,N\}$,

$$S \geq \Big(c_1 \mathrm{diam}(A) \sqrt{\log e\ell} + \mathbb{E} Y_\ell \Big),$$

so by choosing $K = S$, we obtain

$$\begin{aligned}\mathbb{E} \sup_{1 \leq \ell \leq N} Y_\ell &\leq S + \int_S^{+\infty} \mathbb{P}\Big(\exists\, \ell, Y_\ell - \mathbb{E} Y_\ell > u - S + c_1 \mathrm{diam}(A) \sqrt{\log e\ell} \Big) du \\ &\leq S + \sum_{\ell=1}^{N} \int_0^{+\infty} \mathbb{P}\Big(Y_\ell - \mathbb{E} Y_\ell > u + c_1 \mathrm{diam}(A) \sqrt{\log e\ell} \Big) du.\end{aligned}$$

To conclude, we know that for all $t \in A_\ell$, $d(t, t_0) \leq \mathrm{diam}(A)$ and by the concentration inequality (H), we have

$$\begin{aligned}\mathbb{E} \sup_{1 \leq \ell \leq N} Y_\ell &\leq S + 2 \sum_{\ell=1}^{+\infty} \int_0^{+\infty} \exp\left(- c\Big(\frac{u}{\mathrm{diam}(A)} + c_1 \sqrt{\log e\ell} \Big)^2 \right) du \\ &\leq S + \sqrt{\frac{\pi}{c}}\, \mathrm{diam}(A) \sum_{\ell=1}^{+\infty} \exp\big(- c c_1^2 \log(e\ell) \big) \\ &\leq S + \frac{1}{e^2} \sqrt{\frac{\pi}{c}} \mathrm{diam}(A) \sum_{\ell=1}^{+\infty} \frac{1}{\ell^2},\end{aligned}$$

choosing c_1 such that $c c_1^2 = 2$. Because $\log e\ell \geq 1$, we have proved the theorem with

$$C_1 = \frac{1}{\sqrt{c}} \left(\sqrt{2} + \frac{\pi^{3/2}}{6e^2} \right).$$

□

Now, it is very easy to deduce the following result.

Corollary 3. *If the process $(X_t)_{t \in T}$ satisfies a concentration inequality (H) then there exists a constant $C > 1$ (depending only on the constant c in (H)) such that*

$$\mathbb{E} \sup_{t \in T} X_t \leq C \inf_{\mathcal{F} \in Cov(T,d)} \gamma_c(\mathcal{F}, d).$$

Proof. Let $\mathcal{F}$ be a labelled-covering tree of T with respect to the metric d. Then by Theorem 2, we deduce that

$$\mathbb{E} \sup_{t \in T} X_t \leq \sup_{A_i \text{ sons of } T} \Big(C_1\, \mathrm{diam} T \sqrt{\log e\ell_T(A_i)} + \mathbb{E} \sup_{t \in A_i} X_t \Big).$$

Now iterating this procedure over a particular son that realizes this maximum (it is finite because T is finite and note also that by the hypothesis on a labelled-covering tree, the last term of the sum will be $\mathbb{E} X_{t_i} = 0$ because the last sons must be a single point), we deduce that

$$\mathbb{E} \sup_{t \in T} X_t \leq C_1 \sup_{\text{maximal branch}} \sum_{k \geq 1} 2\, r^{-n(A_k)} \sqrt{\log e\ell_{A_k}(A_{k+1})}.$$

This is true for all labelled-covering trees so it gives exactly the stated result.
□

3.2 Relation with the Size of Packing Trees

To study a lower bound of $\mathbb{E}\sup_{t\in T} X_t$, we would like to start with the following theorem due to Talagrand [T5], which will lead us to the idea of how to bound $\mathbb{E}\sup_{t\in T} X_t$, where $(X_t)_{t\in T}$ is a Gaussian process, using packing trees.

Theorem [T5]. *Consider a Gaussian process $(X_t)_{t\in T}$, d the natural submetric associated and sets $\{B_l\}_{l\le N}$ with $N \ge 2$. Assume that $d(B_l, B_{l'}) \ge 15\,u$ for all integers $l, l' \le N, l \ne l'$ and $\operatorname{diam}(B_l) \le u$. Consider $A = \bigcup_{l\le N} B_l$, then*

$$\mathbb{E}\sup_{t\in A} X_t \ge C\,u\sqrt{\log N} + \min_{l\le N}\mathbb{E}\sup_{t\in B_l} X_t,$$

where $C = \pi/\sqrt{2} > 2$.

Proof. The proof of this theorem is based on the following two classical lemmas.

Lemma. *Under the assumptions of the previous theorem, let $t_l \in B_l$ and $Y_\ell = \sup_{t\in B_\ell}(X_t - X_{t_l})$ then*

$$\mathbb{E}\sup_{\ell\in\{1,\ldots,N\}} |Y_\ell - \mathbb{E}Y_\ell| \le u\,\frac{\pi}{\sqrt{2}}\,\sqrt{\log eN}.$$

Remark. This result was also used to obtain the classical Dudley upper bound in terms of entropy numbers [Fe] but is weaker than Theorem 2. As $\sup_{1\le\ell\le N}|Y_\ell - \mathbb{E}Y_\ell|$ is a non-negative random variable,

$$\begin{aligned}
\mathbb{E}\sup_{1\le\ell\le N}|Y_\ell - \mathbb{E}Y_\ell| &= \int_0^{+\infty}\mathbb{P}\big(\exists\,\ell\in\{1\ldots,N\}, |Y_\ell - \mathbb{E}Y_\ell| > t\big)dt \\
&\le K + \sum_{\ell=1}^{N}\int_K^{+\infty}\mathbb{P}\big(|Y_\ell - \mathbb{E}Y_\ell| > t\big)dt \\
&\le K + 2\sum_{\ell=1}^{N}\int_K^{+\infty}\exp\left(-c\Big(\frac{t}{u}\Big)^2\right)dt,
\end{aligned}$$

by the concentration inequality (H) and because $\operatorname{diam}B_l \le u$. The result follows choosing $K = u/\sqrt{c}\sqrt{\log N}$ (and recall that in this case, we could take $c = 2/\pi^2$). □

The next result is a Sudakov type inequality. There are many methods to obtain this kind of inequality. For the Gaussian case, we could see it as an application of Slepian's lemma but there is another method which can be generalized to other processes in the paper of Talagrand [T2] and in the paper of Latała [L].

Lemma. *If $t_1, \dots, t_N$ $(N \geq 2)$ are well separated points in T, i.e. assume that there exists $u > 0$ such that for all $l \neq l'$, $d(t_l, t_{l'}) \geq 15\,u$, then*

$$\mathbb{E} \sup_{i \in \{1,\dots,N\}} X_{t_i} \geq u\,\pi \sqrt{2 \log eN}.$$

Proof. Let $g_1, \dots, g_N$ be i.i.d. random normal Gaussian variables and define the process $Y_1, \dots, Y_N$ by $Y_i = \frac{15}{\sqrt{2}}\,u\,g_i$. Then it is clear that for all $l \neq l'$,

$$\mathbb{E}|X_{t_l} - X_{t_{l'}}|^2 = d(t_l, t_{l'})^2 \geq \mathbb{E}|Y_l - Y_{l'}|^2.$$

As

$$\mathbb{E} \sup(g_1, \dots, g_N) \geq \sqrt{\frac{\log N}{\pi \log 2}} \geq \sqrt{\frac{\log eN}{\pi \log 2e}}$$

for $N \geq 2$ (see for example formula 1.7.1 in [Fe]), the result follows easily by an application of Slepian's comparison property. □

Combining these two lemmas, it is very easy to finish the proof of the previous theorem.

$$\begin{aligned}
\mathbb{E} \sup_{t \in A} X_t &= \mathbb{E} \sup_{l \leq N} (Y_l - \mathbb{E}Y_l) + \mathbb{E}Y_l + X_{t_l} \\
&\geq \min_{l \leq N} \mathbb{E}Y_l + \mathbb{E} \sup(X_{t_1}, \dots, X_{t_N}) - \mathbb{E} \sup_{l \leq N} |Y_l - \mathbb{E}Y_l| \\
&\geq \frac{\pi}{\sqrt{2}} u \sqrt{\log N} + \min_{l \leq N} \mathbb{E} \sup_{t \in B_l} X_t.
\end{aligned}$$

□

Using this theorem we deduce the following corollary.

Corollary 4. *There is a universal constant $C > 0$ such that, if $(X_t)_{t \in T}$ is a Gaussian process and d the natural sub-metric associated, then*

$$C \sup_{\mathcal{F} \in Pac(T,d)} \gamma_p(\mathcal{F}, d) \leq \mathbb{E} \sup_{t \in T} X_t.$$

Proof. Let $\mathcal{F}$ be a packing tree of T with respect to the metric d. For any element A of this packing tree, let $B_1, \dots, B_{N(A)}$ be the sons of $A \in \mathcal{F}$. Then we use the previous lemma with $u = 2r^{-n(A)}$ (because $d(B_l, B_{l'}) \geq 15u$ and $\operatorname{diam}(B_l) \leq 2r^{-n(A)} \leq u$) to get

$$\mathbb{E} \sup_{t \in A} X_t \geq C r^{-n(A)} \sqrt{\log N(A)} + \min_{l \leq N(A)} \mathbb{E} \sup_{t \in B_l} X_t.$$

Now iterate this formula over a particular son which realizes this minimum to deduce that

$$\mathbb{E} \sup_{t \in T} X_t \geq C \inf_{\text{maximal branch}} \sum_{k \geq 1} r^{-n(A_k)} \sqrt{\log\big(N(A_k)\big)}.$$

This is true for all packing trees and it finishes the proof. □

To conclude this part, we just want to state the result we can deduce from Theorem 1, Corollary 3 and Corollary 4 in the case of the Gaussian process.

Theorem 5. *Let T be a finite set, $(X_t)_{t\in T}$ a Gaussian process with $\mathbb{E}X_t = 0$ and d the natural sub-metric associated, then, up to universal constants, the three quantities*

$$\mathbb{E}\sup_{t\in T} X_t, \quad \inf_{\mathcal{F}\in Cov(T,d)} \gamma_c(\mathcal{F}, d) \quad \text{and} \quad \sup_{\mathcal{F}\in Pac(T,d)} \gamma_p(\mathcal{F}, d)$$

are similar.

References

[D] Dudley, R.M. (1967): The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J. Funct. Anal., **1**, 290–330

[Fe] Fernique, X. (1997): Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens. Publications du Centre de Recherches Mathématiques, Montréal

[L] Latała, R. (1997): Sudakov minoration principle and supremum of some processes. Geom. Funct. Anal., **7**, 936–953

[P] Pisier, G. (1989): The Volume of Convex Bodies and Banach Space Geometry. Cambridge Tracts in Math., **94**, Cambridge University Press, Cambridge

[T1] Talagrand, M. (1987): Regularity of Gaussian processes. Acta Math., **159**, 99–149

[T2] Talagrand, M. (1994): The supremum of some canonical processes. American Journal of Mathematics, **116**, 283–325

[T3] Talagrand, M. (1995): Embedding subspaces of L_p in L_p^N. Geometric Aspects of Functional Analysis (Israel, 1992-1994), Oper. Theory Adv. Appl., **77**, Birkhaüser, Basel, 311–325

[T4] Talagrand, M. (1995): Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l'I.H.E.S., **81**, 73–205

[T5] Talagrand, M. (1996): Majorizing measures: the generic chaining. Annals of Probability, **24(3)**, 1049–1103

[Z] Zvavitch, A. (2000): More on embedding subspaces of L_p into ℓ_p^N, $0 < p < 1$. GAFA Seminar 1996-2000, Lecture Notes in Math., **1745**, 269–281