Talagrand's ellipsoid construction

1 Gaussian processes indexed by ellipsoids

S:gaussian

The set $\ell^2 := \{s \in \mathbb{R}^{\mathbb{N}} : \sum_{n\in\mathbb{N}} s_n^2\}$ is a separable HILBERT space under the inner product $\langle s,t\rangle = \sum_{n\in\mathbb{N}} s_n t_n$, with corresponding norm $\|s\|_2 = \left(\sum_{n\in\mathbb{N}} s_n^2\right)^{1/2}$. If $\{g_n : n \in \mathbb{N}\}$ is a sequence of independent $N(0,1)$ random variables then

$$X_t := \sum\nolimits_n t_n g_n \qquad \text{for } t \in \ell^2$$

converges in both the almost sure and L^2 senses. The collection $\{X_t : t \in \ell^2\}$ is a centered gaussian process with

$$\operatorname{cov}(X_s, X_t) = \sum\nolimits_\alpha s_\alpha t_\alpha = \langle s,t\rangle \qquad \text{for } s,t \in \ell^2.$$

Without loss of generality we may assume that the process is DOOB-SEPARABLE.

Talagrand (2021, pp. 73–80) considered the case where T is an ellipsoid, that is,

$$T = \mathcal{E}[a] := \{t \in \ell^2 : \sum\nolimits_n (t_n/a_n)^2 \le 1\} \qquad \text{for some fixed } a \text{ in } \ell^2.$$

The set T is a compact, convex subset of ℓ^2. Talagrand noted (Prop. 2.13.1) that $\mathbb{P}\sup_{t\in T} X_t$ is of order $\|a\|_2$. His argument can be slightly simplified.

eea <1> **Lemma.** *For* $T := \mathcal{E}[a]$,

$$c_0 \|a\|_2 \le \mathbb{P}\sup_t X_t \le \|a\|_2 \qquad \text{where } c_0 := \sqrt{2/\pi} = \mathbb{P}|g_1|.$$

Proof. Rescale X_t as $\langle u, W\rangle$ where $W_i := a_i g_i$ and $u_i := t_i/a_i$, so that

$$\sup_{t\in T} X_t = \sup\{\langle u, W\rangle : \|u\|_2 \le 1\}.$$

By CAUCHY-SCHWARZ, $\langle u, W\rangle \le \|u\|_2 \|W\|_2 = \|W\|_2$ with equality when $u_i = W_i/\|W\|_2$. Thus $\sup_{t\in T} X_t = \|W\|_2$ and

$$\mathbb{P}\sup_{t\in T} X_t = \mathbb{P}\|W\|_2 \le \sqrt{\mathbb{P}\|W\|_2^2} = \|a\|_2 .$$

For the lower bound, let S denote the set of sequences $u = (u_1, u_2, \dots)$ with $u_i = \pm a_i/\|a\|_2$. Each such u has $\|u\|_2^2 = \sum_i a_i^2/\|a\|_2^2 = 1$ and

$$\sup_{u\in S}\langle u, W\rangle = \sup_{u\in A}\sum\nolimits_i \pm a_i^2 g_i/\|a\|_2 = \sum\nolimits_i a_i^2|g_i|/\|a\|_2 ,$$

with expected value $\sum_i a_i^2 c_0/\|a\|_2 = c_0\|a\|_2$. □

Talagrand (2021, pp. 73–80) showed how to construct subsets $T_n \subset \ell^2$ with

$$U_0 = \{0\} \text{ and } |U_n| \le N_n := 2^{2^n} \text{ and}$$

\E@ Tal2.6.1 <2>
$$\sup_{t\in T}\sum\nolimits_{n\ge 0} 2^{n/2} d(t, U_n) \le C_1 \|a\|_2 ,$$

for some constant C_1.

> **Remark.** It suffices to construct the U_n's so that $\log_2|U_n| \le 2^{n+c}|$ for some universal constant c. Then work with the approximating sets $\widetilde{U}_n = \{0\}$ for $n \le m$ and $\widetilde{U}_n = U_{n+m}$ for some integer m greater than c. The subset $\widetilde{U}_n$ could then be replaced by $\{\pi(s) : s \in \widetilde{U}_n\}$, where π denotes the map that takes each s in ℓ^2 to its closest point in T. For each s in $\widetilde{U}_n$ and t in T we have $\|s-t\|_2 \ge \|s-\pi(s)\|_2$, which implies
>
> $$\|t-\pi(s)\|_2 \le \|t-s\|_2 + \|s-\pi(s)\|_2 \le 2\|t-s\|_2 .$$

2 Reduction to blocks

S:blocks

Without loss of generality we can assume that $a_i \searrow 0$. For the purpose of controlling the size of the approximating set U_n we will need to partition $\mathbb{N}$ into disjoint blocks $E_1, E_2, \dots$. My choice of blocks will differ slightly from the blocks used by Talagrand. For each positive integer n define

$$\begin{aligned} E_n &:= \{i \in \mathbb{N} : \alpha_n \le i < \alpha_{n+1}\} \qquad \text{where } \alpha_n := 2^{n-1}, \\ J_n &:= \{i \in \mathbb{N} : 1 \le i < \alpha_{n+1}\} = \{i \in \mathbb{N} : i < 2^n\}, \\ A_n &= a[\alpha_n]. \end{aligned}$$

Then we have $|E_n| = 2^{n-1}$ and $|J_n| = 2^n - 1$. The reason for this particular choice of the α_n's will only become evident when bounds involving those constants are studied. The following calculation is an example.

First note that $\|a\|^2 = \sum_{n\in\mathbb{N}} \sum_{i\in E_n} a_i^2$ and, by monotonicity of the $(a_i : i \in \mathbb{N})$ sequence we have $A_n \ge a_i \ge A_{n+1}$ for $i \in E_n$. Hence

$$\begin{aligned} \text{UPPER} := \sum\nolimits_{n\in\mathbb{N}} |E_n| A_n^2 \ge \|a\|^2 &\ge \sum\nolimits_{n\in\mathbb{N}} |E_n| A_{n+1}^2 \\ &= \tfrac{1}{2}\left(\text{UPPER} - |E_1| A_1^2\right). \end{aligned}$$

From the fact that $|E_1|A_1^2 = a_1^2 \le \|a\|^2$ it then follows that

\E@ norm.bnds <3> $$3\|a\|^2 \ge \sum\nolimits_{n\in\mathbb{N}} |E_n| A_n^2 \ge \|a\|^2.$$

Remark. This form of "condensation argument" is often attributed (Rudin, 1976, Theorem 3.27) to Cauchy.

The argument leading to <2> works separately for each t in T, using a sequence (as described in Lemma <6>) of non-negative integers $p_0 = 0, p_1, p_2, \dots$ depending on t. The construction works a block at a time, starting from

$$R_0 := t = (\tau_1, \tau_2, \dots) \qquad \text{where } \tau_k := (t_i : i \in E_k) \in \mathbb{R}^{E_k}.$$

It helps to rescale the problem so that $\|a\|_2 = 1$. Then the fact that $\sum_i t_i^2/a_i^2 \le 1$ shows that not only do we have $t_i^2 \le a_i^2$ for each i but also

\E@ sum.tau2 <4> $$1 = \sum\nolimits_{i\in\mathbb{N}} a_i^2 \ge \sum\nolimits_{i\in\mathbb{N}} t_i^2 = \sum\nolimits_{n\in\mathbb{N}} |\tau_n|_2^2 \qquad \text{and}$$

\E@ sum.tau2/A2 <5> $$1 \ge \sum\nolimits_{i\in\mathbb{N}} t_i^2/a_i^2 \ge \sum\nolimits_{n\in\mathbb{N}} |\tau_n|_2^2/A_n^2.$$

The argument leading to <2> is recursive. The first step replaces $R_0[J_1] := (R_0[i] : i \in J_1\}$ using an x_1 in $\mathbb{R}^{J_1}$ such that the difference $r_1 := R_0[J_1] - x_1$ has $|r_1|_2 \le 2^{-p_1/2}$. The working vector is then

$$R_1 := (r_1, \tau_2, \tau_3, \dots).$$

The next step finds an x_2 in $\mathbb{R}^{J_2}$ such that the difference $r_2 := R_1[J_2] - x_2$ has $|r_2|_2 \le 2^{-p_2/2}$, with new working vector

$$R_2 := (r_2, \tau_3, \tau_4, \dots).$$

And so on.

Each x_k in $\mathbb{R}^{J_k}$ defines an element $\widetilde{x}_k := (x_k, 0, 0, \dots)$ of ℓ^2. For each t in T the challenge is to choose x_k from a subset V_k of $\mathbb{R}^{J_k}$ with $|V_k|$ suitably small size. The approximating set U_n then consists of all possible sums $\widetilde{x}_1 + \dots + \widetilde{x}_n$ with $\widetilde{x}_k \in V_k$ as t ranges over T.

Remark. The tildes are just to remind you that we need to identify the x_k from $\mathbb{R}^{J_k}$ with an element of ℓ^2. I was rather tempted to abuse notation by writing x_k instead of $\widetilde{x}_k$.

The next Lemma establishes the important properties of the p_n, which I initially write as $p[n,t]$'s to emphasize the dependence on t.

pn <6> **Lemma.** *For a fixed $t = (\tau_1, \tau_2, \dots)$ in $T := \mathcal{E}[a]$ with $\|a\|_2 = 1$ define*

$$\beta[n,t] := \sup\{k \in \mathbb{N}_0 : 2^{-k} \geq |\tau_n|_2^2\}$$

and then recursively define a sequence of nonnegative integers $p[n,t]$ by $p[0,t] := p_0 := 0$ and

$$p[n,t] := \min\Big(p[n-1,t] + 2, \beta[n,t]\Big) \qquad \text{for } n \geq 1.$$

Then:

(i) $p[n,t] \leq p[n-1,t] + 2$ for $n \geq 1$ so that $p[n,t] \leq 2n$ for each n and $t \in T$.

(ii) $2^{-p[n,t]} \geq 2^{-\beta[n,t]} \geq |\tau_n|_2^2$ for each n.

(iii) $\sum_{n\in\mathbb{N}_0} 2^{n/2 - p[n,t]/2} \leq c_1$ where c_1 is a universal constant.

Proof. As the argument works independently for each t let me drop the t from the notation, abbreviating $p[n,t]$ to p_n and so on. Also I'll write h_n for $|\tau_n|^2$.

Assertions (i) and (ii) come directly from the definition of p_n as a minimum of two quantities. Assertion (iii) is a bit more subtle.

Define $\mathbb{N}_\beta := \{n \in \mathbb{N} : p_n = \beta_n\}$. If $n \in \mathbb{N}_\beta$ then β_n is finite and

\E@ beta.finite <7>

$$2^{-p_n} = 2^{-\beta_n} \geq h_n > 2^{-\beta_n - 1} = 2^{-p_n - 1} \qquad \text{implying} \qquad 2h_n > 2^{-p_n}.$$

Consequently,

\E@ sum.NNbeta <8>

$$\sum\nolimits_{n\in\mathbb{N}_\beta} 2^{-p_n}/A_n^2 \leq \sum\nolimits_{n\in\mathbb{N}_\beta} 2h_n/A_n^2 \leq 2 \qquad \text{by inequality } <5>.$$

By CAUCHY-SCHWARZ and inequalities <3> and <8> we then have

\E@ n2p2 <9>

$$\Big(\sum\nolimits_{n\in\mathbb{N}_\beta} 2^{n/2 - p_n/2}\Big)^2 \leq \sum\nolimits_{n\in\mathbb{N}_\beta} \frac{2^{n-p_n}}{2^n A_n^2} \sum\nolimits_{n\in\mathbb{N}_\beta} 2^n A_n^2 \leq 6.$$

If $\mathbb{N}\backslash\mathbb{N}_\beta$ is nonempty then it must be union of disjoint stretches $I_1, I_2, \dots$ of consecutive integers. Suppose

$$I := \{k + j : 1 \leq j \leq \ell\} \subset \mathbb{N}_0\backslash\mathbb{N}_\beta$$

for some $k \in \{0\} \cup \mathbb{N}_\beta$ and some $\ell \in \mathbb{N}$. By assumption $p_n < \beta_n$, which forces $p_n = 2 + p_{n-1}$, for $n \in I$. It then follows that $p_{k+j} = p_k + 2j$ for $1 \leq j \leq \ell$ so that

$$\sum\nolimits_{n\in I} 2^{n/2 - p_n/2} = 2^{k/2 - p_k/2} \sum\nolimits_{j=1}^{\ell} 2^{-j/2} \leq c_0 2^{k/2 - p_k/2},$$

where $c_0 = (\sqrt{2} - 1)^{-1} \approx 2.5$. A similar inequality holds if $|I|$ is infinite. If $k = 0$ the upper bound equals c_0. If $k \in \mathbb{N}_\beta$ then $2^{k/2 - p_k/2}$ is one of the summands on the left-hand side of inequality <9>. Assertion (iii) follows. $\square$

3 The main construction: approximation for ellipsoids

S:approximation

Throughout this Section I assume $\|a\|_2 = 1$ so that by inequalities <4> and <5>,

$$\max\left(\sum\nolimits_{n\in\mathbb{N}} |\tau_n|_2^2, \sum\nolimits_{n\in\mathbb{N}} |\tau_n|_2^2/A_n^2\right) \le 1 \qquad \text{for each } t = (\tau_1, \tau_2, \dots) \text{ in } T.$$

Define $\delta(p) := 2^{-p/2}$ for $p = 0, 1, \dots$. From Lemma 6 with $p_n = p[n, t]$ we have $|\tau_n|_2 \le \delta(p_n)$ and $p_n \le p_{n-1} + 2$ for $n \ge 1$, implying

\E@ delpn

$$\delta(p_n) = 2^{-p_n/2} \ge 2^{-p_{n-1}-1} = \tfrac{1}{2}\delta(p_{n-1}). \qquad \langle 10\rangle$$

For each finite subset J of $\mathbb{N}$ and each non-negative integer p define

$$B(J, p) := \{y \in \mathbb{R}^J : |y|_2 \le 4\delta(p)\}.$$

the euclidean ball with center 0 and radius $4 \times 2^{-p/2}$. Let $V(J, p)$ be a $\delta(p)$ packing set for $B(J, p)$. From PTTM SECTION 10.4

\E@ ball.pack

$$4^{|J|} \le |V(J, p)| := \text{PACK}(\delta(p), B(J, p)) \le 12^{|J|}. \qquad \langle 11\rangle$$

Each point of $B(J, p)$ lies within distance $2^{-p/2}$ of $V(J, p)$. Thus there is a map $\Psi_{J,p} : B(n, p) \to V(J, p)$ for which

$$|y - \Psi_{J,p} y|_2 \le 2^{-p} \qquad \text{for each } y \text{ in } B(J, p).$$

Remark. Without loss of generality it can be assumed that $0 \in V(J, p)$ for all (J, p), which ensures that $\Psi_{J,p} 0 = 0$.

procedure APPROX(t)
  Initialize: $n \leftarrow 1$; $R_0 \leftarrow t$. % fixed $t = (\tau_1, \tau_2, \dots)$ in T
  while $R_{n-1} \ne 0$ **do**
    $y_n \leftarrow R_{n-1}[J_n]$.
    $x_n \leftarrow \Psi_{J_n,p_n}(y_n)$.
    $R_n \leftarrow (r_n, \tau_{n+1}, \tau_{n+2}, \dots)$ where $r_n := y_n - x_n$.
    Increment n by 1.
  end while
end procedure

It helped me to run through a few steps of the while loop.

For $n = 1$:

$$\begin{aligned} |y_1|_2 = |\tau_1|_2 &\le \delta(p_1) \text{ so that } y_1 \in B(J_1, p_1),\\ x_1 &= \Psi_{J_1,p_1} y_1 \in V(J_1, p_1) \quad \text{AND} \quad |r_1|_2 \le \delta(p_1)\\ R_1 &= (r_1, \tau_2, \tau_3, \dots). \end{aligned}$$

For $n = 2$**:**

$$
\begin{aligned}
|y_2|_2 = |(r_1, \tau_2)|_2 \le |r_1|_2 + |\tau_2|_2 \le \delta(p_1) + \delta(p_2) \le 4\delta(p_2) \\
\text{so that } y_2 \in B(J_2, p_2), \\
x_2 = \Psi^{-1}_{J_2,p_2} y_2 \in V_{J_2,p_2} \text{ and } |r_2|_2 \le \delta(p_2), \\
R_2 = (r_2, \tau_3, \tau_4, \dots).
\end{aligned}
$$

For $n = 3$**:**

$$
\begin{aligned}
|y_3|_2 = |(r_2, \tau_3)|_2 \le |r_2|_2 + |\tau_3|_2 \le \delta(p_2) + \delta(p_3) \le 4\delta(p_3) \\
\text{so that } y_3 \in B(J_3, p_3), \\
x_3 = \Psi^{-1}_{J_3,p_3} y_3 \in V_{J_3,p_3} \text{ and } |r_3|_2 \le \delta(p_3), \\
R_3 = (r_3, \tau_3, \tau_5, \dots).
\end{aligned}
$$

In general we have $y_n \in B(J_n, p_n)$ and $x_n \in V(J_n, p_n)$ with that $|r_n|_2 \le 4\delta(p_n)$ with

$$
\begin{aligned}
R_n &= (r_n, \tau_{n+1}, \tau_{n+2}, \dots) = (y_n - x_n, \tau_{n+1}, \tau_{n+2}, \dots) \\
R_{n-1} &= (r_{n-1}, \tau_n, \tau_{n+1}, \dots) = (y_n, \tau_{n+1}, \dots).
\end{aligned}
$$

so that $R_{n-1} - R_n = (x_n, 0, 0, \dots) =: \widetilde{x}_n$. It follows that

$$
t - R_n = \sum\nolimits_{k=1}^{n} (R_{n-1} - R_n) = \sum\nolimits_{k=1}^{n} \widetilde{x}_n =: u_n.
$$

The remainders tend to zero as n goes to infinity because

$$
\|R_n\|_2 \le |r_n|_2 + \sum\nolimits_{k>n} |\tau_k|_2 \le 4 \sum\nolimits_{k \ge n} \delta(p_k).
$$

Finally

$$
\begin{aligned}
\sum\nolimits_{n\in\mathbb{N}} 2^{n/2} d(t, U_n) &\le \sum\nolimits_{n\in\mathbb{N}} 2^{n/2} \|t - u_n\|_2 = \sum\nolimits_{n\in\mathbb{N}} 2^{n/2} \|R_n\|_2 \\
&\le 3 \sum\nolimits_{n\in\mathbb{N}} 2^{n/2} \sum\nolimits_{k\in\mathbb{N}} \{k \le n\} 2^{-p_k/2} \\
&= 3 \sum\nolimits_{k\in\mathbb{N}} 2^{-p_k/2} \sum\nolimits_{n\in\mathbb{N}} 2^{n/2} \{k \le n\} \\
&\le 3 \sum\nolimits_{k\in\mathbb{N}} 2^{-p_k/2} 2^{(k+1)/2} / \left(\sqrt{2} - 1\right),
\end{aligned}
$$

which by Lemma <6> is smaller than $3c_1\sqrt{2}/\left(\sqrt{2} - 1\right)$.

As t ranges over T we get different $\{p[n,t]\}$ sequences, but always in the range $1 \le p_k \le 2k$ and

$$
y_k \in V_k := \cup_{1 \le p \le 2k} V(J_k, p).
$$

By inequality <11>,

$$
|V_k| \le \sum\nolimits_{1\le p\le 2k} 12^{|J_k|} \le 2k \times 12^{2^k} \le 12^{2^{k+1}}.
$$

Similarly, $|U_n| \le \prod_{k=1}^n |V_k|$, implying

$$
\log_2 |U_n| \le \sum\nolimits_{k=1}^{n} 2^{k+C_1} \le 2^{k+C_2},
$$

for some universal constants C_1 and C_2.

S:packing

4 Why chaining with packing numbers can fail

If $a \in \ell^2$ then <3> gives $\sum_{n\in\mathbb{N}} 2^n A_n^2 < \infty$, where $A_n := a[2^{n-1}]$. However it does not imply finiteness of $\sum_{n\in\mathbb{N}} 2^{n/2}(A_n - A_{n+1})$, as shown by the example $a_i := \left(\sqrt{i}\log_2(2i)\right)^{-1}$ where $A_n/2 = 2^{-n/2}n^{-1}$.

For such an a the packing bound gives $\mathbb{P}_{t\in\mathcal{E}[a]}X_t \le \infty$.

Suppose F is an ϵ-packing set for $T := \mathcal{E}[a]$ with $|F| = N$. Define $B_k := \{s \in \mathbb{R}^{E_k} : |s|_2 \le A_k\}$. Consider the vector $t = \widetilde{s} = (0,\dots,0,s,0,\dots)$. By definition of a packing set, there exists an f in F for which

$$|s - f[E_k]|_2 \le \|t - f\|_2 \le \epsilon.$$

Consequently, the set of vectors $\{f[E_k] : f \in F\}$ is an ϵ-covering set for B_k. Inequality <11> gives $N \ge (A_k/\epsilon)^{|E_k|}$.

$$N \ge (A_k/\epsilon)^{|E_k|}.$$

In particular, it follows that

$$\log_2 \text{PACK}(\epsilon, T) \ge |E_k| = 2^{k-1} \qquad \text{if } \epsilon \le A_k/2$$

and hence

$$\int_{A_{k+1}//2}^{A_k/2} \sqrt{\log \text{PACK}(r,T)}\, dr \ge \text{const}.2^{k/2}(A_k - A_{k+1}) \ge \text{const}./k.$$

Sum over k to deduce that

$$\int_0^{\text{diam}(T)} \sqrt{\log \text{PACK}(r,T)}\, dr = +\infty$$

Remark. Check §10.4 to see if pack vs. cover makes a difference.

S:Problems

5 Problems

P:special.t

[1] Suppose $t \in \mathcal{E}[a]$ with $t_i = 0$ for $i \notin E_k$. Assume $0 \in V(J,p)$ for all (J,p), which ensures that $\Psi_{J,p}0 = 0$. Find the sequences $p[n,t]$ and the vectors R_n generated by procedure APPROX. (Note the dependence on k.)

References

Rudin.baby

Rudin, W. (1976). *Principles of Mathematical Analysis* (Third ed.). McGraw-Hill.

Talagrand2021MMbook

Talagrand, M. (2021). *Upper and Lower Bounds for Stochastic Processes: Decomposition Theorems* (Second ed.), Volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag.