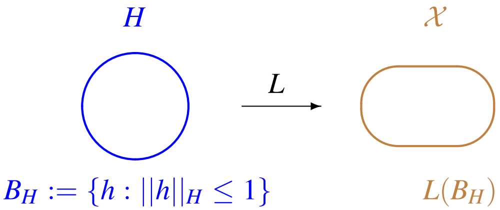
Linear operators, Metric Entropy, and Small Deviation Probabilities

M.A. Lifshits

January, 8
2021

A linear operator

Consider an operator $L : H \mapsto \mathcal{X}$ acting between two normed spaces

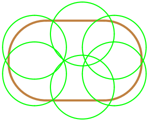


How to measure the "size" of operator L ?

The norm $||L||$ (half-diameter of $L(B_H)$) alone is not enough!

Covering numbers and entropy

We measure compactness of an operator $L : H \to \mathcal{X}$ by metric entropy.



Covering numbers:

$$N_L(\varepsilon) := \inf\left\{n : \exists\{x_j\}_{j\le n}, \{Lh : ||h||_H \le 1\} \subset \bigcup_{j=1}^{n} B_\varepsilon(x_j)\right\}.$$

Metric entropy: $\ln N_L(\varepsilon)$.
Dyadic entropy numbers:

$$e_n(L) := \inf\{\varepsilon > 0 : N_L(\varepsilon) \le 2^n\}.$$

Alternative measures of compactness

- $e_n(L)$ – dyadic entropy numbers;
- Kolmogorov numbers $d_n(L)$;
- linear approximation numbers $a_n(L)$;
- Gelfand numbers;
- Gaussian approximation numbers;

etc ...

Interesting circles of problems:

- compare different characteristics of L.
- duality problem: compare characteristics of L and L^*.

For us:

The main problem in operator language:

Find the behavior of covering numbers $N_L(\varepsilon)$, as $\varepsilon \to 0$.

An example: integration operator

Let $H = L_2[0,1]$, $\mathcal{X} = \mathbb{C}[0,1]$, and let $L : L_2[0,1] \to \mathbb{C}[0,1]$ be a simple integration operator

$$L(f)(t) := \int_0^t f(s)ds, \qquad f \in L_2[0,1].$$

Then $e_n(L) \approx n^{-1}$.
Another point of view on the same result. Define a Sobolev space

$$W_2^1[0,1] := \left\{ f : f \in AC,\, f(0) = 0,\, f' \in L_2[0,1] \right\}.$$

with the Hilbert norm $||f||_{2,1} := ||f'||_2$.
Let $I : f \mapsto f'$ be the natural isometry of Hilbert spaces $W_2^1[0,1]$ and $L_2[0,1]$.
Then the composition $L \circ I$

$$W_2^1[0,1] \overset{I}{\mapsto} L_2[0,1] \overset{L}{\mapsto} \mathbb{C}[0,1]$$

is the natural embedding of $W_2^1[0,1]$ into $\mathbb{C}[0,1]$ and for this embedding we also have $e_n(L \circ I) = e_n(L) \approx n^{-1}$.

An extension: fractional integration

Let $\alpha > 1/2$. Consider Riemann–Liouville fractional integration operator $L : L_2[0, 1] \to \mathbb{C}[0, 1]$ defined by

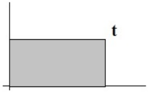
$$L^\alpha(f)(t) := \frac{1}{\Gamma(\alpha)} \int_0^t (t - s)^{\alpha-1} f(s) ds, \qquad f \in L_2[0, 1].$$

Notice that for $\alpha = 1$ we have again the former simple integration. Interestingly, there is a *semigroup* property: $L^\alpha \circ L^\beta = L^{\alpha+\beta}$.

Here we have $e_n (L^\alpha) \approx n^{-\alpha}$.

An open problem

Consider multivariate integration operator on $\mathbb{R}_+^d$.
For $t \in \mathbb{R}_+^d$ define the rectangle $[0,t] = \{s : 0 \le s_j \le t_j,\ 1 \le j \le d\}$.



Let the integration operator $L : L_2([0,1]^d) \mapsto \mathbb{C}([0,1]^d)$ be defined by

$$L(f)(t) := \int_{[0,t]} f(s) ds.$$

Problem:

Find the asymptotics for $e_n(L)$, as $n \to \infty$.

It is only known that

- $e_n(L) \approx n^{-1}$, for $d = 1$;
- $e_n(L) \approx n^{-1} (\ln n)^{3/2}$, for $d = 2$;
- $c_1 n^{-1} (\ln n)^{d-1} \le e_n(L) \le c_2 n^{-1} (\ln n)^{d-1/2}$, for general d.

Small deviations: a definition

Let $(\mathcal{X}, ||\cdot||)$ be a Banach space, for example, $\mathbb{C}[0,1]$, or $L_2[0,1]$, etc.
An $\mathcal{X}$-valued random vector X is understood as a measurable mapping

$$X : (\Omega, \mathbb{P}) \mapsto \mathcal{X}.$$

The small deviation problem (or small ball problem) suggests to explore

$$\mathbb{P}\left(||X|| \le \varepsilon\right), \qquad \varepsilon \to 0.$$

Let P denote the distribution of X, that is a measure in $\mathcal{X}$ given by $P(A) := \mathbb{P}(X \in A)$ and let $U := \{x \in \mathcal{X} : ||x|| \le 1\}$ be the unit ball in $\mathcal{X}$, then we have to study small ball measures

$$P\left(\varepsilon U\right), \qquad \varepsilon \to 0.$$

A typical answer to small deviation problem is

$$\mathbb{P}\left(||X|| \le \varepsilon\right) \sim c_1 \varepsilon^a \exp\{-c_2 \varepsilon^{-b}\}, \qquad \varepsilon \to 0.$$

Gaussian random vectors

Gaussian random vector extends the notion of a normally distributed random variable.
We call a random vector X taking values in a linear topological space $\mathcal{X}$ Gaussian if for every continuous linear functional $g \in \mathcal{X}^*$ the random variable $g(X)$ has a normal distribution.

The distribution of a Gaussian vector X is uniquely determined by the means and covariances of $\{g(X), g \in \mathcal{X}^*\}$.

Main example: Wiener process W – a random element in $\mathbb{C}[0, 1]$ or in $L_2[0, 1]$ such that $\mathbb{E}W(t) = 0$, $cov(W(s), W(t)) = \min\{s, t\}$.

Small deviations: an example

We study small deviation probabilities

$$\mathbb{P}\left(||X|| \le \varepsilon\right), \qquad \varepsilon \to 0.$$

for a random vector X taking values in a Banach space $(\mathcal{X}, ||\cdot||)$. Typically, $\mathcal{X}$ is a function space, and X is a sample path of a random process, such as Wiener process, fractional Brownian motion, Poisson process, etc.
For example, let $X = W$ be a Wiener process. Then

$$\mathbb{P}\left(\sup_{0\le t\le 1} |W(t)| \le \varepsilon\right) \sim \frac{4}{\pi}\ \exp\{-\frac{\pi^2}{8}\,\varepsilon^{-2}\}\,, \qquad \varepsilon \to 0,$$

and for $\mathcal{X} = L_2[0,1]$

$$\mathbb{P}\left(\int_0^1 |W(t)|^2 dt \le \varepsilon^2\right) \sim \frac{4\varepsilon}{\sqrt{\pi}}\ \exp\{-\frac{1}{8}\,\varepsilon^{-2}\}\,, \qquad \varepsilon \to 0.$$

Example: Gaussian vector in a Hilbert space

Let $\mathcal{X}$ be a separable Hilbert space with a basis (e_j). Then any element $x \in \mathcal{X}$ writes as $x = \sum_j x_j e_j$ with $\sum x_j^2 < \infty$. Let a random vector $X \in \mathcal{X}$ be given by

$$X := \sum_j \xi_j \, e_j, \qquad \xi_j \text{ are independent and } N(0, \sigma_j^2) - \text{distributed.}$$

If

$$\sigma_j^2 = C\left(j + \delta + O(j^{-1})\right)^{-\alpha}, \qquad \alpha > 1, \delta \in \mathbb{R},$$

then for small deviation probabilities we have (A.I.Nazarov et al)

$$\mathbb{P}\left(||X|| \le \varepsilon\right) \sim c_1 \, \varepsilon^a \exp\left\{-c_2 \, \varepsilon^{\frac{-2}{\alpha-1}}\right\}, \qquad \varepsilon \to 0.$$

where
$c_2 = c_2(C, \alpha, \delta)$, $a = \frac{2-\alpha-2\delta\alpha}{2(\alpha-1)}$, and c_1 depends on the entire sequence (σ_j).
Quite often, σ_j^2 are the eigenvalues of a Sturm–Liouville problem, thus small deviations are related to asymptotic behavior of these eigenvalues.

Merging two stories: operators and small deviations

If X is a centered Gaussian random vector in a separable Banach space $\mathcal{X}$, then it admits an expansion

$$X = \sum_j \xi_j L(e_j) \qquad \text{almost surely,}$$

where ξ_j are i.i.d. standard normal r.v. and $L : H \to \mathcal{X}$ an appropriate linear operator acting to $\mathcal{X}$ from a Hilbert space H with a basis (e_j).
We say then: the vector X and the operator L are *associated* .
If we replace the basis (e_j) with another basis (e'_j), the vector X will change to another vector X' but the distribution (hence small ball probabilities) will remain the same. Indeed, for every $f \in \mathcal{X}^*$,

$$\mathbb{E} f(X)^2 = \mathbb{E}\Big(\sum_j \xi_j f(Le_j)\Big)^2 = \sum_j f(Le_j)^2 = \sum_j (L^* f, e_j)_H^2 = ||L^* f||_H^2$$

is basis-invariant.
Therefore, any distribution property of a Gaussian vector can be expressed in terms of operator L.

An example of associated operator

Let $\mathcal{X} = \mathbb{C}[0,1]$, $X = W$ – a Wiener process, $H = L_2[0,1]$.
It turns out that an operator $L : L_2[0,1] \to \mathbb{C}[0,1]$ associated to Wiener process is just an integration operator

$$L(f)(t) := \int_0^t f(s)ds, \qquad f \in L_2[0,1].$$

Let us consider the cosine basis in $L_2[0,1]$ given by $e_0(s) := 1$ and

$$e_j(s) := \sqrt{2}\cos(\pi j s), \qquad j \geq 1.$$

Integration yields $Le_0(t) = t$ and

$$Le_j(t) = \sqrt{2}\,\frac{\sin(\pi j t)}{\pi j}, \qquad j \geq 1.$$

We arrive at the expansion

$$W(t) = \xi_0 t + \sqrt{2}\sum_{j=1}^{\infty} \xi_j\,\frac{\sin(\pi j t)}{\pi j}\ .$$

Entropy and Gaussian small deviations

We will concentrate our efforts on the exponential part of asymptotics. Define small deviation function by

$$\varphi(\varepsilon) := -\ln \mathbb{P}(||X|| \le \varepsilon).$$

Relations between $N_L(\varepsilon)$ and $\varphi(\varepsilon)$:

- polynomial growth
  Let $\beta \in (0,2)$. Then

  $$\ln N_L(\varepsilon) \approx \varepsilon^{-\beta} \Leftrightarrow \varphi(\varepsilon) \approx \varepsilon^{-\frac{2\beta}{2-\beta}}, \qquad \text{as } \varepsilon \to 0.$$

  Example: L – integration operator, X – Wiener process, $\beta = 1$.
- logarithmic growth
  Let $\beta > 0, \gamma \in \mathbb{R}$. Then

  $$\ln N_L(\varepsilon) \approx |\ln \varepsilon|^\beta \ln|\ln \varepsilon|^\gamma \Leftrightarrow \varphi(r) \approx |\ln \varepsilon|^\beta \ln|\ln \varepsilon|^\gamma, \qquad \text{as } \varepsilon \to 0.$$

General principle

The following properties are related:

- small deviation probabilities $\mathbb{P}(||X|| \leq \varepsilon)$ are not too small when $\varepsilon \to 0$.
- small deviation function $\varphi(\varepsilon) := -\ln \mathbb{P}(||X|| \leq \varepsilon)$ is growing slowly when $\varepsilon \to 0$.
- sample paths of a process X are rather smooth.
- X has good finite rank approximations

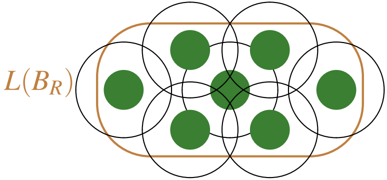
$$X \approx \sum_{j=1}^{n} \xi_j \, L(e_j), \qquad \text{as } n \to \infty.$$

How the connection occurs

We start with an operator $L : H \to \mathcal{X}$. Fix some R, ε. Take the image of the R-ball

$$L(B_R) = \{Lh, ||h||_H \le R\}$$

and construct the pairwise distant points: $h_1, h_2, \dots$ such that $||h_i||_H \le R$ and $||Lh_i - Lh_j||_{\mathcal{X}} > \varepsilon$ for $i \ne j$.



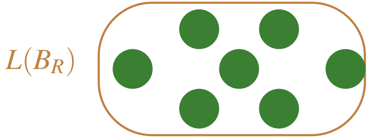
Clearly, we may collect at least $N_{L(B_R)}(\varepsilon)$ points and

$$N_{L(B_R)}(\varepsilon) = N_{L(B_1)}(\varepsilon/R) = N_L(\varepsilon/R).$$

The green balls are disjoint !

How the connection occurs - continued

We have a picture from the former slide



The green balls are $Lh_j + \frac{\varepsilon}{2}\,U$ where U is the unit ball in $\mathcal{X}$.
Christer Borell shift inequality: for every symmetric set $B \subset \mathcal{X}$ and every associated centered Gaussian vector X and operator L, and every $h \in H$

$$\mathbb{P}(X \in B + Lh) \geq \mathbb{P}(X \in B)\exp\{-||h||_H^2/2\}.$$

It follows that

$$\begin{aligned} 1 &\geq \mathbb{P}\Big(X \in \bigcup_j \{Lh_j + \frac{\varepsilon}{2}U\}\Big) = \sum_j \mathbb{P}\Big(X \in \{Lh_j + \frac{\varepsilon}{2}U\}\Big) \\ &\geq N_L(\varepsilon/R)\,\mathbb{P}\big(X \in \tfrac{\varepsilon}{2}U\big)\, e^{-R^2/2} = N_L(\varepsilon/R)\,\mathbb{P}\Big(||X|| \leq \frac{\varepsilon}{2}\Big)\, e^{-R^2/2}. \end{aligned}$$

This reads as $\mathbb{P}\big(||X|| \leq \frac{\varepsilon}{2}\big) \leq e^{R^2/2} N_L(\varepsilon/R)^{-1}$. Optimize the rhs in R !

Example: smooth Gaussian stationary processes

We work with $X(t), t \in \mathbb{R}$, a stationary centered Gaussian process.
Stationarity means that

$$\mathbb{E}X(s)X(t) = K(s-t).$$

Process X is identified by its spectral measure $F(du)$, that is

$$\mathbb{E}X(s)X(t) = \int_{\mathbb{R}} e^{i(s-t)u} F(du)$$

Faster F is decreasing at infinity, smoother is the process X.

In the following we restrict X to the interval $[0,1]$ and evaluate the small deviations of X with respect to the uniform norm

$$||x||_\infty = \max_{t\in[0,1]} |x(t)|.$$

Example: smooth Gaussian stationary processes (continued)

Motivated by Bayesian statistics, we are interested in the case of rather smooth processes.
Namely, consider the family of processes X_ν corresponding to absolutely continuous spectral measures

$$F_\nu(du) = \exp\{-|u|^\nu\}du, \qquad 0 < \nu < \infty.$$

We formally close this family with $F_\infty(du) = \mathbf{1}_{[-1,1]}du$.

The most interesting cases are $\nu = 1$ (exponential spectrum) and $\nu = 2$ (normal spectrum).

Example: smooth Gaussian stationary processes (continued)

By using

$$\ln N_L(\varepsilon) \approx |\ln \varepsilon|^\beta \ln|\ln \varepsilon|^\gamma \Leftrightarrow \varphi(r) \approx |\ln \varepsilon|^\beta \ln|\ln \varepsilon|^\gamma, \qquad \text{as } \varepsilon \to 0.$$

we obtain

Small deviations for exponential continuous spectra

We have, as $\varepsilon \to 0$,

$$\varphi(\varepsilon) \approx \frac{|\ln \varepsilon|^2}{\ln|\ln \varepsilon|}, \qquad 1 < \nu \leq \infty,$$

and

$$\varphi(\varepsilon) \approx |\ln \varepsilon|^{1+\frac{1}{\nu}}, \qquad 0 < \nu \leq 1.$$