Metric entropy and small deviations of Gaussian processes

Thomas Kühn

Universität Leipzig, Germany

2nd International Workshop on Interpolation Theory, Function Spaces and Related Topics, Santiago de Compostela, 18–22 June 2011
18 June 2011

Plan of the talk

1. Metric entropy - basic definitions and properties
2. Gaussian measures and processes
3. Relations between small deviations and metric entropy
4. The d-dimensional Brownian sheet
5. A family of smooth Gaussian processes
6. Covering numbers in Gaussian RKHS

Plan of the talk

1. Metric entropy - basic definitions and properties
2. Gaussian measures and processes
3. Relations between small deviations and metric entropy
4. The d-dimensional Brownian sheet
5. A family of smooth Gaussian processes
6. Covering numbers in Gaussian RKHS

Two main aims of the talk

- to explain how small deviations and metric entropy are connected
- to illustrate this connection by some concrete examples

1. Metric entropy - basic definitions and properties

1. Metric entropy - basic definitions and properties

- Metric entropy
  - – introduced by Pontryagin and Schnirelman (1932)
  - – first systematic treatment by Kolmogorov/Tihomirov (1959)
  - – later by G.G.Lorentz and many others
  - – different versions: ε-entropy, covering numbers, entropy numbers
  - – defined for sets and for operators

1. Metric entropy - basic definitions and properties

- Metric entropy
  – introduced by Pontryagin and Schnirelman (1932)
  – first systematic treatment by Kolmogorov/Tihomirov (1959)
  – later by G.G.Lorentz and many others
  – different versions: ε-entropy, covering numbers, entropy numbers
  – defined for sets and for operators

- Many applications
  – approximation theory
  – functional analysis (eigenvalue distributions)
  – PDEs (spectral properties)
  – probability on Banach spaces (small deviations)
  – mathematical learning theory
  – compressed sensing

- Metric entropy of sets.
  Let (M, d) be a metric space, $A \subseteq M$ and $\varepsilon > 0$. We define

- Metric entropy of sets.

  Let (M, d) be a metric space, $A \subseteq M$ and $\varepsilon > 0$. We define

  covering numbers

  $$\mathcal{N}(\varepsilon, A) = \min\{n \in \mathbb{N} : A \text{ can be covered by } n \text{ balls of radius } \varepsilon\}$$

- Metric entropy of sets.

  Let (M,d) be a metric space, $A \subseteq M$ and $\varepsilon > 0$. We define

  covering numbers

  $\mathcal{N}(\varepsilon, A) = \min\{n \in \mathbb{N} : A$ can be covered by n balls of radius $\varepsilon\}$

  packing numbers

  $\mathcal{M}(\varepsilon, A) = \max\{n \in \mathbb{N} : \exists\ n$ points in A of mutual distance $> \varepsilon\}$

- Metric entropy of sets.

  Let (M, d) be a metric space, $A \subseteq M$ and $\varepsilon > 0$. We define

  covering numbers

  $\mathcal{N}(\varepsilon, A) = \min\{n \in \mathbb{N} : A$ can be covered by n balls of radius $\varepsilon\}$

  packing numbers

  $\mathcal{M}(\varepsilon, A) = \max\{n \in \mathbb{N} : \exists\ n$ points in A of mutual distance $> \varepsilon\}$

  Kolmogorov's ε-entropy $\quad \mathcal{H}(\varepsilon, A) = \log_2 \mathcal{N}(\varepsilon, A)$

- Metric entropy of sets.
  Let (M, d) be a metric space, $A \subseteq M$ and $\varepsilon > 0$. We define

  covering numbers
  $$\mathcal{N}(\varepsilon, A) = \min\{n \in \mathbb{N} : A \text{ can be covered by } n \text{ balls of radius } \varepsilon\}$$

  packing numbers
  $$\mathcal{M}(\varepsilon, A) = \max\{n \in \mathbb{N} : \exists\ n \text{ points in } A \text{ of mutual distance} > \varepsilon\}$$

  Kolmogorov's ε-entropy $\quad \mathcal{H}(\varepsilon, A) = \log_2 \mathcal{N}(\varepsilon, A)$

- Basic inequalities. $\boxed{\mathcal{M}(2\varepsilon, A) \le \mathcal{N}(\varepsilon, A) \le \mathcal{M}(\varepsilon, A)}$

- Metric entropy of sets.
  Let (M, d) be a metric space, $A \subseteq M$ and $\varepsilon > 0$. We define

  covering numbers
  $$\mathcal{N}(\varepsilon, A) = \min\{n \in \mathbb{N} : A \text{ can be covered by } n \text{ balls of radius } \varepsilon\}$$

  packing numbers
  $$\mathcal{M}(\varepsilon, A) = \max\{n \in \mathbb{N} : \exists\ n \text{ points in } A \text{ of mutual distance} > \varepsilon\}$$

  Kolmogorov's ε-entropy $\quad \mathcal{H}(\varepsilon, A) = \log_2 \mathcal{N}(\varepsilon, A)$

- Basic inequalities. $\quad \boxed{\mathcal{M}(2\varepsilon, A) \leq \mathcal{N}(\varepsilon, A) \leq \mathcal{M}(\varepsilon, A)}$

- Often the 'inverse' function of $\quad \varepsilon \mapsto \mathcal{N}(\varepsilon, A) \quad$ is more convenient, i.e. the entropy numbers, defined for $n \in \mathbb{N}$ by
  $$\varepsilon_n(A) = \inf\{\varepsilon > 0 : A \text{ can be covered by } n \text{ balls in } M \text{ of radius } \varepsilon\}$$

- **Metric entropy of sets.**
  Let (M, d) be a metric space, $A \subseteq M$ and $\varepsilon > 0$. We define

  covering numbers
  $$\mathcal{N}(\varepsilon, A) = \min\{n \in \mathbb{N} : A \text{ can be covered by } n \text{ balls of radius } \varepsilon\}$$

  packing numbers
  $$\mathcal{M}(\varepsilon, A) = \max\{n \in \mathbb{N} : \exists\ n \text{ points in } A \text{ of mutual distance} > \varepsilon\}$$

  Kolmogorov's ε-entropy $\quad \mathcal{H}(\varepsilon, A) = \log_2 \mathcal{N}(\varepsilon, A)$

- **Basic inequalities.** $\boxed{\mathcal{M}(2\varepsilon, A) \leq \mathcal{N}(\varepsilon, A) \leq \mathcal{M}(\varepsilon, A)}$

- Often the 'inverse' function of $\varepsilon \mapsto \mathcal{N}(\varepsilon, A)$ is more convenient, i.e. the entropy numbers, defined for $n \in \mathbb{N}$ by
  $$\varepsilon_n(A) = \inf\{\varepsilon > 0 : A \text{ can be covered by } n \text{ balls in } M \text{ of radius } \varepsilon\}$$

- A is precompact $\Longleftrightarrow \mathcal{N}(\varepsilon, A) < \infty$ for all $\varepsilon > 0 \Longleftrightarrow \lim_{n\to\infty} \varepsilon_n(A) = 0$

  $\curvearrowright$ rate of decay of $\varepsilon_n(A)$ describes the 'degree' of compactness of A

- Metric entropy of operators.
  Let X, Y be (quasi-)Banach spaces, B_X the closed unit ball in X and $T : X \to Y$ a (bounded linear) operator. We define

- Metric entropy of operators.
  Let X, Y be (quasi-)Banach spaces, B_X the closed unit ball in X and $T : X \to Y$ a (bounded linear) operator. We define

  covering numbers $\qquad \mathcal{N}(\varepsilon, T) = \mathcal{N}(\varepsilon, T(B_X); Y)$

- Metric entropy of operators.
  Let X, Y be (quasi-)Banach spaces, B_X the closed unit ball in X and $T : X \to Y$ a (bounded linear) operator. We define

  covering numbers $\mathcal{N}(\varepsilon, T) = \mathcal{N}(\varepsilon, T(B_X); Y)$

  (dyadic) entropy numbers $e_k(T) = \varepsilon_{2^{k-1}}(T(B_X))$

- Metric entropy of operators.
  Let X, Y be (quasi-)Banach spaces, B_X the closed unit ball in X and $T : X \to Y$ a (bounded linear) operator. We define

  covering numbers $\qquad \mathcal{N}(\varepsilon, T) = \mathcal{N}(\varepsilon, T(B_X); Y)$

  (dyadic) entropy numbers $\qquad e_k(T) = \varepsilon_{2^{k-1}}(T(B_X))$

- T is compact $\iff \lim_{k\to\infty} e_k(T) = 0 \iff \mathcal{N}(\varepsilon, T) < \infty$ for all $\varepsilon > 0$

- Metric entropy of operators.
  Let X, Y be (quasi-)Banach spaces, B_X the closed unit ball in X and $T: X \to Y$ a (bounded linear) operator. We define

  covering numbers $\qquad \mathcal{N}(\varepsilon, T) = \mathcal{N}(\varepsilon, T(B_X); Y)$

  (dyadic) entropy numbers $\qquad e_k(T) = \varepsilon_{2^{k-1}}(T(B_X))$

- T is compact $\iff \lim_{k\to\infty} e_k(T) = 0 \iff \mathcal{N}(\varepsilon, T) < \infty$ for all $\varepsilon > 0$

- Relation to eigenvalues. If $T: X \to X$ is a compact operator, then its spectrum $\sigma(T)$ has no accumulation points except possibly 0, and all $\lambda \in \sigma(T), \lambda \neq 0$, are eigenvalues of finite multiplicity.

- Metric entropy of operators.
  Let X, Y be (quasi-)Banach spaces, B_X the closed unit ball in X and $T : X \to Y$ a (bounded linear) operator. We define

  covering numbers $\qquad \mathcal{N}(\varepsilon, T) = \mathcal{N}(\varepsilon, T(B_X); Y)$

  (dyadic) entropy numbers $\qquad e_k(T) = \varepsilon_{2^{k-1}}(T(B_X))$

- T is compact $\iff \lim\limits_{k\to\infty} e_k(T) = 0 \iff \mathcal{N}(\varepsilon, T) < \infty$ for all $\varepsilon > 0$

- Relation to eigenvalues. If $T : X \to X$ is a compact operator, then its spectrum $\sigma(T)$ has no accumulation points except possibly 0, and all $\lambda \in \sigma(T), \lambda \neq 0$, are eigenvalues of finite multiplicity.

  $\curvearrowright$ Ordering of the eigenvalues (counting multiplicities)

  $$|\lambda_1(T)| \geq |\lambda_2(T)| \geq ... \geq |\lambda_k(T)| \geq ... \geq 0$$

  If T has $< n$ eigenvalues, we set $\lambda_n(T) = \lambda_{n+1}(T) = \ldots = 0$.

- Metric entropy of operators.
  Let X, Y be (quasi-)Banach spaces, B_X the closed unit ball in X and $T : X \to Y$ a (bounded linear) operator. We define

  covering numbers $\qquad \mathcal{N}(\varepsilon, T) = \mathcal{N}(\varepsilon, T(B_X); Y)$

  (dyadic) entropy numbers $\qquad e_k(T) = \varepsilon_{2^{k-1}}(T(B_X))$

- T is compact $\iff \lim_{k\to\infty} e_k(T) = 0 \iff \mathcal{N}(\varepsilon, T) < \infty$ for all $\varepsilon > 0$

- Relation to eigenvalues. If $T : X \to X$ is a compact operator, then its spectrum $\sigma(T)$ has no accumulation points except possibly 0, and all $\lambda \in \sigma(T), \lambda \neq 0$, are eigenvalues of finite multiplicity.

  $\curvearrowright$ Ordering of the eigenvalues (counting multiplicities)

  $$|\lambda_1(T)| \geq |\lambda_2(T)| \geq ... \geq |\lambda_k(T)| \geq ... \geq 0$$

  If T has $< n$ eigenvalues, we set $\lambda_n(T) = \lambda_{n+1}(T) = \ldots = 0$.

  Carl-Triebel inequality. $\qquad \boxed{|\lambda_k(T)| \leq \sqrt{2}e_k(T)}$

2. Gaussian measures and processes

2. Gaussian measures and processes

- Let E be a real separable Banach space with dual E'.
  A Gaussian measure μ on E is a Borel measure on E such that all its one-dimensional image measures $\mu_a = a(\mu)$, $a \in E'$, are centered Gaussian measures on $\mathbb{R}$.

2. Gaussian measures and processes

- Let E be a real separable Banach space with dual E'.
  A Gaussian measure μ on E is a Borel measure on E such that all its one-dimensional image measures $\mu_a = a(\mu)$, $a \in E'$, are centered Gaussian measures on $\mathbb{R}$.

- μ_a is uniquely determined by its variance $\sigma_a^2 = \int_{\mathbb{R}} t^2 \, d\mu_a(t)$
  $\curvearrowright$ μ is uniquely determined by each of the following quantities:
  – characteristic function $\widehat{\mu}(a) := \int_E e^{i\langle x,a\rangle} \, d\mu(x) = e^{-\sigma_a^2/2}$
  – covariance operator $R : E' \to E$, $Ra := \underbrace{\int_E x\langle x,a\rangle \, d\mu(x)}_{\text{Bochner integral}}$
  – reproducing kernel Hilbert space H_μ, i.e. the completion of $R(E')$ w.r.t. the inner product

$$\langle Ra, Rb\rangle_\mu = \int_E \langle x,a\rangle\langle x,b\rangle \, d\mu(x) \,, \quad a,b \in E'$$

- Remark. For every Gaussian measure μ on E there are a Hilbert space H and an operator $T : H \to E$ s.t. $\mu = T(\gamma_H)$, where γ_H is the canonical Gaussian cylinder measure on H.

  $\curvearrowright$ the covariance operator of μ is $R = TT'$,

  in particular $\sigma_a^2 = \int_E \langle x, a\rangle^2 d\mu(x) = \langle Ra, a\rangle = \|T'a\|^2$

  $\curvearrowright$ the characteristic function of μ is

  $\widehat{\mu}(a) = e^{-\sigma_a^2/2} = e^{-\|T'a\|^2/2}$

  Clearly T is not unique, but one can always take $T = Id : H_\mu \to E$.

- Remark. For every Gaussian measure μ on E there are a Hilbert space H and an operator $T : H \to E$ s.t. $\mu = T(\gamma_H)$, where γ_H is the canonical Gaussian cylinder measure on H.

  $\curvearrowright$ the covariance operator of μ is $R = TT'$,

  in particular $\sigma_a^2 = \int_E \langle x, a\rangle^2 d\mu(x) = \langle Ra, a\rangle = \|T'a\|^2$

  $\curvearrowright$ the characteristic function of μ is

  $\widehat{\mu}(a) = e^{-\sigma_a^2/2} = e^{-\|T'a\|^2/2}$

  Clearly T is not unique, but one can always take $T = Id : H_\mu \to E$.

- A centered Gaussian process is a random process $X = X(t), t \in I$, such that all $X(t)$ are centered normal random variables. The process is uniquely determined by its covariance structure, i.e. the function

$$K(s,t) = \mathbb{E}X(s)X(t) \quad (s, t \in I)\,.$$

  We will consider processes with index set $I = [0,1]^d$, $d \in \mathbb{N}$.

- **The small ball/small deviation problem.** Let μ be a Gaussian measure on a separable Banach space E, e.g. the distribution of a Gaussian process $X = X(t)\,, t \in [0,1]^d\,$, whose sample paths are almost surely in some Banach space E of functions. In most cases $E = L_2([0,1]^d)$ or $C([0,1]^d)$ or some Orlicz space.

- The small ball/small deviation problem. Let μ be a Gaussian measure on a separable Banach space E, e.g. the distribution of a Gaussian process $X = X(t)\,, t \in [0,1]^d$, whose sample paths are almost surely in some Banach space E of functions. In most cases $E = L_2([0,1]^d)$ or $C([0,1]^d)$ or some Orlicz space.

  Problem. Determine the asymptotic behaviour as $\varepsilon \to 0$ of the small ball probabilities of the measure μ, resp. of the small deviation probabilities of the process $X(t)$

$$\varphi(\varepsilon) := \begin{cases} \varphi_\mu(\varepsilon) & = -\log \mu\left(\{x \in E : \|x\|_E \le \varepsilon\}\right) \\ \varphi_X(\varepsilon) & = -\log \mathbb{P}\left(\|X(.)\|_E \le \varepsilon\}\right) . \end{cases}$$

- The small ball/small deviation problem. Let μ be a Gaussian measure on a separable Banach space E, e.g. the distribution of a Gaussian process $X = X(t)\,, t \in [0,1]^d$, whose sample paths are almost surely in some Banach space E of functions. In most cases $E = L_2([0,1]^d)$ or $C([0,1]^d)$ or some Orlicz space.

  Problem. Determine the asymptotic behaviour as $\varepsilon \to 0$ of the small ball probabilities of the measure μ, resp. of the small deviation probabilities of the process $X(t)$

$$\varphi(\varepsilon) := \begin{cases} \varphi_\mu(\varepsilon) & = -\log \mu\left(\{x \in E : \|x\|_E \leq \varepsilon\}\right) \\ \varphi_X(\varepsilon) & = -\log \mathbb{P}\left(\|X(.)\|_E \leq \varepsilon\}\right) . \end{cases}$$

- Many applications in probability and analysis.
  - Law of the iterated logarithm of Chung type
  - Strong limit laws in statistics
  - Quantization (approximation) of stochastic processes
  - Metric entropy of linear operators

3. Relations between small deviations and metric entropy

3. Relations between small deviations and metric entropy

- Notation. E – real separable Banach space
  B – closed unit ball of E
  μ – Gaussian measure on E
  K – unit ball of the RKHS H_μ of μ.

3. Relations between small deviations and metric entropy

- **Notation.** E – real separable Banach space
  B – closed unit ball of E
  μ – Gaussian measure on E
  K – unit ball of the RKHS H_μ of μ.

- **Kuelbs/Li 1993.** Let $\lambda > 0$ and $\varepsilon > 0$. Then

$$\boxed{\varphi(2\varepsilon) + \log \Phi(\lambda + \alpha_\varepsilon) \leq \mathcal{H}(\varepsilon, \lambda K) \leq 2\lambda^2 + \varphi(\varepsilon)}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$ and $\Phi(\alpha_\varepsilon) = \mu(\varepsilon B)$.

3. Relations between small deviations and metric entropy

- Notation. E – real separable Banach space
  B – closed unit ball of E
  μ – Gaussian measure on E
  K – unit ball of the RKHS H_μ of μ.

- Kuelbs/Li 1993. Let $\lambda > 0$ and $\varepsilon > 0$. Then

$$\boxed{\varphi(2\varepsilon) + \log \Phi(\lambda + \alpha_\varepsilon) \leq \mathcal{H}(\varepsilon, \lambda K) \leq 2\lambda^2 + \varphi(\varepsilon)}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$ and $\Phi(\alpha_\varepsilon) = \mu(\varepsilon B)$.

- $\lambda = \sqrt{\frac{\varphi(\varepsilon)}{2}}$ $\curvearrowright$ $\boxed{\varphi(2\varepsilon) \leq \mathcal{H}\left(\varepsilon \sqrt{\frac{2}{\varphi(\varepsilon)}}\, , K\right) \leq 2\varphi(\varepsilon)}$

  This looks quite complicated, but is very useful!

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\quad \|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\quad \|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

$\curvearrowright$ the (closed) balls $B_\varepsilon(x_i)$ are disjoint

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

$\curvearrowright$ the (closed) balls $B_\varepsilon(x_i)$ are disjoint

known in probability: $x \in aK \Longrightarrow \mu(B_\varepsilon(x)) \geq e^{-a^2/2}\mu(B_\varepsilon(0))$

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\quad \|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

$\curvearrowright$ the (closed) balls $B_\varepsilon(x_i)$ are disjoint

known in probability: $x \in aK \Longrightarrow \mu(B_\varepsilon(x)) \geq e^{-a^2/2}\mu(B_\varepsilon(0))$

in our case: $a = 2\lambda \quad \curvearrowright \quad \mu(B_\varepsilon(x_i)) \geq e^{-2\lambda^2}\mu(B_\varepsilon(0))$

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

$\curvearrowright$ the (closed) balls $B_\varepsilon(x_i)$ are disjoint

known in probability: $x \in aK \Longrightarrow \mu(B_\varepsilon(x)) \geq e^{-a^2/2}\mu(B_\varepsilon(0))$

in our case: $a = 2\lambda \quad \curvearrowright \quad \mu(B_\varepsilon(x_i)) \geq e^{-2\lambda^2}\mu(B_\varepsilon(0))$

$\curvearrowright \quad 1 = \mu(E) \geq Me^{-2\lambda^2}\mu(B_\varepsilon(0)) \quad \curvearrowright \quad M \leq e^{2\lambda^2}\frac{1}{\mu(B_\varepsilon(0))}$

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\quad \|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

$\curvearrowright$ the (closed) balls $B_\varepsilon(x_i)$ are disjoint

known in probability: $x \in aK \Longrightarrow \mu(B_\varepsilon(x)) \geq e^{-a^2/2}\mu(B_\varepsilon(0))$

in our case: $a = 2\lambda \quad \curvearrowright \quad \mu(B_\varepsilon(x_i)) \geq e^{-2\lambda^2}\mu(B_\varepsilon(0))$

$\curvearrowright \quad 1 = \mu(E) \geq M e^{-2\lambda^2}\mu(B_\varepsilon(0)) \quad \curvearrowright \quad M \leq e^{2\lambda^2}\frac{1}{\mu(B_\varepsilon(0))}$

$\curvearrowright \quad \mathcal{H}(\varepsilon, \lambda K) = \log \mathcal{N}(\varepsilon, \lambda K) \leq \log \mathcal{M}(\varepsilon, \lambda K) = \log M \leq 2\lambda^2 + \varphi(\varepsilon)$

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\quad \|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

$\curvearrowright$ the (closed) balls $B_\varepsilon(x_i)$ are disjoint

known in probability: $x \in aK \Longrightarrow \mu(B_\varepsilon(x)) \geq e^{-a^2/2}\mu(B_\varepsilon(0))$

in our case: $a = 2\lambda \quad \curvearrowright \quad \mu(B_\varepsilon(x_i)) \geq e^{-2\lambda^2}\mu(B_\varepsilon(0))$

$\curvearrowright \quad 1 = \mu(E) \geq M e^{-2\lambda^2}\mu(B_\varepsilon(0)) \quad \curvearrowright \quad M \leq e^{2\lambda^2}\frac{1}{\mu(B_\varepsilon(0))}$

$\curvearrowright \quad \mathcal{H}(\varepsilon, \lambda K) = \log \mathcal{N}(\varepsilon, \lambda K) \leq \log \mathcal{M}(\varepsilon, \lambda K) = \log M \leq 2\lambda^2 + \varphi(\varepsilon)$

Proof of the lower estimate. more complicated

Proof of the upper estimate.

Let $M = \mathcal{M}(\varepsilon, \lambda K) = \mathcal{M}(2\varepsilon, 2\lambda K)$.
Choose $x_1, ..., x_M \in 2\lambda K$ with $\|x_i - x_j\|_E > 2\varepsilon$ for all $i \neq j$.

$\curvearrowright$ the (closed) balls $B_\varepsilon(x_i)$ are disjoint

known in probability: $x \in aK \implies \mu(B_\varepsilon(x)) \geq e^{-a^2/2}\mu(B_\varepsilon(0))$

in our case: $a = 2\lambda \quad \curvearrowright \quad \mu(B_\varepsilon(x_i)) \geq e^{-2\lambda^2}\mu(B_\varepsilon(0))$

$\curvearrowright \quad 1 = \mu(E) \geq Me^{-2\lambda^2}\mu(B_\varepsilon(0)) \quad \curvearrowright \quad M \leq e^{2\lambda^2}\frac{1}{\mu(B_\varepsilon(0))}$

$\curvearrowright \quad \mathcal{H}(\varepsilon, \lambda K) = \log \mathcal{N}(\varepsilon, \lambda K) \leq \log \mathcal{M}(\varepsilon, \lambda K) = \log M \leq 2\lambda^2 + \varphi(\varepsilon)$

Proof of the lower estimate. more complicated

- Some consequences.
  1. Goodman 1990. $\mathcal{H}(\varepsilon, K) = o(\varepsilon^{-2}) \qquad \curvearrowright H_\mu \hookrightarrow E$ compactly.
  2. Li/Linde 1999. $\varphi(\varepsilon) \sim \varepsilon^{-\alpha} \iff \mathcal{H}(\varepsilon, K) \sim \varepsilon^{-\frac{2\alpha}{2+\alpha}} \qquad (\alpha > 0)$
  3. Aurzada/Ibragimov/Lifshits/van Zanten 2008.

$$\varphi(\varepsilon) \sim (\log \tfrac{1}{\varepsilon})^\alpha \iff \mathcal{H}(\varepsilon, K) \sim (\log \tfrac{1}{\varepsilon})^\alpha$$

Proofs. 1. Goodman: $\mathcal{H}(\varepsilon, \lambda K) \leq 2\lambda^2 + \varphi(\varepsilon)$

$\curvearrowright \quad \left(\frac{\varepsilon}{\lambda}\right)^2 \mathcal{H}\left(\frac{\varepsilon}{\lambda}, K\right) \leq 2\varepsilon^2 + \frac{\varepsilon^2 \varphi(\varepsilon)}{\lambda^2}$ for all $\varepsilon, \lambda > 0$.

Proofs. 1. Goodman: $\mathcal{H}(\varepsilon, \lambda K) \leq 2\lambda^2 + \varphi(\varepsilon)$

$\curvearrowright$ $\left(\frac{\varepsilon}{\lambda}\right)^2 \mathcal{H}\left(\frac{\varepsilon}{\lambda}, K\right) \leq 2\varepsilon^2 + \frac{\varepsilon^2\varphi(\varepsilon)}{\lambda^2}$ for all $\varepsilon, \lambda > 0$.

Keep now $\varepsilon > 0$ fixed, set $\delta = \frac{\varepsilon}{\lambda}$, and let $\lambda \to \infty$. Then $\delta \to 0$ and

$$\limsup_{\delta \to 0} \delta^2 \mathcal{H}(\delta, K) \leq 2\varepsilon^2 .$$

Proofs. 1. Goodman: $\mathcal{H}(\varepsilon, \lambda K) \leq 2\lambda^2 + \varphi(\varepsilon)$

$\curvearrowright \quad \left(\frac{\varepsilon}{\lambda}\right)^2 \mathcal{H}\left(\frac{\varepsilon}{\lambda}, K\right) \leq 2\varepsilon^2 + \frac{\varepsilon^2 \varphi(\varepsilon)}{\lambda^2}$ for all $\varepsilon, \lambda > 0$.

Keep now $\varepsilon > 0$ fixed, set $\delta = \frac{\varepsilon}{\lambda}$, and let $\lambda \to \infty$. Then $\delta \to 0$ and

$$\limsup_{\delta \to 0} \delta^2 \mathcal{H}(\delta, K) \leq 2\varepsilon^2 .$$

This is true for every $\varepsilon > 0$, whence $\lim_{\delta \to 0} \delta^2 \mathcal{H}(\delta, K) = 0$.

Proofs. 1. Goodman: $\mathcal{H}(\varepsilon, \lambda K) \leq 2\lambda^2 + \varphi(\varepsilon)$

$\curvearrowright$ $\left(\frac{\varepsilon}{\lambda}\right)^2 \mathcal{H}\left(\frac{\varepsilon}{\lambda}, K\right) \leq 2\varepsilon^2 + \frac{\varepsilon^2 \varphi(\varepsilon)}{\lambda^2}$ for all $\varepsilon, \lambda > 0$.

Keep now $\varepsilon > 0$ fixed, set $\delta = \frac{\varepsilon}{\lambda}$, and let $\lambda \to \infty$. Then $\delta \to 0$ and

$$\limsup_{\delta \to 0} \delta^2 \mathcal{H}(\delta, K) \leq 2\varepsilon^2 .$$

This is true for every $\varepsilon > 0$, whence $\lim_{\delta \to 0} \delta^2 \mathcal{H}(\delta, K) = 0$.

2. Li/Linde: Assume $\varphi(\varepsilon) \preceq \varepsilon^{-\alpha}$ for some $\alpha > 0$.

Take $\lambda = \varepsilon^{-\alpha/2}$ in the upper Kuelbs-Li estimate and set $\delta = \varepsilon^{1+\alpha/2}$.

$$\curvearrowright \quad \mathcal{H}(\delta, K) = \mathcal{H}(\varepsilon, \varepsilon^{-\alpha/2} K) \preceq \varepsilon^{-\alpha} = \delta^{-\frac{2\alpha}{2+\alpha}}$$

The other implications are less trivial, the proofs are really very tricky.

Proofs. 1. Goodman: $\mathcal{H}(\varepsilon, \lambda K) \leq 2\lambda^2 + \varphi(\varepsilon)$

$\curvearrowright \quad \left(\frac{\varepsilon}{\lambda}\right)^2 \mathcal{H}\left(\frac{\varepsilon}{\lambda}, K\right) \leq 2\varepsilon^2 + \frac{\varepsilon^2 \varphi(\varepsilon)}{\lambda^2}$ for all $\varepsilon, \lambda > 0$.

Keep now $\varepsilon > 0$ fixed, set $\delta = \frac{\varepsilon}{\lambda}$, and let $\lambda \to \infty$. Then $\delta \to 0$ and

$$\limsup_{\delta \to 0} \delta^2 \mathcal{H}(\delta, K) \leq 2\varepsilon^2 .$$

This is true for every $\varepsilon > 0$, whence $\lim_{\delta \to 0} \delta^2 \mathcal{H}(\delta, K) = 0.$

2. Li/Linde: Assume $\varphi(\varepsilon) \preceq \varepsilon^{-\alpha}$ for some $\alpha > 0$.

Take $\lambda = \varepsilon^{-\alpha/2}$ in the upper Kuelbs-Li estimate and set $\delta = \varepsilon^{1+\alpha/2}$.

$$\curvearrowright \quad \mathcal{H}(\delta, K) = \mathcal{H}(\varepsilon, \varepsilon^{-\alpha/2} K) \preceq \varepsilon^{-\alpha} = \delta^{-\frac{2\alpha}{2+\alpha}}$$

The other implications are less trivial, the proofs are really very tricky.

3. Aurzada et al.: Proof is simpler.

4. The d-dimensional Brownian sheet

- This is the centered Gaussian field $B_d = B_d(t), t \in [0,1]^d$, with covariance structure

$$\mathbb{E}\, B_d(s) B_d(t) = \prod_{j=1}^{d} \min(s_j, t_j) \quad (s, t \in [0,1]^d)\,.$$

For dimension $d = 1$ this is the classical Wiener process.

4. The d-dimensional Brownian sheet

- This is the centered Gaussian field $B_d = B_d(t), t \in [0,1]^d$, with covariance structure

$$\mathbb{E}\, B_d(s) B_d(t) = \prod_{j=1}^{d} \min(s_j, t_j) \quad (s, t \in [0,1]^d)\,.$$

  For dimension $d = 1$ this is the classical Wiener process.

- $B_d(t)$ has almost surely continuous paths and is related to the d-dimensional integration operator $T_d : L_2([0,1]^d) \to C([0,1]^d)$

$$T_d f(x) = \int_0^{x_1} ... \int_0^{x_d} f(y_1, ..., y_d)\, dy_d ... dy_1\,.$$

4. The d-dimensional Brownian sheet

- This is the centered Gaussian field $B_d = B_d(t), t \in [0,1]^d$, with covariance structure

$$\mathbb{E}\, B_d(s) B_d(t) = \prod_{j=1}^{d} \min(s_j, t_j) \quad (s, t \in [0,1]^d)\,.$$

For dimension $d = 1$ this is the classical Wiener process.

- $B_d(t)$ has almost surely continuous paths and is related to the d-dimensional integration operator $T_d : L_2([0,1]^d) \to C([0,1]^d)$

$$T_d f(x) = \int_0^{x_1} \ldots \int_0^{x_d} f(y_1, \ldots, y_d)\, dy_d \ldots dy_1\,.$$

- Small deviation results for the Brownian sheet have a long history. Csáki 1982 solved the L_2-case,

$$-\log \mathbb{P}\left(\|B_d\|_{L_2([0,1]^d)} \le \varepsilon \right) \sim \varepsilon^{-2} |\log \varepsilon|^{2d-2}$$

- For small deviations under sup-norm much less is known.

  $d = 1$: classical result for the Wiener process

$$\log \mathbb{P}\left(\|B_1\|_{C([0,1])} \leq \varepsilon\right) \sim -\frac{\pi^2}{8\varepsilon^2}$$

  $d = 2$: Bass 1988 (lower bound) and Talagrand 1994 (upper bound)

$$\log \mathbb{P}\left(\|B_2\|_{C([0,1]^2)} \leq \varepsilon\right) \sim -\varepsilon^{-2}|\log \varepsilon|^3$$

  $d \geq 3$: still open, only recently some progress by Lacey et al. 2009

- For small deviations under sup-norm much less is known.

  $d = 1$: classical result for the Wiener process

$$\log \mathbb{P}\left(\|B_1\|_{C([0,1])} \le \varepsilon\right) \sim -\frac{\pi^2}{8\varepsilon^2}$$

  $d = 2$: Bass 1988 (lower bound) and Talagrand 1994 (upper bound)

$$\log \mathbb{P}\left(\|B_2\|_{C([0,1]^2)} \le \varepsilon\right) \sim -\varepsilon^{-2}|\log \varepsilon|^3$$

  $d \geq 3$: still open, only recently some progress by Lacey et al. 2009

- Dunker/Kühn/Lifshits/Linde JAT 1999. Let $d \geq 2$. Then

$$e_k(T_d : L_2([0,1]^d) \to C([0,1]^d)) \preceq k^{-1}(\log k)^{d-1/2}$$

  Consequently, via the Kuelbs-Li result and by the the L_2-case,

$$\varepsilon^{-2}|\log \varepsilon|^{2d-2} \preceq -\log \mathbb{P}\left(\|B_d\|_{C([0,1]^d)} \le \varepsilon\right) \preceq \varepsilon^{-2}|\log \varepsilon|^{2d-1}$$

- Remark. The last entropy estimate has an interpretation in terms of Sobolev spaces with dominating mixed derivatives,

$$e_k(id : W_2^{(1,\dots 1)}([0,1]^d) \to C([0,1]^d)) = e_k(T_d) \preceq k^{-1}(\log k)^{d-1/2} \,.$$

Proof: via relation to Kolmogorov numbers, using the Haar basis in $L_2([0,1]^d)$, tensor techniques, and hyperbolic cross approximation

- Remark. The last entropy estimate has an interpretation in terms of Sobolev spaces with dominating mixed derivatives,

$$e_k(id : W_2^{(1,\dots 1)}([0,1]^d) \to C([0,1]^d)) = e_k(T_d) \preceq k^{-1}(\log k)^{d-1/2}\,.$$

  Proof: via relation to Kolmogorov numbers, using the Haar basis in $L_2([0,1]^d)$, tensor techniques, and hyperbolic cross approximation

- Intuition. Note that $B_d(0)$ has variance 0, i.e. $B_d(0) = 0$ a.s.

  $\curvearrowright$ small entropy $\mathcal{H}(\varepsilon, K)$ means high concentration of the process near 0

  $\curvearrowright$ large probability $\mathbb{P}(\|B_d\| \le \varepsilon)$

- Remark. The last entropy estimate has an interpretation in terms of Sobolev spaces with dominating mixed derivatives,

$$e_k(id : W_2^{(1,\dots 1)}([0,1]^d) \to C([0,1]^d)) = e_k(T_d) \preceq k^{-1}(\log k)^{d-1/2}\,.$$

  Proof: via relation to Kolmogorov numbers, using the Haar basis in $L_2([0,1]^d)$, tensor techniques, and hyperbolic cross approximation

- Intuition. Note that $B_d(0)$ has variance 0, i.e. $B_d(0) = 0$ a.s.
  $\curvearrowright$ small entropy $\mathcal{H}(\varepsilon, K)$ means high concentration of the process near 0
  $\curvearrowright$ large probability $\mathbb{P}(\|B_d\| \leq \varepsilon)$

- L_2-norm versus $\sup$-norm. The sup-norm case is much harder, since the process may attain its maximum on a very small set.

- Remark. The last entropy estimate has an interpretation in terms of Sobolev spaces with dominating mixed derivatives,

$$e_k(id : W_2^{(1,\ldots 1)}([0,1]^d) \to C([0,1]^d)) = e_k(T_d) \preceq k^{-1}(\log k)^{d-1/2} .$$

  Proof: via relation to Kolmogorov numbers, using the Haar basis in $L_2([0,1]^d)$, tensor techniques, and hyperbolic cross approximation

- Intuition. Note that $B_d(0)$ has variance 0, i.e. $B_d(0) = 0$ a.s.
  $\curvearrowright$ small entropy $\mathcal{H}(\varepsilon, K)$ means high concentration of the process near 0
  $\curvearrowright$ large probability $\mathbb{P}(\|B_d\| \leq \varepsilon)$

- L_2-norm versus $\sup$-norm. The sup-norm case is much harder, since the process may attain its maximum on a very small set.

- Nice interplay probability – analysis.
  probabilistic estimates for small deviations of processes $\leftrightarrow$
  analytical estimates for entropy of operators

5. A family of smooth Gaussian processes

5. A family of smooth Gaussian processes

- Consider the family of centered Gaussian processes $X_{\alpha,\beta}(t)\,, t \geq 0\,,$ defined by the covariance function

$$K(t,s) := \mathbb{E} X_{\alpha,\beta}(s) X_{\alpha,\beta}(t) = \frac{2^{2\beta+1}(st)^{\alpha}}{(s+t)^{2\beta+1}}\,.$$

5. A family of smooth Gaussian processes

- Consider the family of centered Gaussian processes $X_{\alpha,\beta}(t)\,, t \geq 0\,,$ defined by the covariance function

$$K(t,s) := \mathbb{E} X_{\alpha,\beta}(s) X_{\alpha,\beta}(t) = \frac{2^{2\beta+1}(st)^{\alpha}}{(s+t)^{2\beta+1}}\,.$$

- Problem posed at the Workshop in Palo Alto, December 2008: Find the small deviation rates of the processes $X_{\alpha,\beta}$ w.r.t. the
  – L_2 norm, if $\alpha > 0$ and $-1/2 < \beta < \alpha$
  – $\sup$ norm, if $\alpha > \beta + 1/2 > 0\,.$

5. A family of smooth Gaussian processes

- Consider the family of centered Gaussian processes $X_{\alpha,\beta}(t)\,, t \geq 0\,,$ defined by the covariance function

$$K(t,s) := \mathbb{E}X_{\alpha,\beta}(s)X_{\alpha,\beta}(t) = \frac{2^{2\beta+1}(st)^{\alpha}}{(s+t)^{2\beta+1}}\,.$$

- Problem posed at the Workshop in Palo Alto, December 2008: Find the small deviation rates of the processes $X_{\alpha,\beta}$ w.r.t. the
  – L_2 norm, if $\alpha > 0$ and $-1/2 < \beta < \alpha$
  – $\sup$ norm, if $\alpha > \beta + 1/2 > 0\,.$

- Remark. The conditions on α, β are best possible to ensure that the sample paths of the process $X_{\alpha,\beta}(t)$ are almost surely in $L_2[0,1]$, respectively in $C[0,1]$.

- It is not hard to check that the process $X_{\alpha,\beta} = X_{\alpha,\beta}(t), 0 \le t \le 1$ is related to the operator

$$(Sf)(t) = t^{\alpha} \int_0^{\infty} x^{\beta} e^{-xt} f(x)\, dx\,, \quad f \in L_2[0,\infty), t \in [0,1]$$

- It is not hard to check that the process $X_{\alpha,\beta} = X_{\alpha,\beta}(t), 0 \leq t \leq 1$ is related to the operator

$$(Sf)(t) = t^{\alpha} \int_0^{\infty} x^{\beta} e^{-xt} f(x)\, dx\,, \quad f \in L_2[0,\infty), t \in [0,1]$$

- Theorem. (small deviations under L_2-norm)
  Let $\alpha > 0$ and $\alpha > \beta > -1/2$. Then

$$\log \mathbb{P}\left(\int_0^1 X_{\alpha,\beta}(t)^2\, dt \leq \varepsilon^2 \right) \sim -(\log \tfrac{1}{\varepsilon})^3 \quad \text{as } \varepsilon \to 0\,.$$

- It is not hard to check that the process $X_{\alpha,\beta} = X_{\alpha,\beta}(t), 0 \leq t \leq 1$ is related to the operator

$$(Sf)(t) = t^\alpha \int_0^\infty x^\beta e^{-xt} f(x)\, dx\,, \quad f \in L_2[0,\infty), t \in [0,1]$$

- Theorem. (small deviations under L_2-norm)
Let $\alpha > 0$ and $\alpha > \beta > -1/2$. Then

$$\log \mathbb{P}\left(\int_0^1 X_{\alpha,\beta}(t)^2\, dt \leq \varepsilon^2\right) \sim -(\log \tfrac{1}{\varepsilon})^3 \quad \text{as } \varepsilon \to 0\,.$$

- By the Kuelbs-Li entropy connection (in the log-case) it is clear that this would follow from the

  Proposition. For all $\alpha > \beta > -1/2$ the entropy numbers of the operator $S : L_2[0,\infty) \to L_2[0,1]$ satisfy

$$\log e_n(S) \sim -\sqrt[3]{n} \quad \text{as } n \to \infty\,.$$

- Sketch of proof. The operator $T := SS' : L_2[0,1] \to L_2[0,1]$ is

$$(Tf)(t) = \Gamma(2\beta+1) \int_0^1 \frac{(tx)^\alpha}{(t+x)^{2\beta+1}}\, dt\,.$$

The singular numbers of T are known (Laptev 1974)

$$s_n(T) \approx e^{-2c\sqrt{n}}, \quad \text{where } c = c_{\alpha,\beta} = \frac{\sqrt{\alpha-\beta}}{\pi}$$

Now $\quad s_n(T) = s_n(S)^2 \quad$ implies $\quad s_n(S) \approx e^{-c\sqrt{n}}\,.$

Since we are in the Hilbert space setting, the singular numbers and the entropy numbers of S coincide with those of the diagonal operator D_σ in ℓ_2, $(x_n) \mapsto (\sigma_n x_n)$, with $\sigma_n = s_n(S)$.

Gordon-König-Schütt 1987: $\quad e_k(D_\sigma) \sim \sup_{n\geq 1} 2^{-k/n}(\sigma_1 \cdots \sigma_n)^{1/n}$

$$\curvearrowright \quad -\log e_k(S) \sim \inf_{n\geq 1}\left(\tfrac{k}{n} + \tfrac{1}{n}\sum_{j=1}^{n}\sqrt{j}\right) \sim \inf_{n\geq 1}\left(\tfrac{k}{n} + \sqrt{n}\right) \sim \sqrt[3]{n}\,.$$

- Theorem. (small deviations under sup-norm)
  Aurzada/Gao/Kühn/Li/Shao JTP 2011+
  Let $\alpha > \beta + \frac{1}{2} > 0$. Then

$$\log \mathbb{P}\left(\sup_{0\leq t\leq 1} |X_{\alpha,\beta}(t)| \leq \varepsilon\right) \sim -(\log \tfrac{1}{\varepsilon})^3 \,.$$

  (same estimate as in L_2-norm, but for a smaller range of α)

- Theorem. (small deviations under sup-norm)
  Aurzada/Gao/Kühn/Li/Shao JTP 2011+
  Let $\alpha > \beta + \frac{1}{2} > 0$. Then

$$\log \mathbb{P} \left(\sup_{0 \le t \le 1} |X_{\alpha,\beta}(t)| \le \varepsilon \right) \sim -(\log \tfrac{1}{\varepsilon})^3 \,.$$

  (same estimate as in L_2-norm, but for a smaller range of α)

- Of course, the process $X_{\alpha,\beta}$ is related to the same operator S, but now considered as an operator into $C[0,1]$.
  $\curvearrowright$ enough to show: $S : L_2 \to C$ satisfies the same entropy estimate as $S : L_2 \to L_2$, that means $\log e_n(S) \sim -\sqrt[3]{n}$.
  Since $C \hookrightarrow L_2$ with norm one, the lower estimate is clear.

- **Theorem.** (small deviations under sup-norm)
  Aurzada/Gao/Kühn/Li/Shao JTP 2011+
  Let $\alpha > \beta + \frac{1}{2} > 0$. Then

$$\log \mathbb{P}\left(\sup_{0 \le t \le 1} |X_{\alpha,\beta}(t)| \le \varepsilon \right) \sim -(\log \tfrac{1}{\varepsilon})^3 \,.$$

  (same estimate as in L_2-norm, but for a smaller range of α)

- Of course, the process $X_{\alpha,\beta}$ is related to the same operator S, but now considered as an operator into $C[0,1]$.
  $\curvearrowright$ enough to show: $S : L_2 \to C$ satisfies the same entropy estimate as $S : L_2 \to L_2$, that means $\log e_n(S) \sim -\sqrt[3]{n}$.
  Since $C \hookrightarrow L_2$ with norm one, the lower estimate is clear.

- **Idea for the upper estimate.** Observe that S maps even into a smaller space than $C[0,1]$, namely in the Hölder space $C^\lambda[0,1]$ with $\lambda := \min(\alpha - \beta - \frac{1}{2}, \frac{1}{2}) > 0$.
  But how can we take advantage of this fact? By interpolation!

- Lemma. Let u be an operator from some Banach space E into $C^\lambda[0,1]$ for some $0 < \lambda \leq 1$. Then

$$e_n(u : E \to C) \leq 2\|u : E \to C^\lambda\|^{\frac{1}{2\lambda+1}} \cdot e_n(u : E \to L_2)^{\frac{2\lambda}{2\lambda+1}} .$$

- Lemma. Let u be an operator from some Banach space E into $C^\lambda[0,1]$ for some $0 < \lambda \leq 1$. Then

$$e_n(u : E \to C) \leq 2\|u : E \to C^\lambda\|^{\frac{1}{2\lambda+1}} \cdot e_n(u : E \to L_2)^{\frac{2\lambda}{2\lambda+1}} .$$

Sketch of the proof.
First we construct averaging operators $P_\delta : L_2 \to C$ with

$$\|P_\delta : L_2 \to C\| \leq \delta^{-1/2} \quad \text{and} \quad \|id - P_\delta : C^\lambda \to C\| \leq \delta^\lambda$$

By elementary properties of entropy numbers this gives

$$e_n(u : E \to C) \leq \|u : E \to C\| \cdot \delta^\lambda + e_n(u : E \to L_2) \cdot \delta^{-1/2} .$$

Optimizing finally over δ, the result follows.

- Lemma. Let u be an operator from some Banach space E into $C^\lambda[0,1]$ for some $0 < \lambda \le 1$. Then

$$e_n(u : E \to C) \le 2\|u : E \to C^\lambda\|^{\frac{1}{2\lambda+1}} \cdot e_n(u : E \to L_2)^{\frac{2\lambda}{2\lambda+1}} .$$

  Sketch of the proof.
  First we construct averaging operators $P_\delta : L_2 \to C$ with

$$\|P_\delta : L_2 \to C\| \le \delta^{-1/2} \quad \text{and} \quad \|id - P_\delta : C^\lambda \to C\| \le \delta^\lambda$$

  By elementary properties of entropy numbers this gives

$$e_n(u : E \to C) \le \|u : E \to C\| \cdot \delta^\lambda + e_n(u : E \to L_2) \cdot \delta^{-1/2} .$$

  Optimizing finally over δ, the result follows.

- Remark. 1. The direct proof is not very difficult and gives exact constants, so there is no need to use abstract interpolation theory.
  2. Taking logarithms, the exponent $\frac{2\lambda}{2\lambda+1}$ becomes a multiplicative constant, and this solves our problem: $\log e_n(S : L_2 \to C) \preceq -\sqrt[3]{n}$.

6. Covering numbers in Gaussian RKHSs

- The positive definite Gaussian kernels

$$K(x, y) = \exp(-\sigma^2 \|x - y\|_2^2) \quad , \quad x, y \in [0, 1]^d$$

where $\sigma > 0$ is an arbitrary parameter, play an important role in learning theory. They generate RKHSs

$$H_\sigma([0, 1]^d) \hookrightarrow C([0, 1]^d)\,.$$

Of special interest in learning theory are its covering numbers.

6. Covering numbers in Gaussian RKHSs

- The positive definite Gaussian kernels

$$K(x,y) = \exp(-\sigma^2 \|x-y\|_2^2) \quad , \quad x,y \in [0,1]^d$$

where $\sigma > 0$ is an arbitrary parameter, play an important role in learning theory. They generate RKHSs

$$H_\sigma([0,1]^d) \hookrightarrow C([0,1]^d)\,.$$

Of special interest in learning theory are its covering numbers.

- Theorem (Kühn, J. Complexity 2011). The covering numbers of the embedding $I_{\sigma,d} : H_\sigma([0,1]^d) \to C([0,1]^d)$ behave asymptotically like

$$\log \mathcal{N}(\varepsilon, I_{\sigma,d}) \asymp \frac{\left(\log \frac{1}{\varepsilon}\right)^{d+1}}{\left(\log\log \frac{1}{\varepsilon}\right)^{d}} \qquad \textit{as} \quad \varepsilon \to 0\,.$$

The same is true for $I_{\sigma,d} : H_\sigma([0,1]^d) \to L_p([0,1]^d)$, $2 \le p < \infty$.

- Remarks.

1. This improves earlier results of Ding-Xuan Zhou 2002/2003.

He showed $$(\log \frac{1}{\varepsilon})^{\frac{d}{2}} \preceq \log \mathcal{N}(\varepsilon, I_{\sigma,d}) \preceq (\log \frac{1}{\varepsilon})^{d+1}$$

and conjectured that the correct bound is $(\log \frac{1}{\varepsilon})^{\frac{d}{2}+1}$.

2. Our proof uses an explicit description of ONBs in Gaussian RKHSs, due to Steinwart/Hush/Scovel 2006.

- Remarks.

  1. This improves earlier results of Ding-Xuan Zhou 2002/2003.

  He showed $$(\log \tfrac{1}{\varepsilon})^{\frac{d}{2}} \preceq \log \mathcal{N}(\varepsilon, I_{\sigma,d}) \preceq (\log \tfrac{1}{\varepsilon})^{d+1}$$

  and conjectured that the correct bound is $(\log \frac{1}{\varepsilon})^{\frac{d}{2}+1}$.

  2. Our proof uses an explicit description of ONBs in Gaussian RKHSs, due to Steinwart/Hush/Scovel 2006.

- Application to smooth Gaussian processes.

  Let $\sigma > 0$ and $d \in \mathbb{N}$. For the centered Gaussian process $X_{\sigma,d} = (X_{\sigma,d}(t))$, $t \in [0,1]^d$ with covariance structure

  $$\mathbb{E}\, X_{\sigma,d}(s) X_{\sigma,d}(t) = \exp\left(-\sigma^2 \|s-t\|_2^2\right)$$

  the small deviation probabilities under the $\sup$-norm satisfy

  $$-\log \mathbb{P}\left(\sup_{t\in[0,1]^d} |X_{\sigma,d}(t)| \le \varepsilon\right) \sim \frac{\left(\log \frac{1}{\varepsilon}\right)^{d+1}}{\left(\log\log \frac{1}{\varepsilon}\right)^{d}} .$$

  The same estimates hold for all L_p-norms with $2 \le p < \infty$.

THANK YOU FOR YOUR ATTENTION!