Journal of Complexity 27 (2011) 489–499

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

Covering numbers of Gaussian reproducing kernel Hilbert spaces

Thomas Kühn

Mathematisches Institut, Universität Leipzig, Johannisgasse 26, D-04103 Leipzig, Germany

ARTICLE INFO

Article history:
Received 8 September 2010
Accepted 4 January 2011
Available online 22 January 2011

Keywords:
Covering numbers
Gaussian RKHS
Learning theory
Smooth Gaussian processes
Small deviations

ABSTRACT

Metric entropy quantities, like covering numbers or entropy numbers, and positive definite kernels play an important role in mathematical learning theory. Using smoothness properties of the Fourier transform of the kernels, Zhou [D.-X. Zhou, The covering number in learning theory, J. Complexity 18 (3) (2002) 739–767] proved an upper estimate for the covering numbers of the unit ball of Gaussian reproducing kernel Hilbert spaces (RKHSs), considered as a subset of the space of continuous functions.

In this note we determine the *exact asymptotic order* of these covering numbers, exploiting an explicit description of Gaussian RKHSs via orthonormal bases. We show that Zhou's estimate is almost sharp (up to a double logarithmic factor), but his conjecture on the correct asymptotic rate is far too optimistic. Moreover we give an application of our entropy results to small deviations of certain smooth Gaussian processes.

1. Introduction

The pioneering paper by Cucker and Smale [6] gave a new impetus to the statistical theory of learning, which studies how to "learn" unknown objects from random samples. In particular they demonstrated the important role of functional analytic methods in this context. For example, in order to estimate the probabilistic error and the number of samples required for a given confidence level and a given error bound, metric entropy quantities such as covering and entropy numbers are very useful, see also the monographs by Cucker and Zhou [7] and by Steinwart and Christmann [17] and the paper by Williamson et al. [21].

The concept of metric entropy is very basic and general, it has numerous applications in many other branches of mathematics, e.g. in approximation theory, probability theory (small deviation

E-mail address: kuehn@math.uni-leipzig.de.

0885-064X/$ – see front matter
doi:10.1016/j.jco.2011.01.005

problems for stochastic processes), operator theory (eigenvalue distributions of compact operators), PDEs (spectral theory of pseudodifferential operators). For more information on these subjects we refer to the articles by Kuelbs and Li [12] and Li and Linde [14] on small ball problems for Gaussian measures, the monographs by König [11] and Pietsch [16] on eigenvalues of compact operators in Banach spaces, and the book by Edmunds and Triebel [8] on function spaces and spectral theory of PDEs.

If A is a subset of a metric space M and $\varepsilon > 0$, the *covering number* $\mathcal{N}(\varepsilon, A; M)$ is defined as the minimal number of balls in M of radius ε which cover the set A. The centers of these balls form an *ε-net of A in M*. A possibly wider known but closely related notion is Kolmogorov's ε-entropy $\mathcal{H}(\varepsilon, A; M) := \log \mathcal{N}(\varepsilon, A; M)$, see e.g. [10]. Obviously,

$$A \text{ is precompact} \iff \mathcal{N}(\varepsilon, A; M) < \infty \quad \text{for all } \varepsilon > 0.$$

Thus the growth rate of $\mathcal{N}(\varepsilon, A; M)$ or $\mathcal{H}(\varepsilon, A; M)$ as $\varepsilon \to 0$ can be viewed as a measure of the "degree of compactness" or "massiveness" of the set A.

Many modern machine learning methods such as support vector machines use Gaussian radial basis functions, which generate a reproducing kernel Hilbert space. Motivated by these facts, Zhou [23] studied covering numbers of the unit ball of RKHSs, considered as a subset of the space of continuous functions. The results were expressed in terms of smoothness properties of the Fourier transform of the kernels. To illustrate his general results he gave an upper estimate for these covering numbers in the case of Gaussian RKHSs (see Example 4 on p. 761 ff. in [23]) and stated a conjecture about the exact asymptotic behaviour (p. 763 f. in [23]). For further results in this direction see also [24,20].

Using completely different methods we determine the exact asymptotic behaviour of these covering numbers. It turns out that Zhou's upper estimate is almost sharp, up to a double logarithmic factor, but his conjecture is too optimistic. Essential tools in our proof are recent results by Steinwart et al. [18] on the structure of Gaussian RKHSs, in particular we exploit the specific orthonormal bases (ONB) of these spaces given in [18]. A quite interesting detail of the proof of the lower bound is the fact that finite sections of the famous Hilbert matrix come into play.

As an application we obtain the sharp asymptotic rate of small deviation probabilities of certain smooth Gaussian processes. Here we use the close connection between metric entropy and small deviations that was discovered by Kuelbs and Li [12].

The organization of the paper is as follows. In Section 2 we describe the necessary background for our main result. We introduce covering numbers of (bounded linear) operators between Banach spaces, a variant of the covering numbers defined above, and state some simple properties that will be needed in the sequel. Moreover, we briefly recall some general facts from the theory of RKHSs and the result from [18] on ONBs in Gaussian RKHSs. In Section 3 we state and prove our main result on covering numbers, and Section 4 contains the application to small deviation probabilities.

2. Preliminaries

In this section we fix notation and recall some well-known basic facts concerning the two concepts mentioned in the title—covering numbers and reproducing kernel Hilbert spaces. Throughout the paper we consider only *real* Banach spaces, and "operator" always means "bounded linear operator between Banach spaces". The Euclidean norm in any $\mathbb{R}^d$ will be denoted by $\|\cdot\|_2$. For functions $f, g : (0, \infty) \to \mathbb{R}$ we write

$$f(\varepsilon) \sim g(\varepsilon) \text{(strong equivalence)}, \quad \text{if } \lim_{\varepsilon\to 0} \frac{f(\varepsilon)}{g(\varepsilon)} = 1 \quad \text{and}$$

$$f(\varepsilon) \asymp g(\varepsilon) \text{(weak equivalence)}, \quad \text{if } 0 < \liminf_{\varepsilon\to 0} \frac{f(\varepsilon)}{g(\varepsilon)} \le \limsup_{\varepsilon\to 0} \frac{f(\varepsilon)}{g(\varepsilon)} < \infty.$$

The same notation will be used for sequences and $n \to \infty$.