Entropy of Compact Operators with Applications to Landau-Pollak-Slepian Theory and Sobolev Spaces

Thomas Allard
tallard@mins.ee.ethz.ch

Helmut Bölcskei
hboelcskei@ethz.ch

Abstract

We derive a precise general relation between the entropy of a compact operator and its eigenvalues. It is then shown how this result along with the underlying philosophy can be applied to improve substantially on the best known characterizations of the entropy of the Landau-Pollak-Slepian operator and the metric entropy of unit balls in Sobolev spaces.

1 Introduction

Characterizing the metric entropy of function classes is a topic of longstanding interest in the mathematics and engineering literature, spanning across domains as diverse as approximation theory [1, 2], information theory [3, 4], statistics [5, 6], the study of dynamical systems [7, 8], and deep neural network theory [9, 10]. Perhaps somewhat less widely known are the related concepts of entropy and entropy numbers of linear compact operators between Banach spaces [11, 12, 13, 14], finding application in domains as varied as control theory [15], machine learning [16], and the study of Brownian motion [17].

Based on recent advances in the characterization of the metric entropy of ellipsoids by the authors of the present paper [18, 19], we derive a precise relationship between the entropy of a compact operator and the asymptotic behavior of its eigenvalues, thereby improving significantly upon the classical result [11, Proposition 1.3.2]. As a byproduct, we also obtain a relationship between the entropy of a compact operator and its eigenvalue-counting function.

Finally, it is demonstrated how our results along with the underlying general philosophy can be applied to improve substantially on the best known characterizations of the entropy of the Landau-Pollak-Slepian operator and the metric entropy of unit balls in Sobolev spaces.

1.1 Notation and terminology

We write $\mathbb{N}$ for the set of non-negative integers, $\mathbb{N}^*$ for the positive integers, $\mathbb{R}$ for the real numbers, and $\mathbb{R}_+^*$ for the positive real numbers. For $d \in \mathbb{N}^*$, we denote by ω_d the volume of the unit ball in $\mathbb{R}^d$ and by $\mathcal{H}^d$ the d-dimensional Hausdorff measure.

When comparing the asymptotic behavior of the functions f and g as $x \to \ell$, with $\ell \in \mathbb{R} \cup \{-\infty, \infty\}$, we use the standard notation $f = o_{x\to\ell}(g)$ to express that $\lim_{x\to\ell} \frac{f(x)}{g(x)} = 0$.

The authors gratefully acknowledge support by the Lagrange Mathematics and Computing Research Center, Paris, France.

We further indicate asymptotic equivalence according to $f \sim_{x\to\ell} g$ if $\lim_{x\to\ell} \frac{f(x)}{g(x)} = 1$ and $f \asymp_{x\to\ell} g$ if $\lim_{x\to\ell} \frac{f(x)}{g(x)} = C$ for some $C > 0$.

The space of square summable sequences is referred to as ℓ^2. Given a set $\Omega \subseteq \mathbb{R}^d$, we let $L^2(\Omega)$ be the space of square-integrable functions on Ω equipped with the usual inner product $\langle\cdot,\cdot\rangle_{L^2(\Omega)}$. We write $\mathrm{supp}(f)$ for the essential support of $f \in L^2(\Omega)$ and define $C_0^\infty(\Omega)$ to be the space of infinitely differentiable functions with compact support contained in Ω. Further, Id designates the identity operator and $\mathcal{F}$ the Fourier transform operator. Given a Banach space $(E, \|\cdot\|_E)$, we shall write $\mathcal{B}_E := \{x \in E \mid \|x\|_E \le 1\}$ for its closed unit ball.

Finally, log stands for the logarithm to base 2 and $\mathbb{1}_X(\cdot)$ for the indicator function corresponding to the set X.

2 Entropy of compact operators

We first introduce the notions of metric entropy and entropy numbers of a set.

Definition 1 (Metric entropy and entropy numbers of sets)**.** *Let $(\mathcal{X}, d)$ be a metric space and $\mathcal{K} \subseteq \mathcal{X}$ a compact set. An ε-covering of $\mathcal{K}$ with respect to the metric d is a set $\{x_1, \dots, x_N\} \subseteq \mathcal{X}$ such that for each $x \in \mathcal{K}$, there exists an $i \in \{1, \dots, N\}$ so that $d(x, x_i) \le \varepsilon$. The ε-covering number $N(\varepsilon; \mathcal{K}, d)$ is the cardinality of a smallest such ε-covering. The* metric entropy *of the set $\mathcal{K}$ is given by*

$$H(\varepsilon; \mathcal{K}, d) := \log N(\varepsilon; \mathcal{K}, d).$$

*For $m \in \mathbb{N}^*$, the m-*th entropy number *ε_m of $\mathcal{K}$ is defined as the smallest radius $\varepsilon > 0$ required to cover $\mathcal{K}$ with at most 2^m balls of radius ε, i.e.,*

$$\varepsilon_m(\mathcal{K}, d) := \inf\{\varepsilon > 0 \mid H(\varepsilon; \mathcal{K}, d) \le m\}.$$

The extension of the concepts in Definition 1 to compact linear operators $T\colon E \to F$ between Banach spaces E, F proceeds as follows. We first note that the image $T(\mathcal{B}_E)$ of the unit ball in E has compact closure in F. The entropy of the compact operator T, quantifying its compactness, is then simply given by the metric entropy of $\overline{T(\mathcal{B}_E)}$.

Definition 2 (Entropy and entropy numbers of compact linear operators)**.** *Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be Banach spaces, and let $T\colon E \to F$ be a compact linear operator. We define the entropy of T as the metric entropy of the closure of the image of the unit ball, according to*

$$H(\varepsilon; T) := H\Big(\varepsilon; \overline{T(\mathcal{B}_E)}, \|\cdot\|_F\Big).$$

Likewise, for $m \in \mathbb{N}^*$, the m-th entropy number of T is defined as the m-th entropy number of the closure of the image of the unit ball, i.e.,

$$\varepsilon_m(T) := \inf\Big\{\varepsilon > 0 \mid H\Big(\varepsilon; \overline{T(\mathcal{B}_E)}, \|\cdot\|_F\Big) \le m\Big\}.$$

Following the convention in the literature, we shall talk about the "metric entropy" of a set but will simply say the "entropy" of an operator. Although all the results in this paper can be stated in the broader setting of operators between general Banach spaces, for concreteness and simplicity of exposition, we restrict our attention to endomorphisms of separable real Hilbert spaces. Entropy in the complex case differs from that in the real case only by a multiplicative factor of 2. The restriction we impose further has the advantage of allowing direct comparisons with classical results, as e.g. those in [11, 15]. We shall henceforth consider a separable real Hilbert space $\mathcal{H}$ along with the compact linear operator $T\colon \mathcal{H} \to \mathcal{H}$. Arguing through polar decomposition as in [15] or [11, Chapter 3.4], we can further restrict our attention to self-adjoint operators, i.e., operators that can be diagonalized in some suitable basis; the corresponding eigenvalues will be denoted as $\{\lambda_n\}_{n\in\mathbb{N}^*}$. For expositional convenience, eigenvalues will be assumed ordered in a non-increasing fashion throughout the paper.

Relating the entropy numbers of compact linear operators to their eigenvalues has been a topic of longstanding interest. The corresponding results in the area are typically lower and upper bounds on the entropy numbers in terms of the geometric mean of the eigenvalues, see [12] for Banach spaces, [13, Theorem 1.3.4] for quasi-Banach spaces, and [15] for Hilbert spaces. To the best of our knowledge, the sharpest known result [11, Proposition 1.3.2] is

$$\sup_{N\in\mathbb{N}^*}\left\{2^{-m/N}\left[\prod_{n=1}^{N}\lambda_n\right]^{1/N}\right\} \leq \varepsilon_m(T) \leq 6 \sup_{N\in\mathbb{N}^*}\left\{2^{-m/N}\left[\prod_{n=1}^{N}\lambda_n\right]^{1/N}\right\}, \tag{1}$$

for all $m \in \mathbb{N}^*$.

The key to our approach is the simple observation that the image of the unit ball in $\mathcal{H}$ under T is an ellipsoid with semi-axes given by the eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}^*}$ of T. The strategy of using the metric entropy of ellipsoids to characterize the entropy numbers of compact linear operators has already been exploited in the literature, see e.g. [15]. However, recent progress on the characterization of the metric entropy of ellipsoids [18, 19], by the authors of this paper, allows for significant improvements. In particular, we obtain the following general result.

Theorem 3. *Let $\mathcal{H}$ be a separable real Hilbert space and let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator with eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}^*}$ satisfying*

$$\lambda_n = \frac{c_1}{n^{\alpha_1}} + \frac{c_2}{n^{\alpha_2}} + o_{n\to\infty}\left(\frac{1}{n^{\alpha_2}}\right), \tag{2}$$

where $c_1 \in \mathbb{R}_+^*$, $c_2 \in \mathbb{R}$, and $\alpha_1, \alpha_2 \in \mathbb{R}_+^*$ such that

$$\text{either}\quad \alpha_1 < \alpha_2 < \alpha_1\left(1 + \frac{2}{6\alpha_1 + 1}\right), \quad \text{or} \quad \begin{cases}\alpha_1 = \alpha_2, \text{ and}\\ c_2 = 0.\end{cases}$$

Then, the entropy of T satisfies

$$H(\varepsilon; T) = \frac{\alpha_1 {c_1}^{1/\alpha_1}}{\ln(2)}\varepsilon^{-1/\alpha_1} + \frac{c_2\, c_1^{\frac{1-\alpha_2}{\alpha_1}}}{(\alpha_1 + 1 - \alpha_2)\ln(2)}\varepsilon^{\frac{\alpha_2-\alpha_1-1}{\alpha_1}} + o_{\varepsilon\to 0}\left(\varepsilon^{\frac{\alpha_2-\alpha_1-1}{\alpha_1}}\right), \tag{3}$$

which can equivalently be expressed in terms of entropy numbers according to

$$\varepsilon_m(T) = c_1\left(\frac{\alpha_1}{\ln(2)}\right)^{\alpha_1} m^{-\alpha_1} + \frac{c_2\,(\alpha_1/\ln(2))^{\alpha_2}}{\alpha_1 + 1 - \alpha_2} m^{-\alpha_2} + o_{m\to\infty}(m^{-\alpha_2}). \tag{4}$$

Proof. As $\mathcal{H}$ is a separable real Hilbert space and T a compact self-adjoint operator, there exists an orthonormal basis $\{\psi_n\}_{n\in\mathbb{N}^*}$ of $\mathcal{H}$ composed of eigenvectors of T (see e.g. [20, Theorem 6.11]). Using the Bessel-Parseval identity, one obtains the following characterization of the image, under T, of the unit ball $\mathcal{B}$ of $\mathcal{H}$:

$$\begin{aligned} T(\mathcal{B}) &= \{y \in \mathcal{H} \mid y = Tx \text{ with } \|x\|_{\mathcal{H}}^2 \leq 1\} \\ &= \left\{y \in \mathcal{H} \mid y = \sum_{n=1}^{\infty} \lambda_n x_n \psi_n \text{ with } \{x_n\}_{n\in\mathbb{N}^*} \in \ell^2 \text{ s.t. } \sum_{n=1}^{\infty} |x_n|^2 \leq 1\right\} \\ &= \left\{y \in \mathcal{H} \mid y = \sum_{n=1}^{\infty} y_n \psi_n \text{ with } \{y_n\}_{n\in\mathbb{N}^*} \in \ell^2 \text{ s.t. } \sum_{n=1}^{\infty} |y_n/\lambda_n|^2 \leq 1\right\}. \end{aligned}$$

This shows that $T(\mathcal{B})$ is isometric to the ellipsoid in ℓ^2 of semi-axes $\{\lambda_n\}_{n\in\mathbb{N}^*}$. The result (3) then follows as a direct consequence of Lemma 8 in the Appendix.

Now, turning to entropy numbers, it follows from the definition of ε_m and (3) that

$$m = \frac{\alpha_1 {c_1}^{1/\alpha_1}}{\ln(2)} {\varepsilon_m}^{-1/\alpha_1} + \frac{c_2\, {c_1}^{\frac{1-\alpha_2}{\alpha_1}}}{(\alpha_1 + 1 - \alpha_2)\ln(2)} {\varepsilon_m}^{\frac{\alpha_2-\alpha_1-1}{\alpha_1}} + o_{m\to\infty}\left({\varepsilon_m}^{\frac{\alpha_2-\alpha_1-1}{\alpha_1}}\right).$$

Inverting this expression by application of Lemma 7 in the Appendix yields (4). □

In view of the applications considered in Section 3, we decided to restrict the statement of Theorem 3 to regularly varying eigenvalues (in the sense of [21, Definition 1.2.1]). An extension to exponentially decaying eigenvalues can be obtained by replacing the argument in the proof of Theorem 3 relying on Lemma 8 by [19, Theorem 9]. Moreover, we note that an extension to complex Hilbert spaces does not pose any technical difficulties.

As announced, we now show how Theorem 3 leads to a significant improvement of (1). In the case where $\lambda_n \sim c_1 n^{-\alpha_1}$, a direct calculation yields

$$\sup_{N\in\mathbb{N}^*} \left\{2^{-m/N} \left[\prod_{n=1}^{N} \lambda_n\right]^{1/N}\right\} = c_1 \left(\frac{\alpha_1}{\ln(2)}\right)^{\alpha_1} m^{-\alpha_1} + o_{m\to\infty}(m^{-\alpha_1}).$$

We can therefore conclude that (1) characterizes the first term only in the asymptotic expansion of $\varepsilon_m(T)$ and does so up to a multiplicative factor of 6, whereas our result (4) delivers a characterization of the first two terms and with precise constants.

Note that our result is asymptotic in nature, while most results in the literature, including (1), apply for arbitrary $m \in \mathbb{N}^*$. We hasten to add, however, that Theorem 3 does not presuppose a full characterization of the eigenvalues of T, but only requires information on their asymptotic behavior. This is a more widely used and realistic assumption. Typically, when the operator T is (pseudo-)differential, standard results in micro-local analysis (see e.g. [22, Chapter 30]) characterize the asymptotic behavior of the eigenvalue-counting function

$$M_T \colon \gamma \in (0, \infty) \mapsto \#\{n \in \mathbb{N}^* \mid \lambda_n \geq \gamma\} \in \mathbb{N}$$

in the asymptotic regime $\gamma \to 0$. The canonical example is given by the inverse of the Laplacian (used for instance in the study of Sobolev spaces), where the Weyl law for the Laplacian provides the asymptotic scaling of the eigenvalue-counting function (see Section 3.2 for more details). The next result is a consequence of Theorem 3 and illustrates how knowledge of the asymptotic behavior of $M_T(\gamma)$ leads to an asymptotic characterization of the entropy and entropy numbers of T.

Corollary 4. *Let $\mathcal{H}$ be a separable real Hilbert space and let $T\colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator with eigenvalue-counting function satisfying*

$$M_T(\gamma) = \kappa_1\gamma^{-\beta_1} + \kappa_2\gamma^{-\beta_2} + o_{\gamma\to 0}\Big(\gamma^{-\beta_2}\Big), \tag{5}$$

where $\kappa_1 \in \mathbb{R}_+^*$, $\kappa_2 \in \mathbb{R}$, and $\beta_1, \beta_2 \in \mathbb{R}_+^*$ such that

$$\text{either} \quad \beta_1\left(1 - \frac{2}{6+\beta_1}\right) < \beta_2 < \beta_1, \quad \text{or} \quad \begin{cases} \beta_1 = \beta_2, \text{ and} \\ \kappa_2 = 0. \end{cases} \tag{6}$$

Then, the entropy of T satisfies

$$H(\varepsilon; T) = \frac{\kappa_1}{\beta_1 \ln(2)}\varepsilon^{-\beta_1} + \frac{\kappa_2}{\beta_2 \ln(2)}\varepsilon^{-\beta_2} + o_{\varepsilon\to 0}\Big(\varepsilon^{-\beta_2}\Big),$$

which, upon defining $\beta^* := \beta_1/(1+\beta_1-\beta_2)$, can equivalently be expressed as

$$\varepsilon_m(T) = \left(\frac{\kappa_1}{\beta_1\ln(2)}\right)^{1/\beta_1} m^{-1/\beta_1} + \frac{\kappa_2}{\kappa_1\beta_2}\left(\frac{\kappa_1}{\beta_1\ln(2)}\right)^{1/\beta^*} m^{-1/\beta^*} + o_{m\to\infty}\Big(m^{-1/\beta^*}\Big).$$

Proof. Denote the eigenvalues of T, ordered in non-increasing fashion, by $\{\lambda_n\}_{n\in\mathbb{N}^*}$ and let $\{\eta_n\}_{n\in\mathbb{N}^*}$ be an arbitrary sequence of positive real numbers. From the definition of the eigenvalue-counting function, it follows that

$$M_T(\lambda_n + \eta_n) < n \leq M_T(\lambda_n), \quad \text{for all } n \in \mathbb{N}^*.$$

Invoking assumption (5) on the eigenvalue-counting function and choosing $\{\eta_n\}_{n\in\mathbb{N}^*}$ to decay to zero fast enough, we then obtain

$$n = \kappa_1\lambda_n^{-\beta_1} + \kappa_2\lambda_n^{-\beta_2} + o_{n\to\infty}\Big(\lambda_n^{-\beta_2}\Big).$$

Inverting this expression through application of Lemma 7 in the Appendix yields the decay behavior of the eigenvalues according to

$$\lambda_n = \kappa_1^{1/\beta_1} n^{-1/\beta_1} + \frac{\kappa_1^{1/\beta_1-\beta_2/\beta_1}\kappa_2}{\beta_1} n^{\beta_2/\beta_1-1/\beta_1-1} + o_{n\to\infty}\Big(n^{\beta_2/\beta_1-1/\beta_1-1}\Big).$$

Next, introducing the quantities

$$c_1 := \kappa_1^{1/\beta_1}, \quad c_2 := \frac{\kappa_1^{1/\beta_1-\beta_2/\beta_1}\kappa_2}{\beta_1}, \quad \alpha_1 := 1/\beta_1, \quad \text{and} \quad \alpha_2 := 1 + 1/\beta_1 - \beta_2/\beta_1,$$

allows us to reformulate the two cases in (6) as

$$\text{either} \quad \alpha_1 < \alpha_2 < \alpha_1\left(1 + \frac{2}{6\alpha_1+1}\right), \quad \text{or} \quad \begin{cases} \alpha_2 = \alpha_1, \text{ and} \\ c_2 = 0. \end{cases}$$

The hypotheses of Theorem 3 are thus verified and its application yields the desired result according to

$$H(\varepsilon; T) = \frac{\kappa_1}{\beta_1 \ln(2)}\varepsilon^{-\beta_1} + \frac{\kappa_2}{\beta_2 \ln(2)}\varepsilon^{-\beta_2} + o_{\varepsilon\to 0}\Big(\varepsilon^{-\beta_2}\Big)$$

and

$$\varepsilon_m(T) = \left(\frac{\kappa_1}{\beta_1\ln(2)}\right)^{1/\beta_1} m^{-1/\beta_1} + \frac{\kappa_2}{\kappa_1\beta_2}\left(\frac{\kappa_1}{\beta_1\ln(2)}\right)^{1/\beta^*} m^{-1/\beta^*} + o_{m\to\infty}\Big(m^{-1/\beta^*}\Big),$$

with $\beta^* := \beta_1/(1+\beta_1-\beta_2)$, thereby concluding the proof. □

3 Applications

We now put the general results developed in Section 2 to work. Concretely, we derive the entropy of the Landau-Pollak-Slepian operator and we find a precise asymptotic characterization of the metric entropy of unit balls in Sobolev spaces. In both cases, significant improvements over the best known results in the literature are obtained.

3.1 Entropy of the Landau-Pollak-Slepian operator

The classical sampling theorem [23] quantifies the minimum number of samples per unit of time needed to recover a strictly band-limited signal. This result essentially characterizes the information rate of band-limited signals. Landau, Pollak, and Slepian [24, 25] took this idea further by allowing for signals that are effectively band-and time-limited. The object of central interest in this theory is the Landau-Pollak-Slepian operator defined as follows. For $r \in \mathbb{R}_+^*$ and compact subsets Ω and $\mathcal{W}$ of $\mathbb{R}^d$, one considers the sets

$$\begin{aligned} \mathcal{D}(r\Omega) &\coloneqq \Big\{ f \in L^2(\mathbb{R}^d) \mid \operatorname{supp}(f) \subseteq r\Omega \Big\}, \quad \text{and} \\ \mathcal{F}(\mathcal{W}) &\coloneqq \Big\{ f \in L^2(\mathbb{R}^d) \mid \operatorname{supp}(\mathcal{F}f) \subseteq \mathcal{W} \Big\}. \end{aligned}$$

Associating an orthogonal projection operator with each of these sets according to

$$P_{r\Omega} \colon f \mapsto \mathbb{1}_{\{r\Omega\}} f \quad \text{and} \quad P_{\mathcal{W}} \colon f \mapsto \mathcal{F}^{-1} \mathbb{1}_{\{\mathcal{W}\}} \mathcal{F} f,$$

leads to the definition of the Landau-Pollak-Slepian operator as

$$P_{LPS}^{(r)} \coloneqq P_{r\Omega} P_{\mathcal{W}} P_{r\Omega} \colon L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d).$$

We refer to [26, Chapter 2] and [27, Chapter 20] for detailed discussions of the Landau-Pollak-Slepian operator and to [28] for an application to the derivation of uncertainty principles.

Next, we apply the method developed in the previous section to obtain an exact characterization of the entropy rate $\lim_{r\to\infty} H(\varepsilon; P_{LPS}^{(r)})/r^d$ of the Landau-Pollak-Slepian operator, based on which an asymptotic result by Kolmogorov and Tikhomirov on the entropy rate of effectively band-and time-limited signals can be turned into a non-asymptotic statement. To the best of our knowledge, the entropy (rate) of the Landau-Pollak-Slepian operator has not been characterized before in the literature.

Theorem 5. *Let $d \in \mathbb{N}^*$ and let Ω and $\mathcal{W}$ be compact subsets of $\mathbb{R}^d$. Then, we have*

$$\lim_{r\to\infty} \frac{H\Big(\varepsilon; P_{LPS}^{(r)}\Big)}{r^d} = \frac{2\,\mathcal{H}^d(\Omega)\,\mathcal{H}^d(\mathcal{W})}{(2\pi)^d} \log\big(\varepsilon^{-1}\big), \quad \textit{for all } \varepsilon \in (0,1].$$

The proof of Theorem 5 proceeds by first reducing the problem at hand to covering an infinite-dimensional ellipsoid with semi-axes determined by the eigenvalues of the Landau-Pollak-Slepian operator. Indeed, it is well known (see e.g. [29, Lemma 1 and Theorem 1]) that the Landau-Pollak-Slepian operator is compact, self-adjoint, and has positive eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}^*}$ bounded by 1 with associated eigenvalue-counting function satisfying

$$M_r(\gamma) = \left(\frac{r}{2\pi}\right)^d \mathcal{H}^d(\Omega)\,\mathcal{H}^d(\mathcal{W}) + o_{r\to\infty}\Big(r^d\Big), \quad \text{for all } \gamma \in (0,1). \tag{7}$$

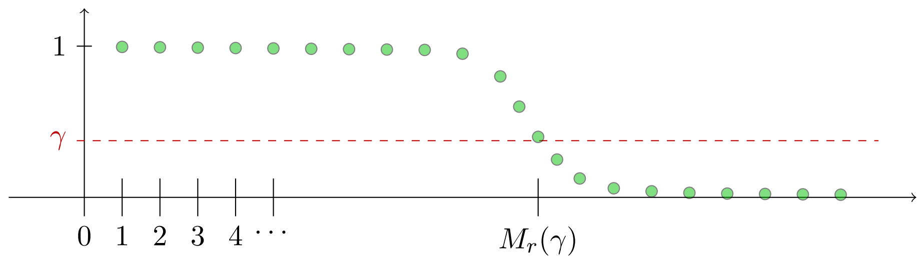

Figure 1: Eigenvalue distribution of the Landau-Pollak-Slepian operator.

In a second step, the problem is further reduced, namely to covering a finite-dimensional ellipsoid obtained by carefully thresholding the infinite-dimensional ellipsoid under consideration. Specifically, the threshold specifies the ε-dependent effective dimension of the infinite-dimensional ellipsoid and is chosen as $M_r(\gamma)$ for γ suitably depending on ε, see Figure 1. Finally, we apply results from [19] on the covering of finite-dimensional ellipsoids.

Proof. The proof will be effected by sandwiching the entropy rate between matching lower and upper bounds. To this end, let us fix $r \in \mathbb{R}_+^*$, $\varepsilon \in (0,1]$, $\gamma \in (0,1)$, and denote the image of the unit ball in $L^2(\mathbb{R}^d)$ under the operator $P_{LPS}^{(r)}$ by $\mathcal{E}^{(r)}$. We have already argued that, in the basis of eigenvectors of $P_{LPS}^{(r)}$, the set $\mathcal{E}^{(r)}$ is isometric to the infinite-dimensional ellipsoid in ℓ^2 of semi-axes given by the eigenvalues of $P_{LPS}^{(r)}$, denoted as $\{\lambda_n\}_{n\in\mathbb{N}^*}$. The associated eigenvalue-counting function satisfies (7). We further let $\mathcal{E}_-^{(r)}$ stand for the finite-dimensional ellipsoid obtained from $\mathcal{E}^{(r)}$ by retaining the $M_r(\gamma)$ largest semi-axes. (Note that $M_r(\gamma)$ is guaranteed to be non-zero for r large enough.) As covering the infinite-dimensional ellipsoid $\mathcal{E}^{(r)}$ requires at least as many covering balls as needed to cover the corresponding finite-dimensional ellipsoid $\mathcal{E}_-^{(r)}$, we have

$$H\Big(\varepsilon;\mathcal{E}^{(r)},\|\cdot\|_2\Big) \geq H\Big(\varepsilon;\mathcal{E}_-^{(r)},\|\cdot\|_2\Big) \tag{8}$$

(for a formalization of this argument, we refer to [19, Lemma 10]). Now, a direct application of Lemma 9 in the Appendix yields

$$\frac{H\Big(\varepsilon;\mathcal{E}_-^{(r)},\|\cdot\|_2\Big)}{2M_r(\gamma)} \geq \log\big(\varepsilon^{-1}\big) + \frac{1}{M_r(\gamma)}\sum_{n=1}^{M_r(\gamma)}\log(\lambda_n). \tag{9}$$

Note that, by definition of $M_r(\gamma)$, the eigenvalues λ_n appearing in (9) are all greater than or equal to γ. This yields the bound

$$\frac{1}{M_r(\gamma)}\sum_{n=1}^{M_r(\gamma)}\log(\lambda_n) \geq \log(\gamma). \tag{10}$$

Now, combining (7)-(10), we get

$$\begin{aligned}\lim_{r\to\infty}\frac{H\big(\varepsilon;\mathcal{E}^{(r)},\|\cdot\|_2\big)}{r^d} &\geq \lim_{r\to\infty}\frac{H\Big(\varepsilon;\mathcal{E}_-^{(r)},\|\cdot\|_2\Big)}{r^d}\\ &\geq 2\log\big(\gamma\,\varepsilon^{-1}\big)\lim_{r\to\infty}\frac{M_r(\gamma)}{r^d}\\ &= \frac{2\,\mathcal{H}^d(\Omega)\,\mathcal{H}^d(\mathcal{W})}{(2\pi)^d}\log\big(\gamma\,\varepsilon^{-1}\big).\end{aligned}$$

In particular, taking γ arbitrarily close to 1, it follows that

$$\lim_{r\to\infty} \frac{H\big(\varepsilon;\mathcal{E}^{(r)},\|\cdot\|_2\big)}{r^d} \geq \frac{2\,\mathcal{H}^d(\Omega)\,\mathcal{H}^d(\mathcal{W})}{(2\pi)^d}\log\big(\varepsilon^{-1}\big). \tag{11}$$

The proof will be completed by establishing an upper bound on the entropy rate matching the lower bound (11). To this end, fix $\tau \in (0,1)$ and consider the ellipsoid $\mathcal{E}_+^{(r)}$ obtained from $\mathcal{E}^{(r)}$ by retaining the $M_r(\tau\varepsilon)$ largest semi-axes. Note that $M_r(\tau\varepsilon)$ is well-defined by $\tau\varepsilon \in (0,1)$, and guaranteed to be non-zero for r large enough. As the semi-axes corresponding to the dimensions not retained in the transition from $\mathcal{E}^{(r)}$ to $\mathcal{E}_+^{(r)}$ have length smaller than $\tau\varepsilon$, every $(1-\tau)\varepsilon$-covering of $\mathcal{E}_+^{(r)}$ can be turned into an ε-covering of $\mathcal{E}^{(r)}$, simply by completing the components of the covering ball centers of $\mathcal{E}_+^{(r)}$ by an infinite sequence of zeros (for a formalization of these arguments, we refer to the techniques developed in [19, Lemmata 10, 11, and 42]). This observation translates into

$$H\Big(\varepsilon;\mathcal{E}^{(r)},\|\cdot\|_2\Big) \leq H\Big((1-\tau)\varepsilon;\mathcal{E}_+^{(r)},\|\cdot\|_2\Big) \leq H\Big((1-\tau)\varepsilon;\mathcal{B}^{(r)},\|\cdot\|_2\Big), \tag{12}$$

where $\mathcal{B}^{(r)}$ denotes the unit ball of dimension $M_r(\tau\varepsilon)$ and the second inequality follows from the fact that the semi-axes have length smaller than or equal to one (which is a direct consequence of the semi-axes being given by the eigenvalues of $P_{LPS}^{(r)}$). From Lemma 10 in the Appendix with $d = 2M_r(\tau\varepsilon)$, we can deduce the existence of a sequence κ with $\lim_{M\to\infty}\kappa(M) = 1$, such that

$$\frac{H\big((1-\tau)\varepsilon;\mathcal{B}^{(r)},\|\cdot\|_2\big)}{2M_r(\tau\varepsilon)} \leq \log\Big(((1-\tau)\varepsilon)^{-1}\Big) + \log(\kappa(2M_r(\tau\varepsilon))). \tag{13}$$

Using (13) in (12) together with (7), we obtain, for all $\tau \in (0,1)$,

$$\begin{aligned}
\lim_{r\to\infty} \frac{H\big(\varepsilon;\mathcal{E}^{(r)},\|\cdot\|_2\big)}{r^d} &\leq \lim_{r\to\infty} \frac{H\big((1-\tau)\varepsilon;\mathcal{B}^{(r)},\|\cdot\|_2\big)}{r^d} \\
&\leq \lim_{r\to\infty} \frac{2M_r(\tau\varepsilon)}{r^d}\Big\{\log\Big([(1-\tau)\varepsilon]^{-1}\Big) + \log(\kappa(2M_r(\tau\varepsilon)))\Big\} \\
&= \frac{2\,\mathcal{H}^d(\Omega)\,\mathcal{H}^d(\mathcal{W})}{(2\pi)^d}\log\Big([(1-\tau)\varepsilon]^{-1}\Big),
\end{aligned}$$

where the last step is by $\lim_{M\to\infty}\kappa(M) = 1$ combined with $\lim_{r\to\infty} M_r(\gamma) = \infty$. We can finally choose τ arbitrarily small to obtain

$$\lim_{r\to\infty} \frac{H\big(\varepsilon;\mathcal{E}^{(r)},\|\cdot\|_2\big)}{r^d} \leq \frac{2\,\mathcal{H}^d(\Omega)\,\mathcal{H}^d(\mathcal{W})}{(2\pi)^d}\log\big(\varepsilon^{-1}\big).$$

This concludes the proof. □

We now discuss the implications of Theorem 5 for $d = 1$, $\Omega = [-1,1]$, and $\mathcal{W} = [-\sigma,\sigma], \sigma \in \mathbb{R}_+^*$. In this case, the image of the unit ball in $L^2(\mathbb{R})$ under $P_{LPS}^{(T)}$, with $T \in \mathbb{R}_+^*$, is obtained by localizing strictly band-limited (namely to $\mathcal{W}$) functions to the time-interval $[-T,T]$. We denote the resulting function class by $B_\sigma^{(T)}$ and recall the following result due to Kolmogorov and Tikhomirov [30, Chapter 7, Theorem XXII], [31, Theorem 8]

$$\lim_{T\to\infty} \frac{H\Big(\varepsilon;B_\sigma^{(T)},\|\cdot\|_2\Big)}{2T} \sim_{\varepsilon\to 0} \frac{2\sigma}{\pi}\log\big(\varepsilon^{-1}\big). \tag{14}$$

This shows that for signals band-limited to $\mathcal{W} = [-\sigma, \sigma]$, as $\varepsilon \to 0$, the number of degrees of freedom per unit of time is given by $\frac{2\sigma}{\pi}$, and hence determined by the bandwidth 2σ. The rationale behind the interpretation in terms of degrees of freedom derives itself from the observation that the metric entropy of finite intervals on the reals is of order $\log(\varepsilon^{-1})$, as $\varepsilon \to 0$. Therefore, the pre-log on the right-hand side of (14) determines the number of information-carrying scalars per unit of time. Now, this interpretation holds only asymptotically as $\varepsilon \to 0$. Consequently, for $\varepsilon \in (0, 1]$, in principle the effective number of degrees of freedom could depend on ε, of course in a manner ensuring compatibility with the asymptotics in (14).

The result in Theorem 5, i.e.,

$$\lim_{T\to\infty} \frac{H\Big(\varepsilon; B_\sigma^{(T)}, \|\cdot\|_2\Big)}{2T} = \frac{2\sigma}{\pi} \log\big(\varepsilon^{-1}\big), \quad \text{for all } \varepsilon \in (0,1], \tag{15}$$

therefore constitutes a substantial improvement over the literature as it proves equality in (14) for all $\varepsilon \in (0, 1]$. Specifically, this shows that the dimension counting argument that has been employed in the literature for decades is, in fact, exact for all $\varepsilon \in (0, 1]$.

3.2 Metric entropy of unit balls in Sobolev spaces

Characterizing the metric entropy of unit balls in function spaces defined through regularity constraints such as Besov, Sobolev, Hölder, or Lipschitz spaces is a prominent problem in approximation theory and other domains of applied mathematics (see e.g. [30, Chapter 7]). The results in this context are typically of the following form. Function classes living on a d-dimensional domain and exhibiting smoothness of degree k have their metric entropy and entropy numbers scale according to

$$H(\varepsilon) \asymp_{\varepsilon\to 0} \varepsilon^{-\frac{d}{k}}, \quad \text{and} \quad \varepsilon_m \asymp_{m\to\infty} m^{-\frac{k}{d}}.$$

We refer to [18], [9, Table 1], and [4, 32, 33] for concrete incarnations of this general behavior.

We next show how our general results on the entropy of linear operators can be applied to fully characterize the first terms in the asymptotic expansion of the metric entropy of unit balls in Sobolev spaces. Concretely, given a bounded open set $\Omega \subset \mathbb{R}^d$, we consider the Sobolev space $W_0^{k,2}(\Omega)$ of regularity $k \in \mathbb{N}^*$ (see [34, Chapter 3] or [20, Chapter 9.4] for rigorous definitions) equipped with the norm

$$\|\cdot\|_{k,\Omega} \colon f \longmapsto \left[\|f\|^2_{L^2(\Omega)} + \sum_{|\alpha|=k} \|D^\alpha f\|^2_{L^2(\Omega)} \right]^{1/2}, \tag{16}$$

where we used the standard multi-index notation for $\alpha = (\alpha_1, \dots, \alpha_d)$ and $D^\alpha = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d}$, with $|\alpha| = \sum_{j=1}^d \alpha_j$. The best known result is due to Donoho (see [35] or the remark after [36, Corollary 2.4]) and provides, in the one-dimensional case with $\Omega = (0, 2\pi)$, the constant in the leading term of the asymptotic expansion of the metric entropy $H(\varepsilon)$ of the unit ball in $W_0^{k,2}(\Omega)$ equipped with the norm (16) according to

$$H(\varepsilon) \sim_{\varepsilon\to 0} \frac{2\,k}{\ln(2)}\, \varepsilon^{-\frac{1}{k}}. \tag{17}$$

In Theorem 6 below, we extend this result in two aspects. First, we allow for general bounded open domains $\Omega \subset \mathbb{R}^d$ in arbitrary finite dimensions d. Second, we provide, under certain regularity conditions on $\Omega \subset \mathbb{R}^d$, an exact characterization of the second term in the asymptotic expansion of metric entropy. Both of these extensions have no counterpart in the existing literature.

The proof of our result relies on a spectral analysis of the Laplacian $-\Delta := -\partial_1^2 - \cdots - \partial_d^2$. More concretely, we resort to Weyl's law for the Laplacian to characterize the asymptotic behavior of its eigenvalue-counting function for bounded domains Ω with smooth boundary $\partial\Omega$ and such that the measure of all periodic billiards is zero. We refer to [37], [38, Corollary 29.3.4], [39, Chapter 1.2], and the survey [40] for a detailed discussion of Weyl's law and the technical condition we require on $\partial\Omega$. To the best of our knowledge, Weyl's law has not been used before in the study of the metric entropy of function classes.

Theorem 6. *Let $d, k \in \mathbb{N}^*$ and let $\Omega \subset \mathbb{R}^d$ be a bounded open subset of $\mathbb{R}^d$. For a given set $S \subset \mathbb{R}^d$, we define a rescaled version of its r-dimensional Hausdorff measure $\mathcal{H}^r(S)$, $r \in \{1, \ldots, d\}$, according to*

$$\chi_r(S) := \frac{\omega_r}{r\,(2\pi)^r\,\ln(2)}\,\mathcal{H}^r(S). \tag{18}$$

Then,

(i) the metric entropy and the entropy numbers of the unit ball in $W_0^{k,2}(\Omega)$ equipped with the norm (16) satisfy

$$H(\varepsilon) = k\,\chi_d(\Omega)\,\varepsilon^{-\frac{d}{k}} + o_{\varepsilon\to 0}\left(\varepsilon^{-\frac{d}{k}}\right)$$

and

$$\varepsilon_m = (k\,\chi_d(\Omega))^{\frac{k}{d}} m^{-\frac{k}{d}} + o_{m\to\infty}\left(m^{-\frac{k}{d}}\right).$$

(ii) if we further assume that $d > 6k/(2k-1)$ and the boundary $\partial\Omega$ is smooth and such that the measure of the periodic billiards in Ω is zero, we have

$$H(\varepsilon) = k\,\chi_d(\Omega)\,\varepsilon^{-\frac{d}{k}} - \frac{k\,\chi_{d-1}(\partial\Omega)}{4}\,\varepsilon^{-\frac{d-1}{k}} + o_{\varepsilon\to 0}\left(\varepsilon^{-\frac{d-1}{k}}\right),$$

and

$$\varepsilon_m = (k\,\chi_d(\Omega))^{\frac{k}{d}} m^{-\frac{k}{d}} - \frac{k\,\chi_{d-1}(\partial\Omega)}{4\,d\,\chi_d(\Omega)}\,(k\,\chi_d(\Omega))^{\frac{k+1}{d}}\,m^{-\frac{k+1}{d}} + o_{m\to\infty}\left(m^{-\frac{k+1}{d}}\right).$$

The statements in Theorem 6 hold identically when the norm (16) is replaced by the equivalent norm

$$\|\cdot\|'_{k,\Omega} \colon f \longmapsto \left[\sum_{|\alpha|=k} \|D^\alpha f\|^2_{L^2(\Omega)}\right]^{1/2}. \tag{19}$$

We refer to [34, Corollary 6.31] for a proof of the equivalence of the norms (16) and (19). Our choice to state Theorem 6 in terms of the norm (16) is motivated by the desire to be compatible with (17). Indeed, for $d = 1$ and $\Omega = (0, 2\pi)$, part (i) of Theorem 6 recovers (17) according to

$$H(\varepsilon) \sim_{\varepsilon\to 0} k\,\chi_1((0,2\pi))\,\varepsilon^{-\frac{1}{k}} \sim_{\varepsilon\to 0} k\,\frac{2\cdot 2\pi}{1\cdot(2\pi)^1\ln(2)}\,\varepsilon^{-\frac{1}{k}} \sim_{\varepsilon\to 0} \frac{2\,k}{\ln(2)}\,\varepsilon^{-\frac{1}{k}}.$$

We finally note that a necessary condition for the assumption $d > 6k/(2k-1)$ in part (ii) of Theorem 6 to hold is $d > 3$ and a sufficient condition is $d > 6$. In particular, this assumption is readily satisfied for functions defined on high-dimensional domains, as is common e.g. in machine learning applications.

Proof. We first introduce the compact operator

$$T := \left[\mathrm{Id} + (-\Delta)^{(k)}\right]^{-1}, \tag{20}$$

where $(-\Delta)^{(k)}$ stands for the k-fold application of the Laplacian and note that, for all $f \in C_0^\infty(\Omega)$,

$$\|f\|_{k,\Omega} = \sqrt{\langle f \mid T^{-1} f\rangle_{L^2(\Omega)}}. \tag{21}$$

By density of $C_0^\infty(\Omega)$ in $W_0^{k,2}(\Omega)$ (see [20, Chapter 9, Remark 18]), we can conclude from (21) that the unit ball under the Sobolev norm $\|\cdot\|_{k,\Omega}$ is an ellipsoid in $L^2(\Omega)$ with semi-axes given by the square root of the eigenvalues of T. Denote the eigenvalue-counting function of T by M_T and let

$$M_\Delta\colon \gamma \mapsto \#\{n \in \mathbb{N}^* \mid \lambda_n \le \gamma, \text{ where } \{\lambda_n\}_{n\in\mathbb{N}^*} \text{ are the eigenvalues of } -\Delta\}.$$

Note that M_Δ counts the number of eigenvalues of Δ below a given threshold, whereas M_T counts those above the threshold. From the definition of T in (20), it follows that

$$M_T(\gamma^2) = M_\Delta\left(\sqrt[k]{\gamma^{-2} - 1}\right).$$

Note that we are interested in $M_T(\gamma^2)$ rather than $M_T(\gamma)$ as the semi-axes of the ellipsoids under consideration are given by the square root of the eigenvalues of T. The Weyl law for the Laplacian ([41, Chapter 9.5]) yields

$$\begin{aligned} M_T\left(\gamma^2\right) &= \frac{\omega_d \mathcal{H}^d(\Omega)}{(2\pi)^d}\left[\gamma^{-2} - 1\right]^{\frac{d}{2k}} + o_{\gamma\to 0}\left(\left[\gamma^{-2} - 1\right]^{\frac{d}{2k}}\right) \\ &= \frac{\omega_d \mathcal{H}^d(\Omega)}{(2\pi)^d}\gamma^{-\frac{d}{k}} + o_{\gamma\to 0}\left(\gamma^{-\frac{d}{k}}\right). \end{aligned}$$

Upon application of Corollary 4, with the choices

$$\kappa_1 = \frac{\omega_d \mathcal{H}^d(\Omega)}{(2\pi)^d}, \quad \kappa_2 = 0, \quad \text{and} \quad \beta_1 = \beta_2 = \frac{d}{k},$$

we obtain the first desired result, namely

$$H(\varepsilon) = \frac{k\,\omega_d\,\mathcal{H}^d(\Omega)}{d\,(2\pi)^d\,\ln(2)}\varepsilon^{-\frac{d}{k}} + o_{\varepsilon\to 0}\left(\varepsilon^{-\frac{d}{k}}\right)$$

and

$$\varepsilon_m = \left[\frac{k\,\omega_d\,\mathcal{H}^d(\Omega)}{d\,(2\pi)^d\,\ln(2)}\right]^{\frac{k}{d}} m^{-\frac{k}{d}} + o_{m\to\infty}\left(m^{-\frac{k}{d}}\right).$$

Under the assumptions of (ii) in the theorem statement, the two-term Weyl law for the Laplacian (see [37]) reads

$$\begin{aligned} M_T\left(\gamma^2\right) &= \frac{\omega_d \mathcal{H}^d(\Omega)}{(2\pi)^d}\left[\gamma^{-2} - 1\right]^{\frac{d}{2k}} \\ &\quad - \frac{\omega_{d-1}\mathcal{H}^{d-1}(\partial\Omega)}{4(2\pi)^{d-1}}\left[\gamma^{-2} - 1\right]^{\frac{d-1}{2k}} + o_{\gamma\to 0}\left(\gamma^{-\frac{d-1}{k}}\right) \\ &= \frac{\omega_d \mathcal{H}^d(\Omega)}{(2\pi)^d}\gamma^{-\frac{d}{k}} - \frac{\omega_{d-1}\mathcal{H}^{d-1}(\partial\Omega)}{4(2\pi)^{d-1}}\gamma^{-\frac{d-1}{k}} + o_{\gamma\to 0}\left(\gamma^{-\frac{d-1}{k}}\right). \end{aligned}$$

It can readily be verified that, under the choice

$$\kappa_1 = \frac{\omega_d \mathcal{H}^d(\Omega)}{(2\pi)^d}, \quad \kappa_2 = -\frac{\omega_{d-1} \mathcal{H}^{d-1}(\partial\Omega)}{4(2\pi)^{d-1}}, \quad \beta_1 = \frac{d}{k}, \quad \text{and} \quad \beta_2 = \frac{d-1}{k}, \tag{22}$$

the assumption $d > 6k/(2k-1)$ implies $\beta_1(1 - 2/(6+\beta_1)) < \beta_2$. We can hence apply Corollary 4 with the choice (22) to obtain the second desired result according to

$$H(\varepsilon) = \frac{k\, \omega_d\, \mathcal{H}^d(\Omega)}{d\,(2\pi)^d\, \ln(2)} \varepsilon^{-\frac{d}{k}} - \frac{k\, \omega_{d-1}\, \mathcal{H}^{d-1}(\partial\Omega)}{4\,(d-1)\,(2\pi)^{d-1}\, \ln(2)} \varepsilon^{-\frac{d-1}{k}} + o_{\varepsilon\to 0}\Big(\varepsilon^{-\frac{d-1}{k}}\Big)$$

and

$$\begin{aligned} \varepsilon_m = & \left[\frac{k\, \omega_d\, \mathcal{H}^d(\Omega)}{d\,(2\pi)^d\, \ln(2)}\right]^{\frac{k}{d}} m^{-\frac{k}{d}} \\ & - \frac{\pi\, k\, \omega_{d-1} \mathcal{H}^{d-1}(\partial\Omega)}{2\,(d-1)\,\omega_d\, \mathcal{H}^d(\Omega)} \left(\frac{k\, \omega_d \mathcal{H}^d(\Omega)}{d\,(2\pi)^d\, \ln(2)}\right)^{\frac{k+1}{d}} m^{-\frac{k+1}{d}} + o_{m\to\infty}\Big(m^{-\frac{k+1}{d}}\Big), \end{aligned}$$

thereby concluding the proof. □

References

[1] G. G. Lorentz, *Approximation of Functions.* Athena Series, 1966, vol. 48.

[2] ——, "Metric entropy and approximation," *Bulletin of the American Mathematical Society*, vol. 72, no. 6, pp. 903–937, 1966.

[3] Y. Yang and A. Barron, "Information-theoretic determination of minimax rates of convergence," *Annals of Statistics*, vol. 27, no. 5, 1999.

[4] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Transactions on Information Theory*, vol. 44, no. 6, 1998.

[5] R. M. Dudley, "Universal Donsker classes and metric entropy," *Annals of Probability*, vol. 15, no. 4, 1987.

[6] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[7] G. Zames, "On the metric complexity of causal linear systems: Estimates of ε-entropy and ε-dimension for continuous time," *IEEE*, pp. 807–810, 1977.

[8] G. Zames and J. Owen, "A note on metric dimension and feedback in discrete time," *IEEE Transactions on Automatic Control*, vol. 38, no. 4, pp. 664–667, 1993.

[9] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei, "Deep neural network approximation theory," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2581–2623, 2021.

[10] P. Grohs, A. Klotz, and F. Voigtlaender, "Phase transitions in rate distortion theory and deep learning," *Foundations of Computational Mathematics*, vol. 23, pp. 329–392, 2021.

[11] B. Carl and I. Stephani, *Entropy, Compactness and the Approximation of Operators*, 1st ed. Cambridge University Press, 1990.

[12] B. Carl and H. Triebel, "Inequalities between eigenvalues, entropy numbers, and related quantities of compact operators in Banach spaces," *Mathematische Annalen*, vol. 251, no. 2, pp. 129–133, 1980.

[13] D. E. Edmunds and H. Triebel, *Function Spaces, Entropy Numbers, Differential Operators*, 1st ed. Cambridge University Press, 1996.

[14] H. König, *Eigenvalue Distribution of Compact Operators*, ser. Operator Theory: Advances and Applications. Birkhäuser Basel, 1986, vol. 16.

[15] R. T. Prosser, "The ε-entropy and ε-capacity of certain time-varying channels," *Journal of Mathematical Analysis and Applications*, vol. 16, pp. 553–573, 1966.

[16] R. C. Williamson, A. J. Smola, and B. Schölkopf, "Generalization performance of regularization networks and support vector machines via entropy numbers of compact operators," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2516–2532, 2001.

[17] M. A. Lifshits and W. Linde, *Approximation and Entropy Numbers of Volterra Operators with Application to Brownian Motion*, ser. Memoirs AMS 745. American Mathematical Society, 2002.

[18] T. Allard and H. Bölcskei, "Covering ellipsoids in infinite dimensions," *in preparation*, 2024.

[19] ——, "Ellipsoid methods for metric entropy computation," *submitted to Constructive Approximation*, 2024. [Online]. Available: https://arxiv.org/pdf/2405.11066

[20] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York, 2011.

[21] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*, 1st ed. Cambridge University Press, 1987.

[22] M. A. Shubin, *Pseudodifferential Operators and Spectral Theory*. Springer Berlin, Heidelberg, 2001.

[23] C. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[24] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty - II," *Bell System Technical Journal*, vol. 40, no. 1, pp. 65–84, 1961.

[25] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty - I," *Bell System Technical Journal*, vol. 40, no. 1, pp. 43–63, 1961.

[26] I. Daubechies, *Ten Lectures on Wavelets*, ser. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992, no. 61.

[27] M. W. Wong, *Wavelet Transforms and Localization Operators.* Birkhäuser Basel, 2002.

[28] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[29] H. J. Landau, "On Szegö's eigenvalue distribution theorem and non-Hermitian kernels," *Journal d'Analyse Mathématique*, vol. 28, no. 1, pp. 335–357, 1975.

[30] A. N. Shiryayev, Ed., *Selected works of A. N. Kolmogorov*, ser. Mathematics and Its Applications. Springer Netherlands, 1993, vol. 27.

[31] D. Jagerman, "ε-entropy and approximation of bandlimited functions," *SIAM Journal on Applied Mathematics*, vol. 17, no. 2, pp. 362–377, 1969.

[32] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, "Tree approximation and optimal encoding," *Applied and Computational Harmonic Analysis*, vol. 11, no. 2, pp. 192–226, 2001.

[33] H. Triebel, *Theory of Function Spaces III*, ser. Monographs in Mathematics. Birkhäuser Basel, 2006, vol. 100.

[34] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces.* Elsevier, 2003.

[35] D. L. Donoho, "Counting bits with Kolmogorov and Shannon," *https://www.researchgate.net/publication/2326699*, 2001.

[36] H. Luschgy and G. Pagès, "Sharp asymptotics of the Kolmogorov entropy for Gaussian measures," *Journal of Functional Analysis*, vol. 212, no. 1, pp. 89–120, 2004.

[37] V. Ivrii, "Second term of the spectral asymptotic expansion of the Laplace - Beltrami operator on manifolds with boundary," *Functional Analysis and Its Applications*, vol. 14, no. 2, pp. 98–106, 1980.

[38] L. Hörmander, *The Analysis of Linear Partial Differential Operators IV*, ser. Classics in Mathematics. Springer Berlin, Heidelberg, 2009.

[39] L. Geisinger, "On the semiclassical limit of the dirichlet laplace operator: Two-term spectral asymptotics and sharp spectral estimates," Ph.D. dissertation, Institut für Analysis, Dynamik und Modellierung Universität Stuttgart, 2011. [Online]. Available: https://elib.uni-stuttgart.de/bitstream/11682/5078/1/diss111101.pdf

[40] V. Ivrii, "100 years of Weyl's law," *Bulletin of Mathematical Sciences*, vol. 6, no. 3, pp. 379–452, 2016.

[41] M. A. Shubin, *Invitation to Partial Differential Equations*, ser. Graduate Studies in Mathematics. American Mathematical Society, 2020, vol. 205.

[42] C. Rogers, "Covering a sphere with spheres," *Mathematika*, vol. 10, no. 2, pp. 157–164, 1963.

Appendix

Lemma 7 (Inversion lemma)**.** *For $\kappa_1 \in \mathbb{R}_+^*$, $\kappa_2 \in \mathbb{R}$, and $\beta_1, \beta_2 \in \mathbb{R}_+^*$ such that $\beta_1 > \beta_2$, let $\{\zeta_n\}_{n\in\mathbb{N}^*}$ be a sequence of real numbers satisfying*

$$n = \kappa_1 \zeta_n^{-\beta_1} + \kappa_2\, \zeta_n^{-\beta_2} + o_{n\to\infty}\left(\zeta_n^{-\beta_2}\right). \tag{23}$$

Then, we have

$$\zeta_n = \kappa_1^{1/\beta_1} n^{-1/\beta_1} + \frac{\kappa_1^{1/\beta_1 - \beta_2/\beta_1}\kappa_2}{\beta_1}\, n^{\beta_2/\beta_1 - 1/\beta_1 - 1} + o_{n\to\infty}\left(n^{\beta_2/\beta_1 - 1/\beta_1 - 1}\right).$$

Proof. We first observe that $\lim_{n\to\infty} \zeta_n = 0$. Therefore, upon rewriting

$$\begin{aligned} n &= \kappa_1 \zeta_n^{-\beta_1} + \kappa_2\, \zeta_n^{-\beta_2} + o_{n\to\infty}\left(\zeta_n^{-\beta_2}\right) & (24)\\ &= \kappa_1 \zeta_n^{-\beta_1}\left[1 + \kappa_1^{-1}\kappa_2\, \zeta_n^{\beta_1-\beta_2} + o_{n\to\infty}\left(\zeta_n^{\beta_1-\beta_2}\right)\right], & (25) \end{aligned}$$

and using $\beta_1 > \beta_2$, we obtain

$$\zeta_n = \kappa_1^{1/\beta_1} n^{-1/\beta_1}(1 + o_{n\to\infty}(1)). \tag{26}$$

Using (26) in (24)-(25) allows us to make the term $(1 + o_{n\to\infty}(1))$ more specific resulting in

$$\begin{aligned} \zeta_n &= \kappa_1^{1/\beta_1} n^{-1/\beta_1}\left[1 + \kappa_1^{-1}\kappa_2\, \zeta_n^{\beta_1-\beta_2} + o_{n\to\infty}\left(\zeta_n^{\beta_1-\beta_2}\right)\right]^{1/\beta_1}\\ &= \kappa_1^{1/\beta_1} n^{-1/\beta_1}\left[1 + \frac{\kappa_1^{-1}\kappa_2}{\beta_1}\, \zeta_n^{\beta_1-\beta_2} + o_{n\to\infty}\left(\zeta_n^{\beta_1-\beta_2}\right)\right]\\ &\overset{(26)}{=} \kappa_1^{1/\beta_1} n^{-1/\beta_1}\left[1 + \frac{\kappa_1^{-\beta_2/\beta_1}\kappa_2}{\beta_1}\, n^{\beta_2/\beta_1 - 1} + o_{n\to\infty}\left(n^{\beta_2/\beta_1 - 1}\right)\right]\\ &= \kappa_1^{1/\beta_1} n^{-1/\beta_1} + \frac{\kappa_1^{1/\beta_1 - \beta_2/\beta_1}\kappa_2}{\beta_1}\, n^{\beta_2/\beta_1 - 1/\beta_1 - 1} + o_{n\to\infty}\left(n^{\beta_2/\beta_1 - 1/\beta_1 - 1}\right), \end{aligned}$$

and thereby finalizing the proof. □

Lemma 8. *Let the sequence $\{\mu_n\}_{n\in\mathbb{N}^*}$ be such that*

$$\mu_n = \frac{c_1}{n^{\alpha_1}} + \frac{c_2}{n^{\alpha_2}} + o_{n\to\infty}\left(\frac{1}{n^{\alpha_2}}\right),$$

with $c_1 \in \mathbb{R}_+^*$, $c_2 \in \mathbb{R}$, and $\alpha_1, \alpha_2 \in \mathbb{R}_+^*$ satisfying

$$\textit{either} \quad \alpha_1 < \alpha_2 < \alpha_1\left(1 + \frac{2}{6\alpha_1 + 1}\right), \quad \textit{or} \quad \begin{cases} \alpha_1 = \alpha_2, \ \textit{and} \\ c_2 = 0. \end{cases}$$

The metric entropy with respect to the norm $\|\cdot\|_2$ of the ellipsoid in ℓ^2 with semi-axes $\{\mu_n\}_{n\in\mathbb{N}^*}$ satisfies

$$H(\varepsilon) = \frac{\alpha_1 {c_1}^{1/\alpha_1}}{\ln(2)}\, \varepsilon^{-1/\alpha_1} + \frac{c_2\, c_1^{\frac{1-\alpha_2}{\alpha_1}}}{(\alpha_1 + 1 - \alpha_2)\ln(2)}\, \varepsilon^{\frac{\alpha_2 - \alpha_1 - 1}{\alpha_1}} + o_{\varepsilon\to 0}\left(\varepsilon^{\frac{\alpha_2 - \alpha_1 - 1}{\alpha_1}}\right).$$

Proof. For $\alpha_1 = \alpha_2$ and $c_2 = 0$, the result is by [18, Theorem 11]. In the case $\alpha_1 < \alpha_2 < \alpha_1\left(1 + \frac{2}{6\alpha_1+1}\right)$, the statement is a consequence of [18, Theorem 12]. $\square$

Lemma 9. *Let $d \in \mathbb{N}^*$ and let $\mathcal{E}^d$ be the ellipsoid in $\mathbb{C}^d$ with positive semi-axes $\mu_1, \dots, \mu_d$. The metric entropy of $\mathcal{E}^d$ with respect to the norm $\|\cdot\|_2$ satisfies*

$$H\Big(\varepsilon; \mathcal{E}^d, \|\cdot\|_2\Big) \geq 2d\left[\log\left(\varepsilon^{-1}\right) + \frac{1}{d}\sum_{n=1}^{d}\log(\mu_n)\right], \quad \text{for all } \varepsilon > 0.$$

Proof. The proof follows directly by application of [19, Theorem 4] with $p = q = 2$, $\mathbb{K} = \mathbb{C}$, and where we set $\underline{\kappa} = 1$, which is possible by the remark at the end of the proof of [19, Theorem 4]. $\square$

Lemma 10. *There exists a sequence $\{\kappa(d)\}_{d\in\mathbb{N}^*}$ satisfying*

$$\kappa(d) = 1 + O_{d\to\infty}\left(\frac{\log d}{d}\right),$$

such that

$$N(\varepsilon; \mathcal{B}_2, \|\cdot\|_2)^{1/d}\,\varepsilon \leq \kappa(d), \quad \text{for all } d \in \mathbb{N}^* \text{ and } \varepsilon > 0,$$

where $\mathcal{B}_2$ is the unit ball in $\mathbb{R}^d$ with respect to the $\|\cdot\|_2$-norm.

Proof. This is a direct consequence of [42, Theorem 3]. $\square$