Kolmogorov n-Widths of Function Classes Induced by a Non-Degenerate Differential Operator: A Convex Duality Approach*

Patrick L. Combettes[1] and Dinh Dũng[2]

[1]Sorbonne Universités – UPMC Univ. Paris 06
UMR 7598, Laboratoire Jacques-Louis Lions
F-75005 Paris, France
plc@ljll.math.upmc.fr

[2]Information Technology Institute
Vietnam National University
Hanoi, Vietnam
dinhzung@gmail.com

Dedicated to Lionel Thibault on the occasion of his 65th birthday

Abstract

The problem of computing the asymptotic order of the Kolmogorov n-width of the unit ball of the space of multivariate periodic functions induced by a differential operator associated with a polynomial in the general case when the ball is compactly embedded into L_2 has been open for a long time. In the present paper, we use convex analytical tools to solve it in the case when the differential operator is non-degenerate.

Keywords. asymptotic order · Kolmogorov n-widths · non-degenerate differential operator · convex duality

Mathematics Subject Classifications (2010) 41A10; 41A50; 41A63

*Contact author: P. L. Combettes, plc@ljll.math.upmc.fr, phone: +33 1 4427 6319, fax: +33 1 4427 7200.

1 Introduction

The problem of evaluating Kolmogorov n-widths naturally arises in various applied mathematics problems such as approximation theory, compressed sensing, neural networks, signal processing, statistics, and numerical analysis; see [3, 9, 10, 16, 18, 20, 21, 28, 30, 31]. The aim of the present paper is to study Kolmogorov n-widths of classes of multivariate periodic functions induced by a differential operator. In order to describe the exact setting of the problem let us introduce some notation.

We first recall the notion of Kolmogorov n-widths [14, 20]. Let $\mathscr{X}$ be a normed space, let F be a nonempty subset of $\mathscr{X}$ such that $F = -F$, and let $\mathscr{G}_n$ be the class of all vector subspaces of $\mathscr{X}$ of dimension at most n. The Kolmogorov n-width of F in $\mathscr{X}$ is

$$d_n(F,\mathscr{X}) = \inf_{G\in\mathscr{G}_n} \sup_{f\in F} \inf_{g\in G} \|f-g\|_{\mathscr{X}}. \tag{1.1}$$

This notion quantifies the error of the best approximation to the elements of F by elements in a vector subspace of $\mathscr{X}$ of dimension at most n [20, 27, 28].

In computational mathematics, the so-called ε-dimension $n_\varepsilon(F,\mathscr{X})$ is used to quantify the computational complexity. It is defined by

$$n_\varepsilon(F,\mathscr{X}) = \inf\left\{ n\in\mathbb{N} \;\middle|\; (\exists\, G\in\mathscr{G}_n) \sup_{f\in F} \inf_{g\in G} \|f-g\|_{\mathscr{X}} \leqslant \varepsilon \right\}. \tag{1.2}$$

This approximation characteristic is the inverse of $d_n(F,\mathscr{X})$ in the sense that the quantity $n_\varepsilon(F,\mathscr{X})$ is the smallest integer n_ε such that the approximation of F by a suitably chosen approximant n_ε-dimensional subspace G in $\mathscr{X}$ gives an approximation error less than ε. Recently, there has been strong interest in applications of Kolmogorov n-widths, and its dual Gelfand n-widths, to compressed sensing [3, 10, 11, 21]. Kolmogorov n-widths and ε-dimensions of classes of functions with mixed smoothness have also been employed in recent high-dimensional approximation studies [5, 9].

We consider functions on $\mathbb{R}^d$ which are 2π-periodic in each variable as functions defined on $\mathbb{T}^d = [-\pi,\pi]^d$. Denote by $L_2(\mathbb{T}^d)$ the Hilbert space of square-integrable functions on $\mathbb{T}^d$ equipped with the standard scalar product, i.e.,

$$(\forall f\in L_2(\mathbb{T}^d))(\forall g\in L_2(\mathbb{T}^d)) \quad \langle f \mid g\rangle = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} f(x)\overline{g(x)}dx, \tag{1.3}$$

and by $\mathscr{S}'(\mathbb{T}^d)$ the space of distributions on $\mathbb{T}^d$. The norm of $f\in L_2(\mathbb{T}^d)$ is $\|f\|_2 = \sqrt{\langle f\mid f\rangle}$ and, given $k\in\mathbb{Z}^d$, the kth Fourier coefficient of $f\in L_2(\mathbb{T}^d)$ is $\hat{f}(k) = \left\langle f \mid e^{i\langle k|\cdot\rangle}\right\rangle$. Every $f\in\mathscr{S}'(\mathbb{T}^d)$ can be identified with the formal Fourier series

$$f = \sum_{k\in\mathbb{Z}^d} \hat{f}(k) e^{i\langle k|\cdot\rangle}, \tag{1.4}$$

where the sequence $(\hat{f}(k))_{k\in\mathbb{Z}^d}$ is a tempered sequence [24, 28]. By Parseval's identity, $L_2(\mathbb{T}^d)$ is the subset of $\mathscr{S}'(\mathbb{T}^d)$ of all distributions f for which

$$\sum_{k\in\mathbb{Z}^d} |\hat{f}(k)|^2 < +\infty. \tag{1.5}$$

Let $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ and let $f \in \mathscr{S}'(\mathbb{T}^d)$. We set

$$\mathbb{Z}_0^d(\alpha) = \big\{(k_1, \dots, k_d) \in \mathbb{Z}^d \;\big|\; (\forall j \in \{1, \dots, d\})\;\; \alpha_j \neq 0 \;\Rightarrow\; k_j \neq 0\big\}. \tag{1.6}$$

As usual, we set $|\alpha| = \sum_{j=1}^d \alpha_j$ and, given $z = (z_1, \dots, z_d) \in \mathbb{C}^d$, we set $z^\alpha = \prod_{j=1}^d z_j^{\alpha_j}$. The αth derivative of $f \in \mathscr{S}'(\mathbb{T}^d)$ is the distribution $f^{(\alpha)} \in \mathscr{S}'(\mathbb{T}^d)$ given through the identification

$$f^{(\alpha)} = \sum_{k \in \mathbb{Z}_0^d(\alpha)} (ik)^\alpha \hat{f}(k) e^{i\langle k | \cdot \rangle}. \tag{1.7}$$

The differential operator D^α on $\mathscr{S}'(\mathbb{T}^d)$ is defined by $D^\alpha \colon f \mapsto (-i)^{|\alpha|} f^{(\alpha)}$. Now let $A \subset \mathbb{N}^d$ be a nonempty finite set, let $(c_\alpha)_{\alpha \in A}$ be nonzero real numbers, and define a polynomial by

$$P \colon x \mapsto \sum_{\alpha \in A} c_\alpha x^\alpha. \tag{1.8}$$

The differential operator $P(D)$ on $\mathscr{S}'(\mathbb{T}^d)$ induced by P is

$$P(D) = \sum_{\alpha \in A} c_\alpha D^\alpha. \tag{1.9}$$

Set

$$W_2^{[P]} = \big\{f \in \mathscr{S}'(\mathbb{T}^d) \;\big|\; P(D)(f) \in L_2(\mathbb{T}^d)\big\}, \tag{1.10}$$

denote the seminorm of $f \in W_2^{[P]}$ by

$$\|f\|_{W_2^{[P]}} = \|P(D)(f)\|_2, \tag{1.11}$$

and let

$$U_2^{[P]} = \Big\{f \in W_2^{[P]} \;\Big|\; \|f\|_{W_2^{[P]}} \leqslant 1\Big\}. \tag{1.12}$$

The problem of computing asymptotic orders of $d_n(U_2^{[P]}, L_2(\mathbb{T}^d))$ in the general case when $W_2^{[P]}$ is compactly embedded into $L_2(\mathbb{T}^d)$ has been open for a long time; see, e.g., [26, Chapter III] for details. Our main contribution is to solve it for a non-degenerate differential operator $P(D)$ (see Definition 2.4). Using convex-analytical tool, we establish the asymptotic order

$$d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \asymp n^{-\varrho} (\log n)^{\nu\varrho}, \tag{1.13}$$

where ϱ and ν depend only on P. In the present paper, we restrict our attention to multivariate periodic functions. One can consider an extension of $d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big)$ to $d_n\big(U_2^{[P]}, L_2(\Omega)\big)$, where Ω is a bounded domain in $\mathbb{R}^d$ (if Ω is unbounded, then $U_2^{[P]}$ is not a compact set and, therefore, $d_n\big(U_2^{[P]}, L_2(\Omega)\big) = +\infty$). The assumption that the differential operator $P(D)$ is non-degenerate plays a crucial role in the proof technique of (1.13), where convex analytical tools are employed. Intuitively, the problem of estimating $d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big)$ may be related to that of estimating $d_n(U_2^A, L_2(\mathbb{T}^d))$ studied in [6], where

U_2^A is the closed unit ball of the space W_2^A of functions with several bounded mixed derivatives (see Subsection 4.4 for a precise definition).

The first exact values of n-widths of univariate Sobolev classes were obtained by Kolmogorov [14] (see also [15, pp. 186–189]). The problem of computing the asymptotic order of $d_n(U_2^{[P]}, L_2(\mathbb{T}^d))$ is directly related to hyperbolic crosses trigonometric approximations and to n-widths of classes multivariate periodic functions with a bounded mixed smoothness. This line of work was initiated by Babenko in [1, 2]. In particular, the asymptotic orders of n-widths in $L_2(\mathbb{T}^d)$ of these classes were established in [1]. Further work on asymptotic orders and hyperbolic cross approximation can be found in [7, 8, 26] and recent developments in [17, 23, 25, 29]. In [6], the strong asymptotic order of $d_n(U_2^A, L_2(\mathbb{T}^d))$ was computed.

The remainder of the paper is organized as follows. In Section 2, we provide as auxiliary results Jackson-type and Bernstein-type inequalities for trigonometric approximations of functions from $W_2^{[P]}$. We also characterize the compactness of $U_2^{[P]}$ in $L_2(\mathbb{T}^d)$ and the non-degenerateness of $P(D)$. In Section 3, we present the main result of the paper, namely the asymptotic order of $d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big)$ in the case when $P(D)$ is non-degenerate. In Section 4, we derive norm equivalences relative to $\|\cdot\|_{W_2^{[P]}}$ and, based on them, we provide examples of n-widths $d_n(U_2^{[P]}, L_2(\mathbb{T}^d))$ for non-degenerate differential operators.

2 Preliminaries

2.1 Notation, standing assumption, and definitions

We set $\mathbb{N} = \{0, 1, \dots, \}$, $\mathbb{N}^* = \{1, 2, \dots, \}$, $\mathbb{R}_+ = [0, +\infty[$, and $\mathbb{R}_{++} =]0, +\infty[$. Let Θ be an abstract set, and let Φ and Ψ be functions from Θ to $\mathbb{R}$. Then we write

$$(\forall \theta \in \Theta) \quad \Phi(\theta) \asymp \Psi(\theta) \tag{2.1}$$

if there exist $\gamma_1 \in \mathbb{R}_{++}$ and $\gamma_2 \in \mathbb{R}_{++}$ such that $(\forall \theta \in \Theta)$ $\gamma_1 \Phi(\theta) \leqslant \Psi(\theta) \leqslant \gamma_2 \Phi(\theta)$. For every $j \in \{1, \dots, d\}$, u^j denotes the j standard unit vector of $\mathbb{R}^d$ and

$$\mathscr{R}^j = \big\{ \lambda u^j \;\big|\; \lambda \in \mathbb{R}_{++} \big\} \tag{2.2}$$

the jth standard strict ray.

Definition 2.1 Let B be a nonempty finite subset of $\mathbb{N}^d$. The convex hull conv(B) of B is the polyhedron spanned by B,

$$\Delta(B) = \Big\{ \alpha \in B \;\Big|\; \big\{ \lambda\alpha \;\big|\; \lambda \in [1, +\infty[\big\} \cap \operatorname{conv}(B) = \{\alpha\} \Big\}, \tag{2.3}$$

and $\vartheta(B)$ is the set of vertices of conv$(\Delta(B))$. In addition,

$$(\forall t \in \mathbb{R}_+) \quad \Omega_B(t) = \Big\{ k \in \mathbb{N}^d \;\Big|\; \max_{\alpha \in B} k^\alpha \leqslant t \Big\}. \tag{2.4}$$

Throughout the paper, the convention 0^0 is adopted and the following standing assumption is made.

Assumption 2.2 *A is a nonempty finite subset of $\mathbb{N}^d$ and $(c_\alpha)_{\alpha\in A}$ are nonzero real numbers. We set*

$$P\colon x\mapsto \sum_{\alpha\in A} c_\alpha x^\alpha \quad\text{and}\quad \tau=\inf_{k\in\mathbb{Z}^d}|P(k)|. \tag{2.5}$$

Moreover, for every $t\in\mathbb{R}_+$, we set

$$K(t)=\big\{k\in\mathbb{Z}^d \;\big|\; |P(k)|\leqslant t\big\} \quad\text{and}\quad V(t)=\bigg\{f\in\mathscr{S}'(\mathbb{T}^d)\;\bigg|\; f=\sum_{k\in K(t)}\hat{f}(k)e^{i\langle k|\cdot\rangle}\bigg\}. \tag{2.6}$$

Remark 2.3 If $0\in A$, then $0\in\vartheta(A)$ and $\Delta(\mathrm{conv}(A))=\Delta(A)$, so that $\vartheta(\mathrm{conv}(A))=\vartheta(A)$. Now suppose that $t\in\,]\tau,+\infty[$. Then $K(t)\neq\varnothing$ and $\dim V(t)=\mathrm{card}\,K(t)$, where $\mathrm{card}\,K(t)$ denotes the cardinality of $K(t)$. In addition, if $\mathrm{card}\,K(t)<+\infty$, then $V(t)$ is the space of trigonometric polynomials with frequencies in $K(t)$.

Definition 2.4 The *Newton diagram* of P is $\Delta(A)$ and the *Newton polyhedron* of P is $\mathrm{conv}(A)$. The intersection of $\mathrm{conv}(A)$ with a supporting hyperplane of $\mathrm{conv}(A)$ is a *face* of $\mathrm{conv}(A)$; $\Sigma(A)$ is the set of intersections of A with a face of $\mathrm{conv}(A)$. The differential operator $P(D)$ is *non-degenerate* if P and, for every $\sigma\in\Sigma(A)$, $P_\sigma\colon\mathbb{R}^d\to\mathbb{R}\colon x\mapsto\sum_{\alpha\in\sigma}c_\alpha x^\alpha$ do not vanish outside the coordinate planes of $\mathbb{R}^d$, i.e.,

$$\big(\forall x\in\mathbb{R}^d\big)\quad\Bigg(\prod_{j=1}^d x_j\neq 0 \quad\Rightarrow\quad \big(\forall\sigma\in\Sigma(A)\big)\quad P(x)P_\sigma(x)\neq 0\Bigg). \tag{2.7}$$

Remark 2.5 Suppose that P is non-degenerate and let $\alpha\in\vartheta(A)$. Then it follows from (2.7) that all the components of α are even.

2.2 Trigonometric approximations

We first prove a Jackson-type inequality.

Lemma 2.6 *Let $t\in\mathbb{R}_{++}$ and define a linear operator $S_t\colon\mathscr{S}'(\mathbb{T}^d)\to\mathscr{S}'(\mathbb{T}^d)$ by*

$$\big(\forall f\in\mathscr{S}'(\mathbb{T}^d)\big)\quad S_t(f)=\sum_{k\in K(t)}\hat{f}(k)e^{i\langle k|\cdot\rangle}. \tag{2.8}$$

Let $f\in W_2^{[P]}$ and suppose that $t>\tau$. Then the distribution $f-S_t(f)$ represents a function in $L_2(\mathbb{T}^d)$ and

$$\|f-S_t(f)\|_2\leqslant t^{-1}\|f\|_{W_2^{[P]}}. \tag{2.9}$$

Proof. Set $g=f-S_t(f)$. Then $g\in\mathscr{S}'(\mathbb{T}^d)$. On the other hand, Parseval's identity yields

$$\|f\|^2_{W_2^{[P]}}=\sum_{k\in\mathbb{Z}^d}|P(k)|^2|\hat{f}(k)|^2. \tag{2.10}$$

Hence,

$$\sum_{k\in\mathbb{Z}^d} |\hat{g}(k)|^2 = \sum_{k\in\mathbb{Z}^d\setminus K(t)} |\hat{f}(k)|^2$$
$$\leqslant \Big(\sup_{k\in\mathbb{Z}^d\setminus K(t)} |P(k)|^{-2}\Big) \sum_{k\in\mathbb{Z}^d\setminus K(t)} |P(k)|^2|\hat{f}(k)|^2$$
$$\leqslant t^{-2}\|f\|^2_{W_2^{[P]}}, \tag{2.11}$$

which means that $f - S_t(f)$ represents a function in $L_2(\mathbb{T}^d)$ for which (2.9) holds. □

Corollary 2.7 *Let $t \in]\tau,+\infty[$. Then*

$$\sup_{f\in U_2^{[P]}} \inf_{\substack{g\in V(t)\\ f-g\in L_2(\mathbb{T}^d)}} \|f-g\|_2 \leqslant t^{-1}. \tag{2.12}$$

Next, we prove a Bernstein-type inequality.

Lemma 2.8 *Let $t \in]\tau,+\infty[$ and let $f \in V(t)\cap L_2(\mathbb{T}^d)$. Then*

$$\|f\|_{W_2^{[P]}} \leqslant t\|f\|_2. \tag{2.13}$$

Proof. By (2.10), we have

$$\|f\|^2_{W_2^{[P]}} = \sum_{k\in K(t)} |P(k)|^2|\hat{f}(k)|^2 \leqslant \Big(\sup_{k\in K(t)} |P(k)|^2\Big) \sum_{k\in K(t)} |\hat{f}(k)|^2 \leqslant t^2\|f\|_2^2, \tag{2.14}$$

which establishes (2.13). □

2.3 Compactness and non-degenerateness

We start with a characterization of the compactness of the unit ball defined in (1.12).

Lemma 2.9 *The set $U_2^{[P]}$ is a compact subset of $L_2(\mathbb{T}^d)$ if and only if the following hold:*

(i) *For every $t \in]\tau,+\infty[$, $K(t)$ is finite.*

(ii) $\tau > 0$.

Proof. To prove sufficiency, suppose that (i) and (ii) hold, and fix $t \in]\tau,+\infty[$. By (i), $V(t)$ is a set of trigonometric polynomials and, consequently, a subset of $L_2(\mathbb{T}^d)$. In particular, using the notation (2.8), $(\forall f \in \mathscr{S}'(\mathbb{T}^d))\ S_t(f) \in L_2(\mathbb{T}^d)$. Hence, by Lemma 2.6,

$$\Big(\forall f \in W_2^{[P]}\Big) \quad f = (f - S_t(f)) + S_t(f) \in L_2(\mathbb{T}^d). \tag{2.15}$$

Thus, $W_2^{[P]} \subset L_2(\mathbb{T}^d)$. On the other hand, (2.10) implies that $U_2^{[P]}$ is a closed subset of $L_2(\mathbb{T}^d)$. Therefore, $U_2^{[P]}$ is compact in $L_2(\mathbb{T}^d)$ if, for every $\varepsilon \in \mathbb{R}_{++}$, it has a finite ε-net in $L_2(\mathbb{T}^d)$ or, equivalently, if the following following two conditions are satisfied:

(iii) For every $\varepsilon \in \mathbb{R}_{++}$, there exists a finite-dimensional vector subspace G_ε of $L_2(\mathbb{T}^d)$ such that

$$\sup_{f \in U_2^{[P]}} \inf_{g \in G_\varepsilon} \|f - g\|_2 \leqslant \varepsilon. \tag{2.16}$$

(iv) $U_2^{[P]}$ is bounded in $L_2(\mathbb{T}^d)$.

It follows from (2.10) that (ii)⇔(iv). On the other hand, since $\dim V(t) = \operatorname{card} K(t)$, Corollary 2.7 yields (i)⇒(iii). To prove necessity, suppose that (i) does not hold. Then $\dim V(\tilde{t}) = \operatorname{card} K(\tilde{t}) = +\infty$ for some $\tilde{t} \in \mathbb{R}_{++}$. By Lemma 2.8, $\widetilde{U} = \big\{ f \in V(\tilde{t}) \cap L_2(\mathbb{T}^d) \;\big|\; \|f\|_2 \leqslant 1/\tilde{t} \big\}$ is a subset of $U_2^{[P]}$ which is not compact in $L_2(\mathbb{T}^d)$. If (ii) does not hold, then $U_2^{[P]} \cap L_2(\mathbb{T}^d)$ is unbounded and, consequently, not compact in $L_2(\mathbb{T}^d)$. ☐

The following lemma characterizes the non-degenerateness of $P(D)$.

Lemma 2.10 *$P(D)$ is non-degenerate if and only if*

$$(\exists\, \gamma \in \mathbb{R}_{++})(\forall x \in \mathbb{R}^d) \quad |P(x)| \geqslant \gamma \max_{\alpha \in \vartheta(A)} |x^\alpha|. \tag{2.17}$$

Proof. As proved in [12, 19], $P(D)$ is non-degenerate if and only if

$$(\exists\, \gamma_1 \in \mathbb{R}_{++})(\forall x \in \mathbb{R}^d) \quad |P(x)| \geqslant \gamma_1 \sum_{\alpha \in \vartheta(A)} |x^\alpha|. \tag{2.18}$$

Hence, since there exist $\gamma_2 \in \mathbb{R}_{++}$ and $\gamma_3 \in \mathbb{R}_{++}$ such that

$$(\forall x \in \mathbb{R}^d) \quad \gamma_2 \max_{\alpha \in \vartheta(A)} |x^\alpha| \leqslant \sum_{\alpha \in \vartheta(A)} |x^\alpha| \leqslant \gamma_3 \max_{\alpha \in \vartheta(A)} |x^\alpha|, \tag{2.19}$$

the proof is complete. ☐

Lemma 2.11 *Let B be a nonempty finite subset of $\mathbb{N}^d$ and let $t \in \mathbb{R}_+$. Then*

$$\Omega_B(t) = \left\{ k \in \mathbb{N}^d \;\middle|\; \max_{\alpha \in B} k^\alpha \leqslant t \right\} \tag{2.20}$$

is finite if and only if

$$(\forall j \in \{1, \ldots, d\}) \quad B \cap \mathscr{R}^j \neq \varnothing. \tag{2.21}$$

Proof. If (2.21) holds, then $(\forall j \in \{1, \ldots, d\})(\exists\, a_j \in \mathbb{R}_{++})\ a_j u^j \in B \cap \mathscr{R}^j$. Hence, (2.4) implies that $\Omega_B(t) \subset \bigcap_{j=1}^d \big\{ k \in \mathbb{N}^d \;\big|\; k_j \leqslant t^{1/a_j} \big\}$ and, therefore, $\Omega_B(t)$ is bounded. Conversely, if (2.21) does not hold, then there exists $j \in \{1, \ldots, d\}$ such that $\big\{ m u^j \;\big|\; m \in \mathbb{N} \big\} \subset \Omega_B(t)$, which shows that $\Omega_B(t)$ is unbounded. ☐

Theorem 2.12 *Suppose that $P(D)$ is non-degenerate. Then $U_2^{[P]}$ is a compact subset of $L_2(\mathbb{T}^d)$ if and only if (2.21) is satisfied and $0 \in A$.*

Proof. Let us prove that there exists $\gamma_1 \in \mathbb{R}_{++}$ such that

$$\left(\forall k \in \mathbb{Z}^d\right) \quad |P(k)| \leqslant \gamma_1 \max_{\alpha \in \vartheta(A)} |k^\alpha|. \tag{2.22}$$

Since there exists $\gamma_1 \in \mathbb{R}_{++}$ such that

$$\left(\forall k \in \mathbb{Z}^d\right) \quad |P(k)| \leqslant \gamma_1 \max_{\alpha \in A} |k^\alpha|, \tag{2.23}$$

and since (2.22) trivially holds if there exists $j \in \{1, \ldots, d\}$ such that $k_j = 0$, it is enough to show that

$$\left(\forall \alpha \in A\right)\left(\forall k \in \mathbb{N}^{*d}\right) \quad k^\alpha \leqslant \max_{\beta \in \vartheta(A)} k^\beta, \tag{2.24}$$

and *a fortiori* that

$$\left(\forall \alpha \in A\right)\left(\forall x \in \mathbb{R}^d_+\right) \quad \langle \alpha \mid x \rangle \leqslant \max_{\beta \in \vartheta(A)} \langle \beta \mid x \rangle. \tag{2.25}$$

Indeed, since $\alpha \in \operatorname{conv}(\vartheta(A))$, by Carathéodory's theorem [22, Theorem 17.1], α is a convex combination of points $(\beta^j)_{1 \leqslant j \leqslant d+1}$ in $\vartheta(B)$, say

$$\alpha = \sum_{j=1}^{d+1} \lambda_j \beta^j, \quad \text{where} \quad (\lambda_j)_{1 \leqslant j \leqslant d+1} \in \mathbb{R}^{d+1}_+ \quad \text{and} \quad \sum_{j=1}^{d+1} \lambda_j = 1. \tag{2.26}$$

Therefore

$$\left(\forall x \in \mathbb{R}^d_+\right) \quad \langle \alpha \mid x \rangle = \sum_{j=1}^{d+1} \lambda_j \left\langle \beta_j \mid x \right\rangle \leqslant \sum_{j=1}^{d+1} \lambda_j \max_{\beta \in \vartheta(A)} \langle \beta \mid x \rangle = \max_{\beta \in \vartheta(A)} \langle \beta \mid x \rangle. \tag{2.27}$$

Hence, Lemma 2.10 asserts that there exists $\gamma_2 \in \mathbb{R}_{++}$ such that

$$\left(\forall k \in \mathbb{Z}^d\right) \quad \gamma_2 \max_{\alpha \in \vartheta(A)} |k^\alpha| \leqslant |P(k)| \leqslant \gamma_1 \max_{\alpha \in \vartheta(A)} |k^\alpha|. \tag{2.28}$$

Consequently, by Lemma 2.9, $U_2^{[P]}$ is a compact set in $L_2(\mathbb{T}^d)$ if and only if, for every $t \in \mathbb{R}_+$, $\Omega_A(t)$ is finite and

$$\inf_{k \in \mathbb{N}^d} \max_{\alpha \in A} k^\alpha > 0. \tag{2.29}$$

In view of Lemma 2.11, the first condition is equivalent to (2.21) and the second to $0 \in A$. ☐

3 Main result

3.1 Convex-analytical results

Several important convex-analytical facts underly our analysis (see [4, 22] for background on convex analysis). We start with the following corollary.

Corollary 3.1 *Suppose that $P(D)$ is non-degenerate. Then* $\big(\forall k\in\mathbb{Z}^d\big)\ |P(k)|\asymp\max_{\alpha\in\vartheta(A)}|k^\alpha|$.

Proof. Combine (2.28) and Lemma 2.10. □

Next, we investigate the geometry of our problem from the view-point of convex duality. Let C be a subset of $\mathbb{R}^d$. Recall that the *polar set* of C is

$$C^{\odot}=\big\{x\in\mathbb{R}^d\;\big|\;(\forall\alpha\in C)\ \langle\alpha\mid x\rangle\leqslant 1\big\},\tag{3.1}$$

and the *indicator function* of C is

$$\iota_C\colon\mathbb{R}^d\to\left]-\infty,+\infty\right]\colon x\mapsto\begin{cases}0, & \text{if } x\in C;\\ +\infty, & \text{otherwise.}\end{cases}\tag{3.2}$$

Moreover, if C is convex and $0\in C$, the *Minkowski gauge* of C is the lower semicontinuous convex function

$$m_C\colon\mathbb{R}^d\to\left]-\infty,+\infty\right]\colon x\mapsto\inf\big\{\xi\in\mathbb{R}_{++}\;\big|\;x\in\xi C\big\}.\tag{3.3}$$

Finally, the domain of a function $\varphi\colon\mathbb{R}^d\to\left]-\infty,+\infty\right]$ is $\operatorname{dom}\varphi=\big\{x\in\mathbb{R}^d\;\big|\;\varphi(x)<+\infty\big\}$.

Lemma 3.2 *Let B be a nonempty finite subset of $\mathbb{R}^d_+$ such that*

$$0\in B\quad\textit{and}\quad(\forall j\in\{1,\ldots,d\})\quad B\cap\mathscr{R}^j\neq\varnothing.\tag{3.4}$$

Set $\mathbf{1}=(1,\ldots,1)\in\mathbb{R}^d$, let $\mu(B)$ be the optimal value of the problem

$$\underset{x\in B^{\odot}}{\text{maximize}}\;\sum_{j=1}^{d}x_j,\tag{3.5}$$

and set

$$\varrho(B)=\max\big\{\rho\in\mathbb{R}_{++}\;\big|\;\rho\mathbf{1}\in\operatorname{conv}(B)\big\}.\tag{3.6}$$

Then $\varrho(B)\in\mathbb{R}_{++}$ and $\mu(B)=1/\varrho(B)$.

Proof. It follows from (3.4) that

$$\mathbb{R}^d_+\cap B^{\odot}=\mathbb{R}^d_+\cap\bigcap_{\alpha\in B}\big\{x\in\mathbb{R}^d\;\big|\;\langle x\mid\alpha\rangle\leqslant 1\big\}\tag{3.7}$$

is a nonempty compact set and hence (3.5) does have a solution. Now fix $j \in \{1,\dots,d\}$. Then $(\exists\, a_j \in \mathbb{R}_{++})$ $a_j u^j \in B$. Hence $x^j = (1/a_j)u^j \in B^{\odot}$ and therefore $\mu(B) = \max_{x\in B^{\odot}} \langle x \mid \mathbf{1}\rangle \geqslant \langle x^j \mid \mathbf{1}\rangle = 1/a_j > 0$. Altogether $\mu(B) \in \mathbb{R}_{++}$. Likewise, (3.4) implies that $\varrho(B) \in \mathbb{R}_{++}$. Let us set $\varphi = m_{\operatorname{conv}(B)}$ and $\psi = \iota_{\{\mathbf{1}\}}$. Then it follows from (3.4) that $\operatorname{dom}\varphi = \operatorname{dom} m_{\operatorname{conv}(B)} = \mathbb{R}^d_+$. Furthermore, the conjugate of φ is $\varphi^* = \iota_{(\operatorname{conv}(B))^{\odot}} = \iota_{B^{\odot}}$ [4, Propositions 14.12 and 7.14(vi)] and the conjugate of ψ is $\psi^* = \langle\cdot \mid \mathbf{1}\rangle$. Hence, since $\mathbf{1} \in \operatorname{int}\operatorname{dom}\varphi = \mathbb{R}^d_{++}$, $\operatorname{dom}\psi \cap \operatorname{int}\operatorname{dom}\varphi \neq \varnothing$ and the Fenchel duality formula [4, Proposition 15.13] yields

$$
\begin{aligned}
\mu(B) &= \max_{x\in B^{\odot}} \sum_{j=1}^{d} x_j \\
&= -\min_{x\in B^{\odot}} \langle -x \mid \mathbf{1}\rangle \\
&= -\min_{x\in\mathbb{R}^d} \big(\iota_{B^{\odot}}(x) + \langle -x \mid \mathbf{1}\rangle\big) \\
&= -\min_{x\in\mathbb{R}^d} \big(\varphi^*(x) + \psi^*(-x)\big) \\
&= \inf_{\alpha\in\mathbb{R}^d} \big(\varphi(\alpha) + \psi(\alpha)\big) \\
&= \inf_{\alpha\in\mathbb{R}^d} \big(m_{\operatorname{conv}(B)}(\alpha) + \iota_{\{\mathbf{1}\}}(\alpha)\big) \\
&= m_{\operatorname{conv}(B)}(\mathbf{1}) \\
&= \inf\big\{\xi \in \mathbb{R}_{++} \;\big|\; \mathbf{1} \in \xi\operatorname{conv}(B)\big\} \\
&= \frac{1}{\sup\big\{\rho \in \mathbb{R}_{++} \;\big|\; \rho\mathbf{1} \in \operatorname{conv}(B)\big\}}.
\end{aligned}
\tag{3.8}
$$

We conclude that $\mu(B) = 1/\varrho(B)$. ☐

To illustrate the duality principles underlying Lemma 3.2, we consider two examples.

Example 3.3 We consider the case when $d = 2$ and $B = \{(6,0),(0,6),(4,4),(0,0)\}$ (see Figure 1). Then (3.4) is satisfied, $\mu(B) = 1/4$, and $\varrho(B) = 4$. The set of solutions to (3.5) is the set S represented by the solid red segment: $S = \big\{(x_1,x_2) \in [1/12,1/6]^2 \;\big|\; x_1 + x_2 = 1/4\big\}$.

Example 3.4 In this example we consider the case when $B = \{(0,6),(2,4),(4,0),(0,0)\}$. Then (3.4) is satisfied, $\mu(B) = 3/8$, and $\varrho(B) = 8/3$. The set of solutions to (3.5) reduces to the singleton $S = \{(1/4,1/8)\}$.

Lemma 3.5 *Let B be a nonempty finite subset of $\mathbb{R}^d_+$ and suppose that*

$$
(\forall j \in \{1,\dots,d\}) \quad B \cap \mathscr{R}^j \neq \varnothing. \tag{3.9}
$$

Let $\mu(B)$ be the optimal value of the problem

$$
\underset{x\in B^{\odot}}{\text{maximize}} \; \sum_{j=1}^{d} x_j, \tag{3.10}
$$

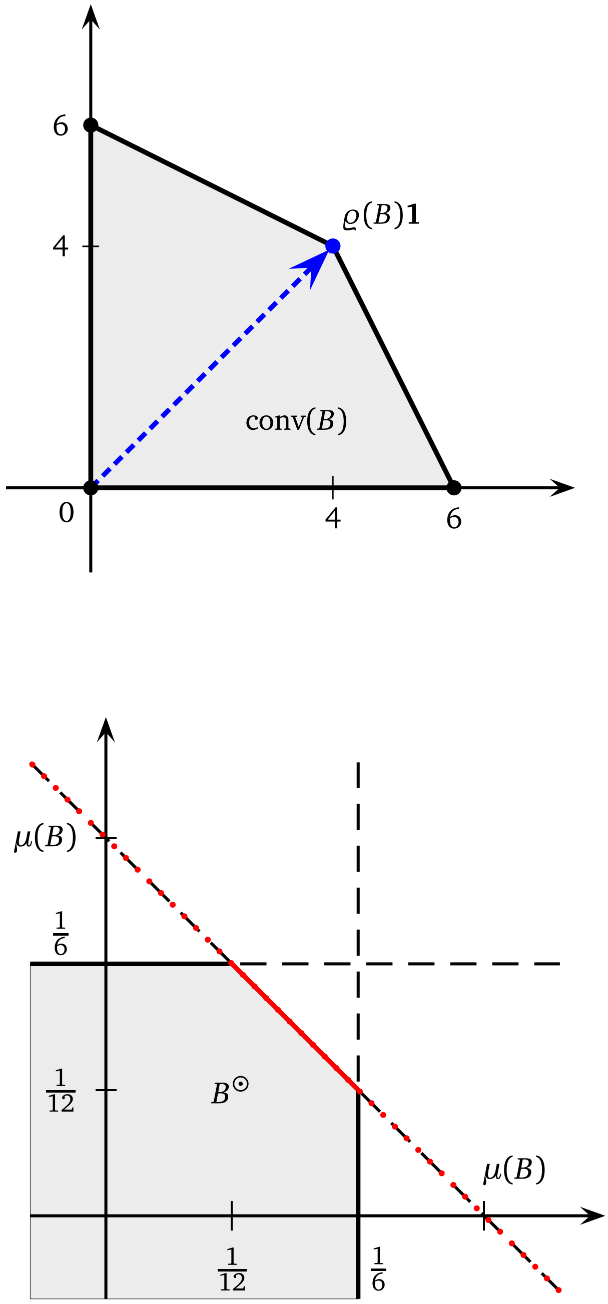

Figure 1: Graphical illustration of Example 3.3: In gray, the Newton polyhedron (top) and its polar (bottom). The dashed lines are the hyperplanes delimiting the polar set $B^{\odot}$ and the dotted line represents the optimal level curve of the objective function $x \mapsto \langle x \mid \mathbf{1} \rangle$ in (3.5). The solid red segment depicts the solution set of (3.5).

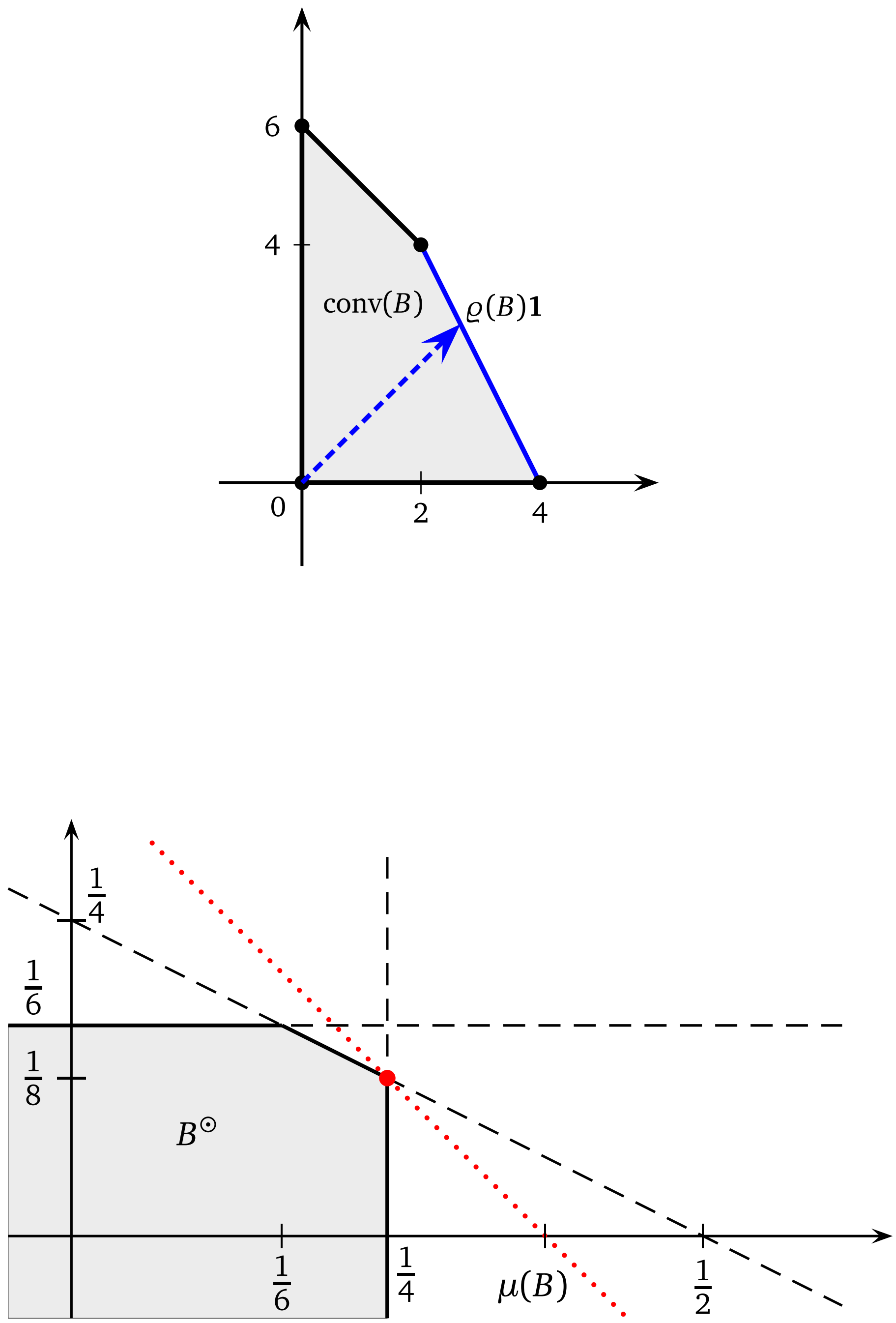

Figure 2: Graphical illustration of Example 3.4: In gray, the Newton polyhedron (top) and its polar (bottom). The dashed lines are the hyperplanes delimiting the polar set $B^{\odot}$ and the dotted line represents the optimal level curve of the objective function $x \mapsto \langle x \mid \mathbf{1} \rangle$ in (3.5). The red dot locates the unique solution to (3.5).

and let $\nu(B)$ *be the dimension of its set of solutions. Then* $\mu(B) \in \mathbb{R}_{++}$ *and*

$$(\forall t \in [2, +\infty[) \quad \operatorname{card} \Omega_B(t) \asymp t^{\mu(B)} \big(\log t\big)^{\nu(B)}. \tag{3.11}$$

Proof. The fact that $\mu(B) \in \mathbb{R}_{++}$ was proved as in Lemma 3.2. Now fix $t \in [2, +\infty[$ and set $\Lambda_B(t) = \big\{x \in \mathbb{R}^d_+ \;\big|\; \max_{\alpha \in B} x^\alpha \leqslant t\big\}$. Then, as in the proof of Lemma 2.11, one can see that $\Lambda_B(t)$ is a bounded subset of $\mathbb{R}^d_+$. If we denote by $\operatorname{vol} \Lambda_B(t)$ the volume of $\Lambda_B(t)$, then it follows from [6, Theorem 1] that

$$\operatorname{vol} \Lambda_B(t) \asymp t^{\mu(B)} (\log t)^{\nu(B)}. \tag{3.12}$$

Furthermore, proceeding as in the proof of [6, Theorem 2], one shows that

$$\operatorname{card} \Omega_B(t) \asymp \operatorname{vol} \Lambda_B(t). \tag{3.13}$$

These asymptotic relations prove the claim. □

3.2 Main result: asymptotic order of Kolmogorov n-width

Our main result can now be stated and proved.

Theorem 3.6 *Suppose that* $P(D)$ *is non-degenerate and that*

$$0 \in A \quad \textit{and} \quad (\forall j \in \{1, \dots, d\}) \quad A \cap \mathscr{R}^j \neq \varnothing. \tag{3.14}$$

Let μ *be the optimal value of the problem*

$$\underset{x \in \vartheta(A)^{\ominus}}{\text{maximize}} \; \sum_{j=1}^{d} x_j, \tag{3.15}$$

let ν *be the dimension of its set of solutions, and set*

$$\varrho = \max\big\{\rho \in \mathbb{R}_{++} \;\big|\; \rho \mathbf{1} \in \operatorname{conv}(\vartheta(A))\big\}. \tag{3.16}$$

Then $\mu = 1/\varrho \in \mathbb{R}_{++}$ *and, for n sufficiently large,*

$$d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \asymp n^{-\varrho} \big(\log n\big)^{\nu\varrho}. \tag{3.17}$$

Equivalently, using (1.2)*, for* $\varepsilon \in \mathbb{R}_{++}$ *sufficiently small,*

$$n_\varepsilon\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \asymp \varepsilon^{-1/\varrho} |\log \varepsilon|^{\nu}. \tag{3.18}$$

Proof. Since A satisfies (3.14), so does $\vartheta(A)$. Hence the fact that $\mu = 1/\varrho \in \mathbb{R}_{++}$ follows from Lemma 3.2. We also note that the equivalence between (3.17) and (3.18) follows from (1.1) and (1.2). To show (3.17), set $\bar{t} = \max\{2, \tau\}$. Then we derive from Corollary 3.1 that

$$(\forall t \in [\bar{t}, +\infty[) \quad \operatorname{card} \Omega_{\vartheta(A)}(t) \asymp \operatorname{card} K(t). \tag{3.19}$$

Applying Lemma 3.5 to $\vartheta(A)$ yields

$$(\forall t \in [\bar{t}, +\infty[) \quad \dim V(t) = \operatorname{card} K(t) \asymp t^{1/\varrho}\big(\log t\big)^{\nu}. \tag{3.20}$$

Hence, for every $n \in \mathbb{N}$ large enough, there exists $t \in \mathbb{R}_{++}$ depending on n such that

$$\begin{aligned} \gamma_1 \dim V(t) \leqslant \gamma_3 t^{1/\varrho}\big(\log t\big)^{\nu} \leqslant n < \gamma_3 (t+1)^{1/\varrho}\big(\log(t+1)\big)^{\nu} \\ \leqslant \gamma_2 \dim V(t+1) \leqslant \gamma_4 t^{1/\varrho}\big(\log t\big)^{\nu}, \end{aligned} \tag{3.21}$$

where γ_1, γ_2, γ_3, and γ_4 are strictly positive real parameters that are independent from n and t. Therefore,

$$n \asymp t^{1/\varrho}\big(\log t\big)^{\nu}. \tag{3.22}$$

or, equivalently,

$$t^{-1} \asymp n^{-\varrho}\big(\log n\big)^{\nu\varrho}. \tag{3.23}$$

It therefore follows from (1.1) and Corollary 2.7 that

$$d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \leqslant t^{-1} \asymp n^{-\varrho}\big(\log n\big)^{\nu\varrho}, \tag{3.24}$$

which establishes the upper bound in (3.17). To establish the lower bound, let us recall from [27] that, for every $n+1$-dimensional vector subspace G_{n+1} of $L_2(\mathbb{T}^d)$ and every $\eta \in \mathbb{R}_{++}$, we have

$$d_n\big(B_{n+1}(\eta), L_2(\mathbb{T}^d)\big) = \eta, \quad \text{where} \quad B_{n+1}(\eta) = \big\{f \in G_{n+1} \;\big|\; \|f\|_{L_2(\mathbb{T}^d)} \leqslant \eta\big\}. \tag{3.25}$$

Arguing as in (3.20)–(3.23), for $n \in \mathbb{N}$ sufficiently large, there exists $t \in \mathbb{R}_{++}$ such that

$$\dim V(t) \geqslant \gamma_5 t^{1/\varrho}\big(\log t\big)^{\nu} > n \geqslant \gamma_6 t^{1/\varrho}\big(\log t\big)^{\nu}, \tag{3.26}$$

where $\gamma_5 \in \mathbb{R}_{++}$ and $\gamma_6 \in \mathbb{R}_{++}$ are independent from n and t. Now set

$$U(t) = \big\{f \in V(t) \;\big|\; \|f\|_2 \leqslant t^{-1}\big\}. \tag{3.27}$$

By Lemma 2.8, $U(t) \subset U_2^{[P]}$. Consequently, it follows from (3.25)–(3.27) and (3.23) that

$$d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \geqslant d_n\big(U(t), L_2(\mathbb{T}^d)\big) \geqslant t^{-1} \asymp n^{-\varrho}\big(\log n\big)^{\nu\varrho}, \tag{3.28}$$

which concludes the proof of (3.17). Next, let us prove (3.18). Given a sufficiently small $\varepsilon \in \mathbb{R}_{++}$, take $t \in \mathbb{R}_{++}$ such that $0 < t-1 < \varepsilon^{-1} \leqslant t$ and $\dim V(t) > 1$. From the above results, it can be seen that

$$\dim V(t) - 1 \leqslant n_\varepsilon\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \leqslant \dim V(t) \tag{3.29}$$

which, together with (3.20), proves (3.18). □

Remark 3.7 We have actually proven a bit more than Theorem 3.6. Namely, suppose that $P(D)$ satisfies the conditions of compactness for $U_2^{[P]}$ stated in Lemma 2.9 and, for every $n \in \mathbb{N}$, let $t(n)$ be the largest number such that $\operatorname{card} K(t(n)) \leqslant n$. Then, for n sufficiently large, we have

$$d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \asymp \frac{1}{t(n)}. \tag{3.30}$$

4 Examples

We first establish norm equivalences and use them to provide examples of asymptotic orders of $d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big)$ for non-degenerate and degenerate differential operators.

Theorem 4.1 *Suppose that $P(D)$ is non-degenerate and set*

$$Q\colon x \mapsto \sum_{\alpha\in\vartheta(A)} x^\alpha. \tag{4.1}$$

Then

$$\big(\forall f \in W_2^{[P]}\big) \quad \|f\|^2_{W_2^{[P]}} \asymp \|f\|^2_{W_2^{[Q]}} \asymp \sum_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2 \asymp \max_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2. \tag{4.2}$$

Moreover, the seminorms in (4.2) *are norms if and only if $0 \in A$.*

Proof. Let $f \in W_2^{[P]}$. It is clear that

$$\sum_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2 \asymp \max_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2. \tag{4.3}$$

Parseval's identity yields

$$\max_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2 = \max_{\alpha\in\vartheta(A)} \sum_{k\in\mathbb{Z}^d} |k|^{2\alpha} |\hat{f}(k)|^2 \leqslant \sum_{k\in\mathbb{Z}^d} \Big(\max_{\alpha\in\vartheta(A)} |k^\alpha|\Big)^2 |\hat{f}(k)|^2. \tag{4.4}$$

Now let $(\mathbb{Z}^d(\alpha))_{\alpha\in\vartheta(A)}$ be a partition of $\mathbb{Z}^d$ such that

$$\max_{\beta\in\vartheta(A)} |k^\beta| = |k^\alpha|, \quad k \in \mathbb{Z}^d(\alpha). \tag{4.5}$$

Then

$$\begin{aligned}\max_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2 &= \max_{\alpha\in\vartheta(A)} \sum_{\alpha'\in\vartheta(A)} \sum_{k\in\mathbb{Z}^d(\alpha')} |k^{2\alpha}|\, |\hat{f}(k)|^2 \\ &\geqslant \sum_{\alpha'\in\vartheta(A)} \sum_{k\in\mathbb{Z}^d(\alpha')} |k^{2\alpha'}|\, |\hat{f}(k)|^2 \\ &= \sum_{k\in\mathbb{Z}^d} \max_{\alpha\in\vartheta(A)} |k^\alpha|^2\, |\hat{f}(k)|^2.\end{aligned} \tag{4.6}$$

Thus,

$$\max_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2 = \sum_{k\in\mathbb{Z}^d} \max_{\alpha\in\vartheta(A)} |k^\alpha|^2\, |\hat{f}(k)|^2. \tag{4.7}$$

Hence, appealing to Corollary 3.1 and (2.10), we obtain

$$\max_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2 \asymp \|f\|^2_{W_2^{[P]}}. \tag{4.8}$$

The relation

$$\max_{\alpha\in\vartheta(A)} \|D^\alpha f\|_2^2 \asymp \|f\|^2_{W_2^{[Q]}} \tag{4.9}$$

follows from the last seminorm equivalence and the identity $\vartheta(\vartheta(A)) = \vartheta(A)$. Therefore, we derive from (4.2) that the seminorms in (4.2) are norms if and only if $0 \in A$. □

4.1 Isotropic Sobolev classes

Let $s \in \mathbb{N}^*$. The isotropic Sobolev space H^s is the Hilbert space of functions $f \in L_2(\mathbb{T}^d)$ equipped with the norm

$$\|\cdot\|_{H^s} : f \mapsto \sqrt{\|f\|_2^2 + \sum_{|\alpha|=s} \|f^{(\alpha)}\|_2^2}. \tag{4.10}$$

Consider

$$P : x \mapsto 1 + \sum_{|\alpha|=s} x^\alpha = \sum_{\alpha\in A} x^\alpha, \tag{4.11}$$

where $A = \{0\} \cup \left\{\alpha \in \mathbb{N}^d \;\middle|\; |\alpha| = s\right\}$. If s is even, it follows directly from Lemma 2.10 that the differential operator $P(D)$ is non-degenerate, and consequently, by Theorem 4.1, $\|\cdot\|_{H^s}$ is equivalent to one of the norms appearing in (4.2) with $\vartheta(A) = \{0\} \cup \left\{su^j \;\middle|\; 1 \leqslant j \leqslant d\right\}$ and

$$Q : x \mapsto 1 + \sum_{j=1}^d x_j^s. \tag{4.12}$$

Moreover, we have $\varrho(A) = s/d$ and $\nu(a) = 0$. Therefore, we retrieve from Theorem 3.6 the well-known result

$$d_n\big(U^s, L_2(\mathbb{T}^d)\big) \asymp n^{-s/d}, \tag{4.13}$$

where U^s denotes the closed unit ball in H^s. This result is a direct generalization of the first result on n-widths established by Kolmogorov in [14].

4.2 Anisotropic Sobolev classes

Given $\beta = (\beta_1, \ldots, \beta_d) \in \mathbb{N}^{*d}$, the anisotropic Sobolev space H^β is the Hilbert space of functions $f \in L_2$ equipped with the norm

$$\|\cdot\|^2_{H^\beta} : f \mapsto \sqrt{\|f\|_2^2 + \sum_{j=1}^d \|f^{(\beta_j u^j)}\|_2^2}. \tag{4.14}$$

Consider the polynomial

$$P\colon x \mapsto 1 + \sum_{j=1}^{d} x_j^{\beta_j} = \sum_{\alpha\in A} x^\alpha, \tag{4.15}$$

where $A = \{0\} \cup \{\beta_j u^j \mid 1 \leqslant j \leqslant d\}$. If the coordinates of β are even, the differential operator $P(D)$ is non-degenerate. Consequently, by Theorem 4.1, $\|\cdot\|_{H^\beta}$ is equivalent to one of the norms in (4.2) with $\vartheta(A) = A$ and

$$Q = P. \tag{4.16}$$

We have

$$\varrho = \varrho(A) = \left(\sum_{j=1}^{d} 1/\beta_j\right)^{-1} \tag{4.17}$$

and $\nu(A) = 0$, and therefore, from Theorem 3.6 we retrieve the known result [13]

$$d_n\big(U^\beta, L_2(\mathbb{T}^d)\big) \asymp n^{-\varrho}, \tag{4.18}$$

where U^β denotes the unit ball in in H^β.

4.3 Classes of functions with a bounded mixed derivative

Let $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ with $0 < \alpha_1 = \cdots = \alpha_{\nu+1} < \alpha_{\nu+2} = \cdots = \alpha_d$ for some $\nu \in \{0, \ldots, d-1\}$. Given a set $e \subset \{1, \ldots, d\}$, let the vector $\alpha(e) \in \mathbb{N}^d$ be defined by $\alpha(e)_j = \alpha_j$ if $j \in e$, and $\alpha(e)_j = 0$ otherwise (in particular, $\alpha(\varnothing) = 0$ and $\alpha(\{1, \ldots, d\}) = \alpha$). The space W_2^α is the Hilbert space of functions $f \in L_2$ equipped with the norm

$$\|\cdot\|_{W_2^\alpha}\colon f \mapsto \sqrt{\sum_{e\subset\{1,\ldots,d\}} \|f^{(\alpha(e))}\|_2^2}. \tag{4.19}$$

Consider

$$P\colon x \mapsto \sum_{e\subset\{1,\ldots,d\}} x^{\alpha(e)} = \sum_{\alpha\in A} x^\alpha, \tag{4.20}$$

where $A = \{\alpha(e) \mid e \subset \{1, \ldots, d\}\}$. If the coordinates of α are even, the differential operator $P(D)$ is non-degenerate and hence, by Theorem 4.1, $\|\cdot\|_{W_2^\alpha}$ is equivalent to one of the norms in (4.2) with $\vartheta(A) = A$ and $Q = P$. We have $\varrho(A) = \alpha_1$ and $\nu(A) = \nu$, and therefore, from Theorem 3.6 we recover the result proven in [1], namely that for n sufficiently large

$$d_n\big(U_2^\alpha, L_2(\mathbb{T}^d)\big) \asymp n^{-\alpha_1}\big(\log n\big)^{\nu\alpha_1}, \tag{4.21}$$

where U_2^α denotes the unit ball in W_2^α. In the particular case when $\alpha = \varrho\mathbf{1}$, we have

$$d_n\big(U_2^{\varrho\mathbf{1}}, L_2(\mathbb{T}^d)\big) \asymp n^{-\varrho}\big(\log n\big)^{(d-1)\varrho}. \tag{4.22}$$

4.4 Classes of functions with several bounded mixed derivatives

Suppose that (3.14) is satisfied. Let W_2^A be the Hilbert space of functions $f \in L_2(\mathbb{T}^d)$ equipped with the norm

$$\|\cdot\|_{W_2^A} : f \mapsto \sqrt{\sum_{\alpha \in A} \|f^{(\alpha)}\|_2^2}. \tag{4.23}$$

Notice that spaces H^s, H^r, and W_2^α are a particular cases of W_2^A. Now consider

$$P : x \mapsto \sum_{\alpha \in A} x^\alpha. \tag{4.24}$$

If the coordinates of every $\alpha \in \vartheta(A)$ are even, the differential operator $P(D)$ is non-degenerate and it follows from Theorem 4.1 that $\|\cdot\|_{W_2^A}$ is equivalent to one of the norms in (4.2). If $\varrho = \varrho(\vartheta(A))$ and $\nu = \nu(\vartheta(A))$, we again retrieve from Theorem 3.6 the result proven in [6], namely that for n sufficiently large

$$d_n\big(U_2^A, L_2(\mathbb{T}^d)\big) \asymp n^{-\varrho}\big(\log n\big)^{\nu\varrho}, \tag{4.25}$$

where U_2^A denotes the unit ball in W_2^A.

4.5 Classes of functions induced by a differential operator

We give two examples of spaces $W_2^{[P]}$ with non-degenerate differential operator $P(D)$ for $d = 2$. Consider the polynomials

$$\begin{cases} P_1 : x \mapsto 8x_1^4 - 4x_1^3 - 3x_1^3 x_2 - 2x_1^2 x_2 - 4x_1 x_2 + 6x_2^2 - 4x_1 - 3x_2 + 13 \\ P_2 : x \mapsto 6x_1^6 + x_1^4 x_2^2 - 6x_1^5 - x_1^3 x_2^2 + 5x_2^4 - 4x_2^3 + 3. \end{cases} \tag{4.26}$$

We have

$$\begin{cases} A_1 & = \{(4,0),(3,0),(2,1),(2,0),(1,1),(0,2),(1,0),(0,1),(0,0)\} \\ \vartheta(A_1) & = \{(4,0),(0,2),(0,0)\} \\ A_2 & = \{(6,0),(4,2),(5,0),(3,2),(0,4),(0,3),(0,0)\} \\ \vartheta(A_2) & = \{(6,0),(4,2),(0,4),(0,0)\}. \end{cases} \tag{4.27}$$

It is easy to verify that $P_1(D)$ and $P_2(D)$ are non-degenerate and that (3.14) holds. Moreover, $\varrho(\vartheta(A_1)) = 4/3$, $\nu(\vartheta(A_1)) = 0$, $\varrho(\vartheta(A_2)) = 8/3$, and $\nu(\vartheta(A_2)) = 1$. We derive from Theorem 3.6 that

$$d_n\big(U^{[P_1]}, L_2(\mathbb{T}^2)\big) \asymp n^{-4/3}, \tag{4.28}$$

and

$$d_n\big(U^{[P_2]}, L_2(\mathbb{T}^2)\big) \asymp n^{-8/3}\big(\log n\big)^{8/3}. \tag{4.29}$$

Let us give an example of a degenerate differential operator. For

$$P_3 : x \mapsto x_1^4 - 2x_1^3 x_2 + x_1^2 x_2^2 + x_1^2 + x_2^2 + 1, \tag{4.30}$$

the differential operator $P_3(D)$ is degenerate, although $P_3 \geqslant 1$ on $\mathbb{R}^2$, and $U^{[P_3]}$ is a compact set in $L_2(\mathbb{T}^2)$. Therefore, we cannot compute $d_n(U^{[P_3]}, L_2(\mathbb{T}^2))$ by using Theorem 3.6. However, by a direct computation we get $\operatorname{card} K(t) \asymp t^{1/2} \log t$. Hence, (3.30) yields

$$d_n\big(U^{[P_3]}, L_2(\mathbb{T}^2)\big) \asymp n^{-2}\big(\log n\big)^2. \tag{4.31}$$

4.6 A conjecture

Suppose that $U_2^{[P]}$ is compact in $L_2(\mathbb{T}^d)$. In view of Lemma 2.9, this is equivalent to the conditions:

(i) For every $t \in \mathbb{R}_+$, $K(t)$ is finite.

(ii) $\tau > 0$.

As mentioned in (3.30), for every $n \in \mathbb{N}$ sufficiently large, if $t(n) \in \mathbb{R}_{++}$ is the maximal number such that $\operatorname{card} K(t(n)) \leqslant n$, then

$$d_n\big(U_2^{[P]}, L_2(\mathbb{T}^d)\big) \asymp \frac{1}{t(n)}. \tag{4.32}$$

This means that the problem of computing the asymptotic order of $d_n(U_2^{[P]}, L_2(\mathbb{T}^d))$ is equivalent to the problem of computing that of $\operatorname{card} K(t)$ when $t \to +\infty$. Let us formulate it as the following conjecture.

Conjecture 4.2 Suppose that, for every $t \in \mathbb{R}_+$, $K(t)$ is finite (the condition $\tau > 0$ is not essential). Then there exist integers α, β, and ν such that $0 < \alpha \leqslant \beta$, $0 \leqslant \nu < d$, and, for t large enough,

$$\operatorname{card} K(t) \asymp t^{\alpha/\beta}\big(\log t\big)^{\nu}. \tag{4.33}$$

In view of (3.20), we know that the conjecture is true when P satisfies conditions (2.7) and (3.9).

Acknowledgment. Dinh Dung's research work is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.01-2014.02, and a part of it was done when Dinh Dung was working as a research professor and Patrick Combettes was visiting at the Vietnam Institute for Advanced Study in Mathematics (VIASM). Both authors thank the VIASM for providing fruitful research environment and working condition. They also thank the LIA CNRS Formath Vietnam for providing travel support.

References

[1] K. I. Babenko, Approximation of periodic functions of many variables by trigonometric polynomials, Soviet Math. Dokl. 1 (1960) 513–516.

[2] K. I. Babenko, Approximation by trigonometric polynomials in a certain class of periodic functions of several variables, Soviet Math. Dokl. 1 (1960) 672–675.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, A simple proof of the restricted isometry property for random matrices, Constr. Approx. 28 (2008) 253–263.

[4] H. H. Bauschke and P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, New York, 2011.

[5] A. Chernov and Dinh Dũng, New explicit-in-dimension estimates for the cardinality of high-dimensional hyperbolic crosses and approximation of functions having mixed smoothness, http://arxiv.org/abs/1309.5170

[6] Dinh Dũng, The number of integral points in some sets and approximation of functions of several variables, Mat. Zametki 36 (1984) 479–491.

[7] Dinh Dũng, Approximation of functions of several variables on a torus by trigonometric polynomials, Math. USSR-Sb. 59 (1988) 247–267.

[8] Dinh Dũng, Best multivariate approximations by trigonometric polynomials with frequencies from hyperbolic crosses, J. Approx. Theory 91 (1997) 205–225.

[9] Dinh Dũng and T. Ullrich, n-widths and ε-dimensions for high-dimensional approximations, Found. Comput. Math. 13 (2013) 965–1003.

[10] D. L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006) 1289–1306.

[11] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich, The Gelfand widths of ℓ_p-balls for $0 < p \leqslant 1$, J. Complexity 26 (2010) 629–640.

[12] S. G. Gindikin, Energy estimates connected with Newton polyhedron, Trudy Moskov. Mat. Obshch. 31 (1974) 189–236.

[13] K. Höllig, Diameters of classes of smooth functions, in: Quantitative Approximation, Academic, New York, 1980, 163–175.

[14] A. N. Kolmogorov, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, Ann. of Math. 37 (1936) 107–110.

[15] A. N. Kolmogorov. Selected Works – Volume 1, Mathematics and Mechanics, Nauka, Moscow 1985 (in Russian).

[16] V. Kůrková and M. Sanguineti, Comparison of worst case errors in linear and neural network approximation, IEEE Trans. Inform. Theory, 48 (2002) 262–275.

[17] A. Kushpel and S. A. Tozoni, Entropy and widths of multiplier operators on two-point homogeneous spaces, Constr. Approx. 35 (2012) 137–180.

[18] J. Liang and T. W. Parks, Kolmogorov n-widths and wavelet representations for signal classes, IEEE Trans. Signal Process. 44 (1996) 1693–1703.

[19] V. P. Mihaĭlov, Behavior at infinity of a certain class of polynomials, Proc. Steklov Inst. Math. 91 (1967) 61–82.

[20] A. Pinkus, n-Widths in Approximation Theory, Springer, New York, 1985.

[21] A. Pinkus, Sparse representations and approximation theory, J. Approx. Theory 163 (2011) 388–412.

[22] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ 1970.

[23] H.-J. Schmeisser and W. Sickel, Spaces of functions of mixed smoothness and approximation from hyperbolic crosses, J. Approx. Theory 128 (2004) 115–150.

[24] L. Schwartz, Théorie des Distributions, 2nd ed., Hermann & Cie, Paris, 1966.

[25] W. Sickel and T. Ullrich, Tensor products of Sobolev–Besov spaces and applications to approximation from the hyperbolic cross, J. Approx. Theory 161 (2009) 748–786.

[26] V. Temlyakov, Approximation of Periodic Functions, Nova Science Publishers, Inc., New York, 1993.

[27] V. M. Tikhomirov, Diameters of sets in functional spaces and the theory of best approximations, Russian Math. Survey 15 (1960) 75–111.

[28] V. M. Tikhomirov, Some Problems in Approximation Theory (in Russian), Moscow State University, 1985.

[29] H. Wang, Widths between the anisotropic spaces and the spaces of functions with mixed smoothness, J. Approx. Theory 164 (2012) 406–430.

[30] P. Wojtaszczyk, On greedy algorithm approximating Kolmogorov widths in Banach spaces, J. Math. Anal. Appl. 424 (2015) 685–695.

[31] L. Zhang, Nearly optimal minimax estimator for high-dimensional sparse linear regression, Ann. Stat. 41 (2013) 2149–2175.