Chapter 2

Preliminaries

In this chapter, we introduce key preliminary material needed later. After some initial notation, we define the domains (the symmetric hypercubes), probability measures (the uniform and Chebyshev measures, respectively) and the Lebesgue–Bochner spaces. We next formalize our main smoothness assumption: namely, holomorphy in suitable (unions of) Bernstein polyellipses. We then introduce orthogonal polynomial expansions and best s-term polynomials approximations, before discussing sequence spaces and best s-term approximations of sequences. Finally, we conclude by reviewing algebraic and exponential rates of convergence for best s-term polynomial approximations, before a short discussion on lower and anchored sets.

2.1 Notation

We first introduce some notation. For $d \in \mathbb{N}$, we write $[d] = \{1, \ldots, d\}$. We also extend this to allow for $d = \infty$, in which case $[d] = \mathbb{N}$ is the set of positive integers. For $d \in \mathbb{N} \cup \{\infty\}$, we write $\boldsymbol{e}_j$, $j \in [d]$, for the standard basis vectors, i.e., $\boldsymbol{e}_j = (\delta_{jk})_{k \in [d]}$. Also for $d \in \mathbb{N} \cup \{\infty\}$, we write $\mathbb{R}^d$ or $\mathbb{C}^d$ for the vector space of real or complex vectors of length d. Note that when $d = \infty$, $\mathbb{R}^d$ and $\mathbb{C}^d$ are the vector spaces $\mathbb{R}^{\mathbb{N}}$ and $\mathbb{C}^{\mathbb{N}}$ of real- or complex-valued sequences indexed over $\mathbb{N}$.

For $1 \leq p \leq \infty$, we write $\|\cdot\|_p$ for the usual vector ℓ^p-norm and for the induced matrix ℓ^p-norm. When $0 < p < 1$, we use the same notation to denote the ℓ^p-quasinorm. For $1 \leq p, q < \infty$ we define the matrix $\ell^{p,q}$-norm of an $m \times n$ matrix $\boldsymbol{G} = (G_{ij})_{i,j=1}^{m,n}$ as $\|\boldsymbol{G}\|_{p,q}^q := \sum_{j=1}^n (\sum_{i=1}^m |G_{ij}|^p)^{q/p}$, and similarly when $p = \infty$ or $q = \infty$.

Throughout this work, we consider sets of multi-indices. Let $d \in \mathbb{N}$. Then we define the multi-index set $\mathcal{F}$ as the set of nonnegative multi-indices, i.e.,

$$\mathcal{F} := \mathbb{N}_0^d = \left\{ \boldsymbol{\nu} = (\nu_k)_{k=1}^d : \nu_k \in \mathbb{N}_0 \right\}, \quad d < \infty. \tag{2.1}$$

When $d = \infty$, we consider multi-indices in $\mathbb{N}_0^{\mathbb{N}}$ with at most finitely many nonzero terms, i.e., we define

$$\mathcal{F} := \left\{ \boldsymbol{\nu} = (\nu_k)_{k=1}^{\infty} \in \mathbb{N}_0^{\mathbb{N}} : |\{k : \nu_k \neq 0\}| < \infty \right\}, \quad d = \infty. \tag{2.2}$$

In either finite or infinite dimensions, we write $\mathbf{0}$ and $\mathbf{1}$ for the multi-indices consisting of all zeros and all ones, respectively. Finally, the inequality $\boldsymbol{\mu} \leq \boldsymbol{\nu}$ is understood componentwise for any multi-indices $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$.

2.2 Domains and function spaces

Let $\varrho = \varrho^{(1)}$ be a probability measure on $[-1, 1]$. In this work, we focus on two main examples, the uniform and Chebyshev (arcsine) measures. These are defined by

$$\mathrm{d}\varrho(y) = 2^{-1}\,\mathrm{d}y, \quad \text{and} \quad \mathrm{d}\varrho(y) = \frac{1}{\pi\sqrt{1-y^2}}\,\mathrm{d}y, \quad y \in \mathcal{U}, \tag{2.3}$$

respectively. See Chapter 11 for a short discussion on other domains and measures. In finite dimensions, we let $\mathcal{U} = [-1, 1]^d$ be the symmetric d-dimensional hypercube and write $\boldsymbol{y} = (y_1, \ldots, y_d) \in \mathcal{U}$ for the variable in this domain. We define a probability measure on $\mathcal{U}$ as the product measure

$$\varrho = \varrho^{(d)} := \varrho^{(1)} \times \cdots \times \varrho^{(1)}.$$

In particular, the d-dimensional uniform and Chebyshev measures are given by

$$\mathrm{d}\varrho(\boldsymbol{y}) = 2^{-d}\,\mathrm{d}\boldsymbol{y} \quad \text{and} \quad \mathrm{d}\varrho(\boldsymbol{y}) = \prod_{k=1}^{d} \frac{1}{\pi\sqrt{1-y_k^2}}\,\mathrm{d}\boldsymbol{y}, \quad \forall \boldsymbol{y} \in \mathcal{U},$$

respectively. In infinite dimensions, we consider the domain $\mathcal{U} = [-1, 1]^{\mathbb{N}}$ and write $\boldsymbol{y} = (y_1, y_2, \ldots) \in \mathcal{U}$ for the variable in this domain. The Kolmogorov extension theorem (see, e.g., [136, Section 2.4]) guarantees the existence of a tensor-product probability measure on $\mathcal{U}$, which we denote as

$$\varrho = \varrho^{(\infty)} = \prod_{k \in \mathbb{N}} \varrho^{(1)}.$$

In either finite or infinite dimensions, for $1 \leq p \leq \infty$ we write $L^p_\varrho(\mathcal{U})$ for the corresponding weighted Lebesgue spaces of complex scalar-valued functions over $\mathcal{U}$ and $\|\cdot\|_{L^p_\varrho(\mathcal{U})}$ for their norms.

Throughout, we let $\mathcal{V}$ be a separable Hilbert space over $\mathbb{C}$ (it presents few difficulties to consider a complex field instead of the real field). We write $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ and $\|\cdot\|_{\mathcal{V}}$ for its inner product and norm. We define the weighted (Lebesgue-)Bochner space $L^p_\varrho(\mathcal{U}; \mathcal{V})$ as the space consisting of (equivalence classes of) strongly ϱ-measurable functions $f : \mathcal{U} \to \mathcal{V}$ for which $\|f\|_{L^p_\varrho(\mathcal{U};\mathcal{V})} < \infty$, where

$$\|f\|_{L^p_\varrho(\mathcal{U};\mathcal{V})} := \begin{cases} \left(\int_{\mathcal{U}} \|f(\boldsymbol{y})\|^p_{\mathcal{V}}\,\mathrm{d}\varrho(\boldsymbol{y})\right)^{1/p} & 1 \leq p < \infty, \\ \operatorname{ess\,sup}_{\boldsymbol{y} \in \mathcal{U}} \|f(\boldsymbol{y})\|_{\mathcal{V}} & p = \infty. \end{cases}$$

Note that $L^p_\varrho(\mathcal{U})$ is a special case of $L^p_\varrho(\mathcal{U}; \mathcal{V})$ corresponding to $\mathcal{V} = (\mathbb{C}, |\cdot|)$.

When $\mathcal{V}$ is infinite dimensional, we usually cannot work directly with it. Hence, we consider a finite-dimensional discretization

$$\mathcal{V}_h \subseteq \mathcal{V}. \tag{2.4}$$

Here $h > 0$ denotes a discretization parameter, e.g., the mesh size in the case of a finite element discretization (as is common in parametric DEs). In the context of finite elements, assuming (2.4) corresponds to considering so-called *conforming* discretizations. We let $\{\varphi_k\}_{k=1}^K$ be a (not necessarily orthonormal) basis of $\mathcal{V}_h$, where $K = K(h) = \dim(\mathcal{V}_h)$. We write $\mathcal{P}_h : \mathcal{V} \to \mathcal{V}_h$ for the orthogonal projection onto $\mathcal{V}_h$ and, for $f \in L^2_\varrho(\mathcal{U}; \mathcal{V})$, we let $\mathcal{P}_h f \in L^2_\varrho(\mathcal{U}; \mathcal{V}_h)$ be the function defined almost everywhere as

$$(\mathcal{P}_h f)(\boldsymbol{y}) = \mathcal{P}_h(f(\boldsymbol{y})), \quad \boldsymbol{y} \in \mathcal{U}. \tag{2.5}$$

2.3 Holomorphy

Here we recall the definitions of holomorphy and holomorphic extension for Hilbert-valued functions. We note that equivalent definitions are possible (see, e.g., [77, Chapter 2]) and that the definition employed in this work is based on the notion of the Gateaux partial derivative. For other details on differentiability of Hilbert-valued functions we refer to [22, Chapter 17], and the references therein. Note the following definitions apply in both the finite- ($d \in \mathbb{N}$) and infinite- ($d = \infty$) dimensional settings, where we recall that $[d] = \mathbb{N}$ and $\mathbb{C}^d = \mathbb{C}^\mathbb{N}$ when $d = \infty$.

Definition 2.1 (Holomorphy; finite- or infinite-dimensional case). Let $d \in \mathbb{N} \cup \{\infty\}$, $\mathcal{O} \subseteq \mathbb{C}^d$ be an open set and $\mathcal{V}$ be a separable Hilbert space. A function $f : \mathcal{O} \to \mathcal{V}$ is *holomorphic in* $\mathcal{O}$ if it is holomorphic with respect to each variable in $\mathcal{O}$. That is to say, for any $z \in \mathcal{O}$ and any $j \in [d]$, the following limit exists in $\mathcal{V}$:

$$\lim_{\substack{h \in \mathbb{C} \\ h \to 0}} \frac{f(z + h\boldsymbol{e}_j) - f(z)}{h} \in \mathcal{V}.$$

Let $f : \mathcal{U} \to \mathcal{V}$ and $\mathcal{U} \subset \mathcal{O} \subseteq \mathbb{C}^d$ be an open set. If there is a function $\tilde{f} : \mathcal{O} \to \mathcal{V}$ that is holomorphic in $\mathcal{O}$ and for which $\tilde{f}|_{\mathcal{U}} = f$, then we say that f has a *holomorphic extension to* $\mathcal{O}$, or simply, that f is *holomorphic in* $\mathcal{O}$. In this case, we also define $\|f\|_{L^\infty(\mathcal{O};\mathcal{V})} := \|\tilde{f}\|_{L^\infty(\mathcal{O};\mathcal{V})}$ or, when $\mathcal{V} = \mathbb{C}$, simply $\|f\|_{L^\infty(\mathcal{O})}$. If $\mathcal{O}$ is a closed set, then we say that f is holomorphic in $\mathcal{O}$ if it has a holomorphic extension to some open neighborhood of $\mathcal{O}$.

We are interested in approximating Hilbert-valued functions $f : \mathcal{U} \to \mathcal{V}$ that are holomorphic in suitable complex regions containing $\mathcal{U}$ – specifically, regions defined by *Bernstein (poly)ellipses*. When $d = 1$ the Bernstein ellipse of parameter $\rho > 1$ is given by

$$\mathcal{E}_\rho = \left\{ \frac{1}{2}(z + z^{-1}) : z \in \mathbb{C},\ 1 \leq |z| \leq \rho \right\} \subset \mathbb{C}.$$

This is an ellipse with ± 1 as its foci and major and minor semi-axis lengths $\frac{1}{2}(\rho \pm \rho^{-1})$. For $d \in \mathbb{N} \cup \{\infty\}$, given $\boldsymbol{\rho} = (\rho_j)_{j=1}^d \in \mathbb{R}^d$ with $\boldsymbol{\rho} > \boldsymbol{1}$, we define the Bernstein polyellipse as the Cartesian product

$$\mathcal{E}(\boldsymbol{\rho}) = \mathcal{E}(\rho_1) \times \mathcal{E}(\rho_2) \times \cdots \subset \mathbb{C}^d.$$

We denote the class of Hilbert-valued functions that are holomorphic in $\mathcal{E}(\boldsymbol{\rho})$ with norm at most one as

$$\mathcal{B}(\boldsymbol{\rho}) = \left\{ f : \mathcal{U} \to \mathcal{V},\ f \text{ holomorphic in } \mathcal{E}(\boldsymbol{\rho}),\ \|f\|_{L^\infty(\mathcal{E}(\boldsymbol{\rho});\mathcal{V})} \leq 1 \right\}. \tag{2.6}$$

In infinite dimensions, we also consider a class of functions that are holomorphic in a certain union of Bernstein polyellipses. Let $0 < p < 1$, $\varepsilon > 0$ and $\boldsymbol{b} = (b_j)_{j \in \mathbb{N}} \in \ell^p(\mathbb{N})$. We define

$$\mathcal{R}(\boldsymbol{b}, \varepsilon) = \bigcup \left\{ \mathcal{E}(\boldsymbol{\rho}) : \boldsymbol{\rho} \geq \mathbf{1},\ \sum_{j=1}^{\infty} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) b_j \leq \varepsilon \right\}.$$

In analogy with $\mathcal{B}(\boldsymbol{\rho})$, we write

$$\mathcal{B}(\boldsymbol{b}, \varepsilon) = \left\{ f : \mathcal{U} \to \mathcal{V},\ f \text{ holomorphic in } \mathcal{R}(\boldsymbol{b}, \varepsilon), \|f\|_{L^\infty(\mathcal{R}(\boldsymbol{b},\varepsilon);\mathcal{V})} \leq 1 \right\} \tag{2.7}$$

for the corresponding space of functions that are holomorphic in $\mathcal{R}(\boldsymbol{b}, \varepsilon)$ with norm at most one.

2.4 Orthogonal polynomials, polynomial expansions and best s-term polynomial approximation

Under mild assumptions on $\varrho^{(1)}$ (see, e.g., [106, Section 2.1] or [132, Section 2.2]), there exists a unique orthonormal polynomial basis $\{\Psi_\nu\}_{\nu \in \mathbb{N}_0}$ of $L^2_\varrho([-1, 1])$, where $\Psi_\nu = \Psi_\nu^{(1)}$ is a polynomial of degree ν. For the measures (2.3), these are the Legendre and Chebyshev polynomials, respectively. Given the corresponding tensor-product measure ϱ on $\mathcal{U} = [-1, 1]^d$, we construct an orthonormal basis

$$\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \mathcal{F}} \subset L^2_\varrho(\mathcal{U})$$

of $L^2_\varrho(\mathcal{U})$ via tensorization

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{k \in [d]} \Psi_{\nu_k}(y_k), \quad \boldsymbol{y} \in \mathcal{U},\ \boldsymbol{\nu} \in \mathcal{F}.$$

Note that $\Psi_0^{(1)} = 1$ since $\varrho^{(1)}$ is a probability measure. Therefore, since $\boldsymbol{\nu} \in \mathcal{F}$ has only finitely many nonzero entries, in infinite dimensions this equivalent to

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{k : \nu_k \neq 0} \Psi_{\nu_k}(y_k),$$

which is a product of finitely many terms.

Let $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$. Then it has the convergent expansion (in $L^2_\varrho(\mathcal{U};\mathcal{V})$) given by

$$f = \sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}, \quad c_{\boldsymbol{\nu}} := \int_{\mathcal{U}} f(\boldsymbol{y})\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})\,\mathrm{d}\varrho(\boldsymbol{y}) \in \mathcal{V}, \tag{2.8}$$

where the *coefficients* $c_{\boldsymbol{\nu}}$ are elements of $\mathcal{V}$. Now let $S \subset \mathcal{F}$ be a finite index set and

$$\mathcal{P}_{S;\mathcal{V}} = \Big\{\sum_{\boldsymbol{\nu}\in S} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}} : c_{\boldsymbol{\nu}} \in \mathcal{V}\Big\} \subset L^2_\varrho(\mathcal{U};\mathcal{V}). \tag{2.9}$$

Then the $L^2(\mathcal{U};\mathcal{V})$-norm *best s-term polynomial approximation* f_s of f is defined as

$$f_s \in \operatorname{argmin}\big\{\|f-g\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} : g \in \mathcal{P}_{S,\mathcal{V}},\ S \subset \mathcal{F},\ |S| = s\big\}. \tag{2.10}$$

Note that f_s has the explicit expression

$$f_s = \sum_{\boldsymbol{\nu}\in S^*} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}},$$

where $S^* \subset \mathcal{F}$, $|S^*| = s$, is a set of consisting of the multi-indices of the largest s values of the coefficient norms $(\|c_{\boldsymbol{\nu}}\|_{\mathcal{V}})_{\boldsymbol{\nu}\in\mathbb{N}_0^d}$. By Parseval's identity, the error of this approximation satisfies

$$\|f - f_s\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \sqrt{\sum_{\boldsymbol{\nu}\notin S^*} \|c_{\boldsymbol{\nu}}\|^2_{\mathcal{V}}}. \tag{2.11}$$

2.5 Sequence spaces and best s-term approximation of sequences

The expression (2.11) motivates studying s-term approximation of sequences of polynomial coefficients. To do this, we now introduce some further notation.

Let $\Lambda \subseteq \mathcal{F}$ denote a (possibly infinite) multi-index set. We write $\boldsymbol{v} = (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ for a sequence with $\mathcal{V}$-valued entries $v_{\boldsymbol{\nu}} \in \mathcal{V}$. For $1 \le p \le \infty$, we define the space $\ell^p(\Lambda;\mathcal{V})$ as the set of those sequences $\boldsymbol{v} = (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ for which $\|\boldsymbol{v}\|_{p;\mathcal{V}} < \infty$, where

$$\|\boldsymbol{v}\|_{p;\mathcal{V}} := \begin{cases} \left(\sum_{\boldsymbol{\nu}\in\Lambda} \|v_{\boldsymbol{\nu}}\|^p_{\mathcal{V}}\right)^{1/p} & 1 \le p < \infty, \\ \sup_{\boldsymbol{\nu}\in\Lambda} \|v_{\boldsymbol{\nu}}\|_{\mathcal{V}} & p = \infty. \end{cases}$$

Note that $\ell^2(\Lambda;\mathcal{V})$ is a Hilbert space with inner product

$$\langle \boldsymbol{u}, \boldsymbol{v}\rangle_{2;\mathcal{V}} = \sum_{\boldsymbol{\nu}\in\Lambda} \langle u_{\boldsymbol{\nu}}, v_{\boldsymbol{\nu}}\rangle_{\mathcal{V}}.$$

On occasion, we will consider complex, scalar-valued sequences. In this case, $\mathcal{V} = (\mathbb{C}, |\cdot|)$ in the various definitions above. For ease of notation, we simply write $\ell^p(\Lambda)$, $\|\cdot\|_p$, $\langle\cdot,\cdot\rangle_2$ and so forth in this case.

Definition 2.2 (Sparsity). Let $\Lambda \subseteq \mathcal{F}$ and $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ be a $\mathcal{V}$-valued sequence. The *support* of $\boldsymbol{c}$ is the set

$$\operatorname{supp}(\boldsymbol{c}) = \{\boldsymbol{\nu} \in \Lambda : \|c_{\boldsymbol{\nu}}\|_{\mathcal{V}} \neq 0\}.$$

A sequence is s-*sparse* for some $s \in \mathbb{N}_0$ satisfying $s \leq |\Lambda|$ if it has at most s nonzero entries, i.e.,

$$|\operatorname{supp}(\boldsymbol{c})| \leq s.$$

Definition 2.3 (Best s-term approximation error). Let $\Lambda \subseteq \mathcal{F}$, $0 < p \leq \infty$, $\boldsymbol{c} \in \ell^p(\Lambda; \mathcal{V})$ and $s \in \mathbb{N}_0$ with $s \leq |\Lambda|$. The *ℓ^p-norm best s-term approximation error* of $\boldsymbol{c}$ is

$$\sigma_s(\boldsymbol{c})_{p;\mathcal{V}} = \min\left\{\|\boldsymbol{c} - \boldsymbol{z}\|_{p;\mathcal{V}} : \boldsymbol{z} \in \ell^p(\Lambda; \mathcal{V}),\ |\operatorname{supp}(\boldsymbol{z})| \leq s\right\}.$$

Let $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}}$ be the coefficients of some function $f \in L^2_{\varrho}(\mathcal{U}; \mathcal{V})$, as defined in (2.8). Then, when $p = 2$, we have the following:

$$\sigma_s(\boldsymbol{c})_{2;\mathcal{V}} = \|f - f_s\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})},$$

where f_s is its best s-term polynomial approximation (2.10). Therefore, we can study the error of f_s by studying the quantity $\sigma_s(\boldsymbol{c})_{2;\mathcal{V}}$. For notational purposes, we denote this quantity in terms of the coefficients $\boldsymbol{c}$. However, on some occasions, this term is expressed as $\sigma_s(f)_{2;\mathcal{V}}$ instead.

2.6 Rates of best s-term polynomial approximation

As noted, best s-term polynomial approximation of holomorphic functions is a well-studied subject, especially in the context of solutions of parametric DEs. See, e.g., [23–25, 27, 36, 42, 43, 75, 115, 139, 141] and, in particular, [41] and [8, Chapter 3]. In this section, we recap two standard types of error decay rates for this approximation, those of *algebraic* and *exponential* type, respectively. Note that these results are for Chebyshev and Legendre polynomial approximations – the main focus of the work. The latter type of decay rate holds in finite dimensions, while the former holds in both finite and infinite dimensions. In this work, these error decay rates serve as the optimal benchmark against which to compare the approximations computed from sample values.

The following two results are standard, and have appeared in various different guises in the aforementioned works.

Theorem 2.4 (Algebraic rates of convergence; finite-dimensional case). *Let $0 < p \leq 1$ and $f \in \mathcal{B}(\boldsymbol{\rho})$ for some $\boldsymbol{\rho} > \mathbf{1}$. Let $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathbb{N}_0^d}$ be as in (2.8). Then, for every $s \geq 1$ there are sets $S_1, S_2 \subset \mathcal{F}$, $|S_1|, |S_2| \leq s$, such that*

$$\|f - f_{S_1}\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1/2-1/p}, \quad \|f - f_{S_2}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1-1/p}, \quad (2.12)$$

where $f_{S_i} = \sum_{\boldsymbol{\nu} \in S_i} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ *for* $i = 1, 2$ *and* $C = C(d, p, \boldsymbol{\rho}) > 0$ *depends on* d, p *and* $\boldsymbol{\rho}$ *only.*

Theorem 2.5 (Algebraic rates of convergence; infinite-dimensional case). *Let* $0 < p < 1$, $\varepsilon > 0$, $\boldsymbol{b} = (b_j)_{j \in \mathbb{N}} \in \ell^p(\mathbb{N})$ *and* $f \in \mathcal{B}(\boldsymbol{b}, \varepsilon)$, *where* $\mathcal{B}(\boldsymbol{b}, \varepsilon)$ *is as in* (2.7). *Then, for every* $s \geq 1$ *there are sets* $S_1, S_2 \subset \mathcal{F}$, $|S_1|, |S_2| \leq s$, *such that*

$$\|f - f_{S_1}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1/2-1/p}, \quad \|f - f_{S_2}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1-1/p}, \tag{2.13}$$

where $f_{S_i} = \sum_{\boldsymbol{\nu} \in S_i} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ *for* $i = 1, 2$ *and* $C = C(\boldsymbol{b}, \varepsilon, p) > 0$ *depends on* $\boldsymbol{b}$, ε *and* p *only.*

Observe that the curse of dimensionality is not avoided in the constant $C(d, p, \boldsymbol{\rho})$ in (2.12), but it is avoided in the rate. Conversely, (2.13) holds in infinite dimensions.

We next state a result on exponential convergence in finite dimensions. Such rates have been established in various different works (see, e.g., [23, 24, 41, 115, 141]). The following result is a minor modification of [8, Theorem 3.25], in which we allow arbitrary $s \geq 1$ at the expense of a constant C in the error bound.

Theorem 2.6 (Exponential rates of convergence; finite-dimensional case). *Let* $f \in \mathcal{B}(\boldsymbol{\rho})$ *for some* $\boldsymbol{\rho} > \mathbf{1}$ *and* $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathbb{N}_0^d}$ *be as in* (2.8). *Then, for every* $s \geq 1$ *there is a set* $S \subset \mathcal{F}$, $|S| \leq s$, *such that*

$$\|f - f_S\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f - f_S\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq C \cdot \exp\left(-\gamma s^{1/d}\right), \tag{2.14}$$

for all

$$0 < \gamma < (d+1)^{-1} \left(d! \prod_{j=1}^{d} \ln(\rho_j) \right)^{1/d}, \tag{2.15}$$

where $f_S = \sum_{\boldsymbol{\nu} \in S} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ *and* $C = C(d, \gamma, p, \boldsymbol{\rho}) > 0$ *is a constant depending on* d, γ, p *and* $\boldsymbol{\rho}$ *only.*

In Appendix A we show how these three theorems can be obtained as immediate consequences of several more general results.

Remark 2.7. It is possible to improve the rate (2.14) by removing the $(d+1)^{-1}$ factor in (2.15) [141]. The difficulty in doing this is that such rates are not necessarily attained in lower sets (this is, however, true if $\boldsymbol{\rho}$ is sufficiently large – see [8, Lemma 7.20]). As we discuss next, lower sets are a crucial ingredient in our analysis. Fortunately, the rates described in Theorem 2.6 can always be attained in lower sets.

2.7 Lower and anchored sets

Our objective in this work is to construct a polynomial approximation that attains error bounds that are similar to those of the best s-term approximation f_s, for any holomorphic function f. Hence, ideally, we would have access to the multi-index set S corresponding to the largest s coefficients of f (measured in the $\mathcal{V}$-norm). As discussed, this is not possible in general, since the only information we have about f is its values at a finite number of sample points. Another problem is that such coefficients could occur at arbitrarily large multi-indices, thus necessitating a search over infinitely many multi-indices. Fortunately, it is well known that near-best s-term polynomial approximations can be constructed using sets of multi-indices with additional structure. These are *lower* sets (used in the finite-dimensional case) and *anchored* sets (used in the infinite-dimensional case). Classical references for lower and anchored sets include [50, 90, 96, 138]. More recently, these structures have been used extensively in the construction of interpolation, least-squares and compressed sensing schemes for polynomial approximation with desirable sample complexity bounds (see, e.g., [8] and references therein).

Definition 2.8. A set $\Lambda \subseteq \mathcal{F}$ is *lower* if the following holds for every $\boldsymbol{\nu}, \boldsymbol{\mu} \in \mathcal{F}$:

$$(\boldsymbol{\nu} \in \Lambda \text{ and } \boldsymbol{\mu} \leq \boldsymbol{\nu}) \Rightarrow \boldsymbol{\mu} \in \Lambda.$$

A set $\Lambda \subseteq \mathcal{F}$ is *anchored* if it is lower and if the following holds for every $j \in \mathbb{N}$:

$$\boldsymbol{e}_j \in \Lambda \Rightarrow \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_j\} \subseteq \Lambda.$$

Lower sets are typically used in finite-dimensional settings, with anchored sets being employed in infinite dimensions. They are key concepts that we exploit in this work. To underscore their usefulness, we remark in passing that the rates articulated in not just Theorem 2.6, but also Theorems 2.4 and 2.5, can all be attained using s-term approximations in lower or anchored sets, subject to some modifications. See Appendix A.