Multilevel approximation of parametric and stochastic PDEs*

Jakob Zech[†] Dinh Dũng[‡] Christoph Schwab[†]

March 27, 2018

Abstract

We analyze the complexity of the sparse-grid interpolation and sparse-grid quadrature of countably-parametric functions which take values in separable Banach spaces with unconditional bases. Under the provision of a suitably quantified holomorphic dependence on the parameters, we establish dimension-independent convergence rate bounds for sparse-grid approximation schemes. Analogous results are shown in the case that the parametric solutions are obtained as solutions of corresponding parametric-holomorphic, nonlinear operator equations as considered in [A. Cohen and A. Chkifa and Ch. Schwab: Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs, Journ. Math. Pures et Appliquees **103**(2) 400-428 (2015)] by means of stable, finite dimensional approximations, for example nonlinear Petrov-Galerkin projections. Error and convergence rate bounds for constructive and explicit multilevel, sparse tensor approximation schemes combining sparse-grid interpolation in the parameter space and general, multilevel discretization schemes in the physical domain are proved. The results considerably generalize several earlier works in terms of the admissible multilevel approximations in the physical domain (comprising general stable Petrov-Galerkin and discrete Petrov-Galerkin schemes, collocation and stable domain approximations) and in terms of the admissible operator equations (comprising smooth, nonlinear locally well-posed operator equations). Additionally, a novel, general computational strategy to localize sequences of nested index sets is given for the anisotropic Smolyak scheme realizing best n-term benchmark convergence rates. We also consider Smolyak-type quadratures in this general setting, for which we establish improved convergence rates based on cancellations in gpc expansions due to symmetries of the probability measure [J. Zech and Ch. Schwab: Convergence rates of high dimensional Smolyak quadrature, Report 2017-27, SAM ETH Zürich].

Several examples illustrating the abstract theory include domain uncertainty quantification ("UQ" for short) for general, linear, second order, elliptic advection-reaction-diffusion equations on polygonal domains, where optimal convergence rates of FEM are known to require local mesh refinement near corners. For these equations, we also consider a combined sparse-grid scheme in physical and parameter space, affording complexity similar to the recent multiindex stochastic collocation approach. Further applications of the presently developed theory comprise evaluations of posterior expectations in Bayesian inverse problems.

*Jakob Zech and Christoph Schwab are supported in part by the Swiss National Science Foundation under Grant No. 159940. Dinh Dũng's research work was founded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.01-2017.05. A part of this work was done when Dinh Dũng was working at the Vietnam Institute for Advanced Study in Mathematics (VIASM). He would like to thank the VIASM for providing a fruitful research environment and working condition.

[†]Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland (jakob.zech@sam.math.ethz.ch, christoph.schwab@sam.math.ethz.ch).

[‡]Information Technology Institute, Vietnam National University, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam (dinhzung@gmail.com).

1 Introduction

A key task in computational UQ is the efficient numerical treatment of partial differential equations (PDEs) with uncertain, distributed input data. This is to say that uncertain input data takes values in (generally infinite-dimensional, separable) Banach spaces rather than in $\mathbb{R}^n$. Cases in point are for example coefficients describing material properties of heterogeneous media, source terms or geometries of physical domains to name but a few. In recent years, substantial attention has been focused on the mathematical and numerical analysis of partial differential equation models in the sciences and engineering with parametric dependence of *distributed* uncertain input data, taking values in function spaces on domains. Choosing (computationally convenient) bases in these spaces, thereby parametrizing the distributed uncertainty, results in *parametric UQ on possibly high-dimensional parameter spaces*, the parameter dimension being given by the number of active basis elements in uncertainty parametrization.

1.1 Previous approximation results for parametric solution families

Uncertainty parametrization renders the solution of the forward problem likewise parametric, thereby leading to the mathematical and computational approximation problem of many-parametric "forward" solution maps. A set of results, starting with [49], exploited *analyticity* of the parametric solution at the origin, and via bounds on the radius of convergence of Taylor generalized polynomial chaos ("gpc for short") expansions, inferred convergence rates of n-term truncated Taylor gpc expansions of the parametric solution. This line of work was expanded in [18, 19, 15], where *holomorphy* of the parameter-to-solution maps was quantified through radii of poly-discs respectively of poly-ellipses, and dimension independent rates of convergence of best N-term approximations were obtained. Best N-term approximations of, generally, infinite-dimensional gpc expansions are, as a rule, not constructive, but provide a benchmark for convergence rates that can, in principle, be achieved.

An important role in the approximation of such many-parametric solutions has been taken in recent years by *generalized polynomial chaos* expansions, and their numerical approximations by collocation and Galerkin projections. We refer to [44, 31] and the references there for anisotropic spectral interpolation methods in parameter space. Under mild additional assumptions, constructive *stochastic collocation* schemes have been proposed. They were shown in [12] to produce sequences of gpc approximations with the same convergence rates as N-term approximations by sampling the parameter spaces on unisolvent *generalized sparse-grids*, adapted to the structure of the parametric solution. Algorithms for the computational construction of rate-optimal (w.r.t. a cost-benefit criterion) index sets were proposed in [3] and analyzed in [43]. The construction in [3] involved the numerical solution of a certain knapsack problem, based on apriori given bounds on gpc coefficients.

The results in the mentioned references were mostly (with the exception of [34, 15]) for model linear parametric elliptic diffusion problems. An extension of the convergence analysis to generic parametric solution families taking values in (separable) Hilbert spaces was developed in [43]: the focus of this work is on the collocation error analysis in the parameter space. Discretization in physical space was, in principle, covered in a single-level setting, i.e., with the same discretization used for all collocation points. It has, however, been stipulated already early on (see, e.g., [6, 5]) that substantial efficiency could be gained by adapting the discretization resolution in physical space to the size of the details with respect to the parameters.

Computational implementation of interpolatory or collocation approximations in the parameter domain requires, as a rule, also *discretization* of the corresponding operator equation. Here, *sparsity of sequences of discretization schemes* has been identified as a crucial ingredient in viable numerical approximation schemes; cases in point are multilevel Monte Carlo Methods (see, e.g., [27] and references therein), and generalized sparse-grid approaches. Recently, the so-called "multi-index stochastic collocation" method (MISC) for the solution of many-parametric PDEs was proposed in [33, 32]. This method combines sparse-grid approximations in both the parameter space and the physical space, yielding a multilevel method that was shown to achieve dimension independent convergence rates for the numerical integration of a model elliptic problem with a random diffusion coefficient. Moreover, benchmark approximation rates for fully discrete, multilevel approximations have been recently proved, in [2].

1.2 Contributions of the present paper

The contributions of the present paper consist, on the one hand, in a generalization of [43, 32] in that Banach space valued, countably-parametric solution families of general, locally well-posed nonlinear operator equations with holomorphic dependence on the uncertain inputs are admitted. This setting accommodates certain nonlinear partial differential equations where growth conditions for the (analytic) nonlinearity preclude a Hilbert space setting. Moreover, we propose novel algorithms for the construction of nested, downward closed index sets affording near-optimal convergence rates for stochastic collocation which obviate the use of computational knapsack solvers as proposed, for example, in [3, 43]. They provide an apriori, constructive localization of near optimal index sets (in the sense that best N-term rates are achieved by sparse tensor gpc interpolants on the corresponding unisolvent sparse-grids). The present results contain, finally, a "fully discrete", sparse tensor version of [15], in that analogous nonlinear, implicit holomorphic-parametric operator equations are admitted. They contain in particular earlier results for certain operator equations in [17, 19, 28, 43]. We illustrate the abstract framework by extending earlier results for coefficient and domain uncertainty for second order, diffusion equations with uncertain, parametric coefficients to general advection-reation-diffusion PDEs in polygons, with uncertain coefficients and/or uncertain domains. Here, the holomorphy of the data-to-solution map is employed in corner-weighted Kondrat'ev space in D, accomodating elliptic regularity shifts and optimal FE convergence rates on corner-refined meshes in D. Further examples include the Maxwell and Navier-Stokes Equations in uncertain domains [21, 38].

1.3 Outline

The present paper is structured as follows. After developing abstract results on multilevel approximation of Banach space valued $(\boldsymbol{b}, \varepsilon)$-holomorphic maps in Section 2, which could be of independent interest, we specialize these results in Section 3 to stable Petrov-Galerkin approximation of solution families of countably-parametric, holomorphic nonlinear operator equations. Section 4 discusses several examples to illustrate the scope of the foregoing, abstract results, covering in particular general parametric advection-reaction-diffusion models with uncertain coefficients, in polygonal domains and domain UQ. Finally, in Section 5, we investigate the sharpness of the theoretical results in several numerical experiments. The appendix contains proofs on sequence approximation results which are the key technical ingredients in Section 2.1.1.

1.4 Notation

We use standard notation. Specifically, we shall use multiindices $\boldsymbol{\nu} = (\nu_j)_{j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$ being sequences of nonnegative integers. Denote by $\operatorname{supp}\boldsymbol{\nu} := \{j \in \mathbb{N} : \nu_j > 0\}$ the support of $\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}}$, and by $\mathcal{F}$ the set of all $\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}}$ such that $\operatorname{supp}\boldsymbol{\nu}$ is a finite subset of $\mathbb{N}$. Evidently, the *order* of a multiindex $\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}}$, which is denoted by $|\boldsymbol{\nu}| := \sum_{j\in\mathbb{N}} \nu_j$, is finite if and only if $\boldsymbol{\nu} \in \mathcal{F}$. For two multiindices $\boldsymbol{\nu}$, $\boldsymbol{\mu} \in \mathbb{N}_0^{\mathbb{N}}$, inequalities such as $\boldsymbol{\nu} \le \boldsymbol{\mu}$ are always understood componentwise, i.e. $\boldsymbol{\nu} \le \boldsymbol{\mu}$ if and only if $\nu_j \le \mu_j$ for all $j \in \mathbb{N}$. For a scalar $c \in \mathbb{C}$ we write $c\boldsymbol{\nu}$ to denote $(c\nu_j)_{j\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$ and similarly $\boldsymbol{\nu}^c := (\nu_j^c)_{j\in\mathbb{N}}$. If $\boldsymbol{c} \in \mathbb{C}^{\mathbb{N}}$ on the other hand, then $\boldsymbol{\nu}^{\boldsymbol{c}} := \prod_{\{j\,:\,\nu_j\neq 0\}} \nu_j^{c_j}$. An important role in the proofs of the convergence rate bounds for sparse-grid interpolants in the present paper is taken by particular subsets of $\mathbb{N}_0^{\mathbb{N}}$ introduced in [52].

Definition 1.1. *For $k \in \mathbb{N}$, $\mathcal{F}_k \subseteq \mathcal{F}$ denotes the sets of finitely supported multiindices defined by*

$$\mathcal{F}_k := \{\boldsymbol{\nu} \in \{0, k, k+1, \dots\}^{\mathbb{N}} \,:\, |\boldsymbol{\nu}| < \infty\}. \tag{1.1}$$

Obviously, $\mathcal{F} = \mathcal{F}_1 \supset \mathcal{F}_2 \supset \dots$ and all $\mathcal{F}_k$ are countable. The sets $\mathcal{F}_k$ arise in indexing generalized polynomial chaos ("gpc" for short) expansions of homogeneity at least $k \ge 1$.

We shall make use of *downward closed sets of multiindices*. A subset $\Lambda \subseteq \mathcal{F}$ is labelled *downward closed*, if $\boldsymbol{\nu} \in \Lambda$ implies $\boldsymbol{\mu} \in \Lambda$ for all $\boldsymbol{\mu} \le \boldsymbol{\nu}$. We shall say that a sequence $(t_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ is monotonically decreasing if it satisfies

$$\boldsymbol{\nu} \le \boldsymbol{\mu} \quad \Rightarrow \quad t_{\boldsymbol{\nu}} \ge t_{\boldsymbol{\mu}} \qquad \forall \boldsymbol{\nu},\ \boldsymbol{\mu} \in \mathcal{F}. \tag{1.2}$$

Additionally we introduce the space $\ell^p_{\mathrm{m}}(\mathcal{F})$ of $\ell^p(\mathcal{F})$ sequences $(t_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \mathbb{C}^{\mathcal{F}}$ which are monotonically decreasing.

For a finite set I, we denote by $|I|$ the cardinality of I. Moreover, if I is at most countable and G some other set, then by G^I we mean the (infinite) product set $\bigtimes_{i\in I} G$.

Throughout, $\mathbb{C}^{\mathbb{N}}$ will be equipped with the product topology and any subset, such as $U := [-1,1]^{\mathbb{N}} \subseteq \mathbb{C}^{\mathbb{N}}$, with the subspace topology. Since we deal with functions $u : U \to X$, for some Banach space X, we also require a measure μ on U. This is defined as the infinite product measure $\mu := \bigotimes_{j\in\mathbb{N}} \lambda/2$ with λ being the Lebesgue measure on $[-1,1]$. Recall that the sigma algebra for μ is generated by all finite rectangles $\prod_{j=1}^{\infty} P_j$ such that only a finite number of the P_j are different from $[-1,1]$ and those that are different are intervals contained in $[-1,1]$. All integrals over U as well as $L^p(U,X)$ norms will be understood with respect to this measure and the Borel sigma algebra on X. Moreover, occasionally we will consider integrals over $[-1,1]^m$, $m \in \mathbb{N}$, in which case, by abuse of notation we also write μ to denote the measure $\bigotimes_{j=1}^m \lambda/2$ on $[-1,1]^m$. The open ball with radius $\rho > 0$ and center $0 \in \mathbb{C}$ is denoted by $B_\rho = \{z \in \mathbb{C} \,:\, |z| < \rho\}$. Additionally, for $m \in \mathbb{N}$ and $\boldsymbol{\rho} = (\rho_j)_{j=1}^m \in (0,\infty)^m$ we set $B_{\boldsymbol{\rho}} := \bigtimes_{j=1}^m B_{\rho_j} \subseteq \mathbb{C}^m$. Similarly, $B_{\boldsymbol{\rho}} := \bigtimes_{j\in\mathbb{N}} B_{\rho_j} \subseteq \mathbb{C}^{\mathbb{N}}$ in case $\boldsymbol{\rho} = (\rho_j)_{j\in\mathbb{N}} \in (0,\infty)^{\mathbb{N}}$. The closure of a subset B of a topological space is denoted by $\operatorname{clos}(B)$.

Finally, for a vector space X over $\mathbb{R}$ or $\mathbb{C}$ we write $X_{\mathbb{C}}$ for its complexification: by this we mean elements in the set $X_{\mathbb{C}} := X + \mathrm{i}X$, with i denoting the imaginary unit in $\mathbb{C}$. If X is a Banach space, the vector space $X_{\mathbb{C}}$ becomes a Banach space being endowed with the norm $\|x_1 + \mathrm{i}x_2\|_{X_{\mathbb{C}}} := \sup_{0\le t\le 2\pi} \|x_1 \cos t - x_2 \sin t\|_X$ (cf. [41]).

2 Multilevel approximation of holomorphic maps

For a Banach space X, consider the approximation of infinite-variate functions $u : U \to X$. We assume u to exhibit certain holomorphic properties as summarized in the next definition. Recall that for $\boldsymbol{\rho} \in (0,\infty)^{\mathbb{N}}$, $B_{\boldsymbol{\rho}} = \times_{j\in\mathbb{N}} B_{\rho_j}$ denotes the cartesian product of the complex open balls $B_{\rho_j} \subseteq \mathbb{C}$ with radius ρ_j around 0.

Definitions analogous to the following one were introduced and employed in [19, 15, 48, 25], also see [52].

Definition 2.1 (($\boldsymbol{b}, \varepsilon, X$)-Holomorphy)**.** *Let X be a Banach space and $u : U \to X$. Assume given a sequence $\boldsymbol{b} = (b_j)_{j\in\mathbb{N}}$ of nonnegative numbers b_j such that $\boldsymbol{b} \in \ell^p(\mathbb{N})$ for some $p \in (0, 1]$, and such that b_j is monotonically decreasing.*

We say that $\boldsymbol{\rho} \in (1,\infty)^{\mathbb{N}}$ is $(\boldsymbol{b}, \varepsilon)$-admissible for some $\varepsilon > 0$ if

$$\sum_{j\in\mathbb{N}} b_j(\rho_j - 1) \le \varepsilon \; . \tag{2.1}$$

With $\overline{B}_{\boldsymbol{\rho}} := \times_{j\in\mathbb{N}} \mathrm{clos}(B_{\rho_j})$ let

$$O_{\boldsymbol{b},\varepsilon} := \bigcup_{\{\boldsymbol{\rho}\,:\,\boldsymbol{\rho} \text{ is } (\boldsymbol{b},\varepsilon)\text{-admissible}\}} \overline{B_{\boldsymbol{\rho}}} \subseteq \mathbb{C}^{\mathbb{N}}. \tag{2.2}$$

The map u is $(\boldsymbol{b}, \varepsilon, X)$-holomorphic (or simply $(\boldsymbol{b}, \varepsilon)$-holomorphic), if for every $(\boldsymbol{b}, \varepsilon)$-admissible $\boldsymbol{\rho}$, u allows a continuous extension $u : B_{\boldsymbol{\rho}} \to X_{\mathbb{C}}$ that is holomorphic as a function of each z_j, and if additionally, there exists a constant $C_{\boldsymbol{b},u} < \infty$ such that

$$\sup_{\boldsymbol{z}\in O_{\boldsymbol{b},\varepsilon}} \|u(\boldsymbol{z})\|_{X_{\mathbb{C}}} \le C_{\boldsymbol{b},u}. \tag{2.3}$$

The implication of $(\boldsymbol{b}, \varepsilon)$-holomorphy of countably-parametric functions on gpc convergence rates is formulated in the following theorem. To state it, we denote the Legendre polynomial of degree n by L_n and assume the normalization $\|L_n\|_{L^\infty(-1,1)} = 1$. For $\boldsymbol{\nu} \in \mathcal{F}$, define the tensorized Legendre polynomial $L_{\boldsymbol{\nu}} := \prod_{j\in\mathbb{N}} L_{\nu_j}$. Note that $(L_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ is an orthogonal basis of $L^2(U, \mathbb{R})$ w.r.t. the measure μ introduced in Section 1 and it holds $\|L_{\boldsymbol{\nu}}\|_{L^2(U)} = \prod_{j\in\mathbb{N}}(2\nu_j + 1)^{-1/2} \le 1$.

Theorem 2.2. *Let $p \in (0, 1]$, let X be a Banach space and let $u : U \to X$ be $(\boldsymbol{b}, \varepsilon, X)$-holomorphic with $\boldsymbol{b} \in \ell^p$. Then u admits the unconditionally in $L^\infty(U, X)$ convergent Legendre gpc expansion $u(\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y})$. The sequence $(\|u_{\boldsymbol{\nu}}\|_X)_{\boldsymbol{\nu}\in\mathcal{F}}$ possesses a majorant in $\ell^p_{\mathrm{m}}(\mathcal{F})$.*

Theorem 2.2 was proven in [15, Theorem 2.2] for $p \in (0, 1)$ under less stringent assumptions than we use here (cp. Remark 2.3). The case $p = 1$ can be shown similarly.

Remark 2.3. *We point out that for simplicity we work here with holomorphy assumed on polydiscs as stated in Definition 2.1. Most results of this paper remain true under the (weaker) assumption that u allows holomorphic extensions to certain polyellipses contained in the polydiscs of Definition 2.1. This will be elaborated in more detail in [51].*

The outline of the remainder of this section is as follows: In Section 2.1 we establish some preliminaries and prove convergence rates w.r.t. the L^q-norm, $q \in [2,\infty]$, for a linear approximation operator. The linearity and boundedness of the operator for $q \in \{2, \infty\}$ will allow us to employ interpolation theory. The accompanying result, Theorem 2.14, is stated under the additional restriction that X is a Hilbert space, since the proof of this theorem will require an orthogonal basis. For this reason, we admit X to be a Banach space throughout, with the sole exception of Theorem 2.14. For the same reason, the proof will employ the Legendre expansion from Theorem 2.19, whereas in the subsequent sections we resort to Taylor expansions. Subsequently, in Section 2.2, we present a constructive interpolation algorithm which maintains the rate of the linear approximant from Theorem 2.14 in the L^∞-norm. Improved convergence rates are shown for the corresponding quadrature algorithm. Finally, in Section 2.3 we discuss the error in terms of the total complexity of our interpolation and quadrature algorithm.

2.1 Linear approximation

We prepare the presentation of the linear approximation results and their proofs by first presenting some results on weighted sequence minimization which are of independent interest. They arise in the derivation of approximation rate bounds and are crucial for our subsequent complexity and convergence rate results. Proofs are provided in the appendix.

2.1.1 Weighted sequence approximation

Let $\mathbf{t}_0 = (t_{0;j})_{j\in\mathbb{N}}$, $\mathbf{t}_1 = (t_{1;j})_{j\in\mathbb{N}}$ be two sequences of nonnegative numbers and let $q, \alpha > 0$. For every $N \in \mathbb{N}$ let $\mathbf{l}_N = (l_{N;j})_{j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$ be a multiindex with $|\mathbf{l}_N| \le N$. We wish to minimize

$$S(\mathbf{t}_0, \mathbf{t}_1, \mathbf{l}_N, \alpha, q) := \sum_{j\in\mathbb{N}} \min\{t_{0;j}, t_{1;j}(l_{N;j}+1)^{-\alpha}\}^q \tag{2.4}$$

or, more precisely, we are interested in the asymptotic behaviour of these sums for $\mathbf{l}_N$ optimal (i.e. minimizing (2.4)), as $N \to \infty$. We provide an answer together with a constructive choice for $l_{N;j}$ in Theorem 2.6 below.

The relation of this problem to sparse-grid approximation of countably-parametric maps u taking values in a Banach space X is as follows. We assume an (unconditionally convergent) expansion of u in a gpc basis (e.g. Taylor or Legendre expansion). To approximate the coefficients in this (gpc type) expansion, we stipulate a numerical method approximating elements of X converging at rate α. For example, we may think of Petrov-Galerkin approximations as in [45]. *The concrete form of the approximation, however, is not essential in the present analysis.* It therefore comprises a broad range of numerical approximations such as collocation or NURBs domain approximations, etc. The coefficient $l_{N;j} \in \mathbb{N}_0$ in (2.4) could then be interpreted *work budget* allowed to approximate the jth largest expansion coefficient, where N stands for a prescribed total amount of work. In this sense, $l_{N;j}$ will also be referred to as the *level* of approximation. The total error of the (multilevel) approximant then boils down to a sum of the type (2.4). For the next result, we admit only discretization levels $l_{N;j}$ in a certain subset $\mathfrak{W}$ of $\mathbb{N}_0$ introduced next.

Assumption 2.4. *The set $\mathfrak{W} = \{\mathfrak{w}_j : j \in \mathbb{N}_0\} \subseteq \mathbb{N}_0$ of work measures consists of the strictly monotonically growing, nonnegative sequence $(\mathfrak{w}_j)_{j\in\mathbb{N}}$, and of $\mathfrak{w}_0 = 0$. There exists a constant $\mathfrak{w}_1 + 1 \le K_{\mathfrak{W}} < \infty$ such that*

$$\forall j \in \mathbb{N}: \qquad \frac{\mathfrak{w}_{j+1}}{\mathfrak{w}_j} \le K_{\mathfrak{W}}. \tag{2.5}$$

Definition 2.5. *For $A \subseteq \mathbb{N}_0$, $0 \in A$, $|A| = \infty$, the operators $\lfloor \cdot \rfloor_A : [0,\infty) \to A$ and $\lceil \cdot \rceil_A : [0,\infty) \to A$ are given by*

$$\lfloor x \rfloor_A := \max\{a \in A \,:\, a \le x\} \qquad \text{and} \qquad \lceil x \rceil_A := \min\{a \in A \,:\, a \ge x\}. \tag{2.6}$$

If applied to sequences, these operators are considered componentwise, and in case $A = \mathbb{N}_0$ we omit the index A.

Theorem 2.6. *Let Assumption 2.4 be fulfilled. Let $\alpha > 0$, $p_0 \in (0,1)$, $p_1 \in [p_0,\infty]$, $q \in (p_0,\infty)$. Let $\mathbf{t}_i = (t_{i;j})_{j\in\mathbb{N}} \in \ell^{p_i}(\mathbb{N})$, $i \in \{0,1\}$ be nonnegative monotonically decreasing sequences. Set*

$$r(p_0,p_1,q,\alpha) := \begin{cases} \alpha q & \text{if } p_1 \le \frac{q}{\alpha q+1}, \\ \beta(\frac{q}{p_0}-1) & \text{if } p_1 > \frac{q}{\alpha q+1}, \end{cases} \quad \text{with} \quad \beta := \frac{\alpha}{\alpha + p_0^{-1} - p_1^{-1}}. \tag{2.7}$$

Set $M = M(N,p_0,p_1,q,\alpha) := \lceil N^\beta \rceil$ and define $w_{N;j}$ via

$$w_{N;j} := \left\lceil N t_{1;j}^{\frac{q}{\alpha q+1}} \left(K_{\mathfrak{W}} \sum_{i=1}^{M} t_{1;i}^{\frac{q}{\alpha q+1}} \right)^{-1} - 1 \right\rceil_{\mathfrak{W}} \tag{2.8}$$

for $j \le M$, respectively $w_{N;j} := 0$ otherwise.

There then exists $0 < C = C(\|\mathbf{t}_0\|_{\ell^{p_0}}, \|\mathbf{t}_1\|_{\ell^{p_1}}, q, \alpha) < \infty$ such that for every $N \in \mathbb{N}$ the multiindex $\mathbf{w}_N \in \mathfrak{W}^{\mathbb{N}}$ satisfies

(i) $|\mathbf{w}_N| \le N$,

(ii) $i \le j$ *implies* $w_{N;i} \ge w_{N;j}$,

(iii) with $S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{w}_N,q,\alpha)$ as in (2.4)

$$S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{w}_N,q,\alpha) \le C N^{-r(p_0,p_1,q,\alpha)}, \tag{2.9}$$

(iv) (2.9) is optimal in the sense that for $\tilde{r} > r$ arbitrary, there exist $\mathbf{t}_0$, $\mathbf{t}_1$ as above, such that $(\mathbf{w}_N)_{N\in\mathbb{N}}$ satisfying (i) and $S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{w}_N,q,\alpha) = O(N^{-\tilde{r}})$ as $N \to \infty$ does not exist.

Before stating the convergence rate bound in Theorem 2.14, we prove an interpolation space result.

2.1.2 Auxiliary results on interpolation of sequence spaces

In the following let $(S,\mathcal{A},\mu)$ be a measure space and let the Banach space X be equipped with the Borel sigma algebra.

Definition 2.7. *For $p \in (0,\infty]$ set $L^p(S,X,\mu) := \{f : S \to X \,:\, f$ strongly measurable, $\|f\|_{L^p(X)} < \infty\}$ where $\|f\|_{L^p(X)} := (\int_S (\|f(x)\|_X)^p \mathrm{d}\mu(x))^{1/p}$, with the usual modification for $p = \infty$ (and identifying equivalence classes of functions agreeing μ-a.e.). We also write $L^p(S,X) := L^p(S,X,\mu)$.*

Definition 2.8. *Let I be a countable index set and let $p \in (0,\infty)$, $\mathbf{r} = (r_i)_{i\in I} \in (0,\infty)^I$. The measure $\mu_{\mathbf{r}}$ on I is given by $\mu_{\mathbf{r}}(i) := r_i^p$ for all $i \in I$. With the measure space $(I, 2^I, \mu_{\mathbf{r}})$, where 2^I denotes the power set of I, we write $\ell^p_{\mathbf{r}}(I,X) := L^p(I,X,\mu_{\mathbf{r}})$. For $\mathbf{r} = (1)_{i\in I}$ we omit the subscript as usual, and in case there is no confusion about X or I, we simply write $\ell^p_{\mathbf{r}}(I)$ or $\ell^p_{\mathbf{r}}$.*

Definition 2.9. *For $1 \le p_0 \le p_1 \le \infty$ and $\theta \in [0,1]$, denote by $[L^{p_0}(S,X), L^{p_1}(S,X)]_\theta$ the complex interpolation space of type θ with respect to $L^{p_0}(S,X)$ and $L^{p_1}(S,X)$ (for the exact definition see, e.g., [37]).*

We recall the well known Marcinkiewicz interpolation theorem.

Theorem 2.10 ([37, Theorem 2.2.6])**.** *Let $1 \le p_0 \le p_1 < \infty$ or $1 \le p_0 < p_1 = \infty$, $\theta \in [0,1]$ and $1/q = \theta/p_0 + (1-\theta)/p_1$. Then there holds*

$$[L^{p_0}(S,X), L^{p_1}(S,X)]_\theta = L^q(S,X) \tag{2.10}$$

isometrically.

Corollary 2.11. *With I, X, $\mathbf{r}$ as in Definition 2.8 and p_0, p_1, q, θ as in Theorem 2.10, it holds*

$$[\ell^{p_0}_{\mathbf{r}}(I,X), \ell^{p_1}_{\mathbf{r}}(I,X)]_\theta = \ell^q_{\mathbf{r}}(I,X) \tag{2.11}$$

isometrically.

Proof. The set of all finitely supported $\mathbf{x}$, i.e.

$$\mathbf{x} = (x_i)_{i\in I} \in X^I, \qquad \text{s.t. } x_i = 0 \text{ for almost every } i \in I \tag{2.12}$$

is dense in $[\ell^{p_0}_{\mathbf{r}}, \ell^{p_1}_{\mathbf{r}}]_\theta$, since $\ell^{p_0}_{\mathbf{r}} \cap \ell^{p_1}_{\mathbf{r}}$ is dense in the interpolation space by [37, Corollary C.2.8], and the finitely supported sequences are dense in $\ell^{p_0}_{\mathbf{r}} \cap \ell^{p_1}_{\mathbf{r}}$. Since the finitely supported sequences are also dense in $\ell^q_{\mathbf{r}}$, in order to prove the corollary, it is sufficient to show that

$$\|\mathbf{x}\|_{\ell^q_{\mathbf{r}}} = \|\mathbf{x}\|_{[\ell^{p_0}_{\mathbf{r}}, \ell^{p_1}_{\mathbf{r}}]_\theta} \tag{2.13}$$

for all $\mathbf{x}$ satisfying (2.12).

For $j \in \{1,2\}$ we consider the two well-defined linear operators

$$A := \begin{cases} \ell^{p_j} \to \ell^{p_j}_{\mathbf{r}} \\ (x_i)_{i\in I} \mapsto (x_i r_i^{-1})_{i\in I} \end{cases} \qquad \text{and} \qquad A^{-1} := \begin{cases} \ell^{p_j}_{\mathbf{r}} \to \ell^{p_j} \\ (x_i)_{i\in I} \mapsto (x_i r_i)_{i\in I}. \end{cases} \tag{2.14}$$

They have norm one and are mutually inverse for $j \in \{1,2\}$. By [37, Theorem C.2.8], for all $\mathbf{x}$ satisfying (2.12),

$$\|A\mathbf{x}\|_{[\ell^{p_0}_{\mathbf{r}}, \ell^{p_1}_{\mathbf{r}}]_\theta} \le \|\mathbf{x}\|_{[\ell^{p_0}, \ell^{p_1}]_\theta} \qquad \text{and} \qquad \|A^{-1}\mathbf{x}\|_{[\ell^{p_0}, \ell^{p_1}]_\theta} \le \|\mathbf{x}\|_{[\ell^{p_0}_{\mathbf{r}}, \ell^{p_1}_{\mathbf{r}}]_\theta}. \tag{2.15}$$

According to Theorem 2.10 we have $\|\mathbf{y}\|_{[\ell^{p_0}, \ell^{p_1}]_\theta} = \|\mathbf{y}\|_{\ell^q}$ for any finitely supported $\mathbf{y} \in X^I$, so that

$$\begin{aligned} \|\mathbf{x}\|_{\ell^q_{\mathbf{r}}} &= \|A^{-1}\mathbf{x}\|_{\ell^q} = \|A^{-1}\mathbf{x}\|_{[\ell^{p_0}, \ell^{p_1}]_\theta} \le \|\mathbf{x}\|_{[\ell^{p_0}_{\mathbf{r}}, \ell^{p_1}_{\mathbf{r}}]_\theta} \\ &= \|AA^{-1}\mathbf{x}\|_{[\ell^{p_0}_{\mathbf{r}}, \ell^{p_1}_{\mathbf{r}}]_\theta} \le \|A^{-1}\mathbf{x}\|_{[\ell^{p_0}, \ell^{p_1}]_\theta} = \|A^{-1}\mathbf{x}\|_{\ell^q} = \|\mathbf{x}\|_{\ell^q_{\mathbf{r}}}, \end{aligned} \tag{2.16}$$

which shows (2.13) and concludes the proof. □

2.1.3 Convergence rate bound

We now state and prove a variant of Theorem 2.6 based on the gpc-index set $\mathcal{F}$. To this end, we require the following auxiliary result on sequence approximation. Its proof is provided in the appendix.

Proposition 2.12. *Let Assumption 2.4 be fulfilled and let* $\mathfrak{W}$ *be as in Assumption 2.4. Let further* $(t_{i;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{p_i}_{\mathrm{m}}(\mathcal{F})$, $i \in \{0,1\}$, $p_0 \in (0,1)$, $p_1 \in [p_0,\infty]$, $q \in [1,2]$, $\alpha > 0$. *There exists a positive constant* $C = C(\|\mathbf{t}_0\|_{\ell^{p_0}}, \|\mathbf{t}_1\|_{\ell^{p_1}}, \alpha)$ *and for every* $N \in \mathbb{N}$, *there exists* $\mathbf{w}_N = (w_{N;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \mathfrak{W}^{\mathcal{F}}$ *such that* $|\mathbf{w}_N| \le N$ *and with the convergence rate* $r = r(p_0,p_1,q,\alpha)$ *as in* (2.7),

$$\sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}\neq 0\}} (w_{N;\boldsymbol{\nu}}+1)^{-\alpha q} t^q_{1;\boldsymbol{\nu}} + \sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}=0\}} t^q_{0;\boldsymbol{\nu}} \le CN^{-r}. \tag{2.17}$$

As is well-known by now, see e.g. [2], multilevel approximations of parametric solution families involve space discretization of instances of the parametric solution, incurring a discretization error. To bound it, we require *"spacial regularity"* of parametric solutions. Our requirement takes the form of the parametric solution belonging to some *"regularity space"* $X^{\mathrm{s}} \subset X$, *with smoothness order "*s*". Concrete examples for choices of* X^{s} *will be presented in Section 4 ahead.*

Assumption 2.13 (Discretization)**.** *The spaces* X *and* X^{s} *are Banach spaces and* X^{s} *is a linear subspace of* X. *There exist* $\alpha > 0$ *and a positive constant* $C(\alpha, X^{\mathrm{s}})$ *such that for every* $N \in \mathbb{N}$ *there exists a space* $X_N \subseteq X$ *of dimension at most* N *and a linear mapping* $\Pi_N : X^{\mathrm{s}} \to X_N$ *such that* $\Pi_0 = 0$ *and for all* $x \in X^{\mathrm{s}}$

$$\|x - \Pi_N x\|_X \le C(\alpha, X^{\mathrm{s}})(N+1)^{-\alpha}\|x\|_{X^{\mathrm{s}}}. \tag{2.18}$$

The next theorem requires X (but not X^{s}) to be a Hilbert space, since we employ the Parseval identity. We also use that best N-term approximation rates are realized by *linear* (quasi-) interpolation operators which facilitates the use of interpolation methods. The complexification $X_{\mathbb{C}} = X + \mathrm{i}X$ of a Hilbert space is obtained through the sesquilinear inner product $\langle a + \mathrm{i}b, c + \mathrm{i}d\rangle := \langle a, b\rangle + \langle b, d\rangle + \mathrm{i}(\langle b, c\rangle - \langle a, d\rangle)$. This is, up to an isomorphism, consistent with the previously introduced complexification for Banach spaces. The statement we now prove is, that the convergence rate of Theorem 2.6 can be achieved subject to a constraint on the total number of degrees of freedom $|\mathbf{w}_N| \le N$ used in the approximant (cp. Assumption 2.13). We point out that the additional spacial regularity of the parametric solution family $\{u(\boldsymbol{y}) : \boldsymbol{y} \in U\} \subset X^{\mathrm{s}}$ will, in general, entail *reduced summability of the* $\|\circ\|_{X^{\mathrm{s}}}$*-norms of the gpc coefficients.* This was observed in previous work [28, 2] to be essential for improved convergence rates when measured in terms of the total number of degrees of freedom. The following theorem accounts for that by providing two summability indices p_0 and p_1 for X resp., for X^{s}-regularity.

Theorem 2.14. *Let Assumptions 2.4 and 2.13 hold, assume that* $p_0 \in (0,1)$, $p_1 \in [p_0,1]$, $q \in [1,2]$ *and let* $u : U \to X^{\mathrm{s}}$ *be* $(\boldsymbol{b}_0, \varepsilon_0, X)$*-holomorphic and* $(\boldsymbol{b}_1, \varepsilon_1, X^{\mathrm{s}})$*-holomorphic, with* $\boldsymbol{b}_0 \in \ell^{p_0}(\mathbb{N})$, $\boldsymbol{b}_1 \in \ell^{p_1}(\mathbb{N})$. *Additionally, assume* X *to be a Hilbert space. Then, there exists* $C < \infty$ *and with* $\mathfrak{W}$ *as in Assumption 2.4 for every* $N \in \mathbb{N}$ *there exists* $\mathbf{w}_N = (w_{N;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \mathfrak{W}^{\mathcal{F}}$ *with* $|\mathbf{w}_N| \le N$ *satisfying* (2.17) *such that with he conjugate* $q^* \in [2,\infty]$ *of* q *(i.e.,* $1/q^* + 1/q = 1$*) and with the rate* $r(p_0,p_1,q,\alpha)$ *as in* (2.7) *there holds*

$$\|u(\boldsymbol{y}) - \sum_{\boldsymbol{\nu}\in\mathcal{F}} L_{\boldsymbol{\nu}}(\boldsymbol{y})\Pi_{w_{N;\boldsymbol{\nu}}} u_{\boldsymbol{\nu}}\|_{L^{q^*}(U,X)} \le C \begin{cases} N^{-r(p_0,p_1,q,\alpha)/q^*} & \text{if } q^* < \infty \\ N^{-r(p_0,p_1,q,\alpha)} & \text{if } q^* = \infty. \end{cases} \tag{2.19}$$

Proof. By Theorem 2.2, $u = \sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}$ with unconditional convergence in $L^\infty(U, X^{\mathrm{s}})$, and with $(\|u_{\boldsymbol{\nu}}\|_X)_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{p_0}_{\mathrm{m}}$, $(\|u_{\boldsymbol{\nu}}\|_{X^{\mathrm{s}}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{p_1}$. Recall that the $L_{\boldsymbol{\nu}}$ are orthogonal in $L^2(U)$ with $\|L_{\boldsymbol{\nu}}\|_{L^2(U)} \le 1$ and $\|L_{\boldsymbol{\nu}}\|_{L^\infty(U)} = 1$ for all $\boldsymbol{\nu} \in \mathcal{F}$. Choose now $\mathbf{w}_N$ as in Proposition 2.12 for $\mathbf{t}_0 = (t_{0;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{p_0}(\mathcal{F})$ a monotonically decreasing majorant of $\|u_{\boldsymbol{\nu}}\|_X$ and $\mathbf{t}_1 = (t_{1;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{p_1}_{\mathrm{m}}(\mathcal{F})$ a monotonically decreasing majorant of $\|u_{\boldsymbol{\nu}}\|_{X^{\mathrm{s}}}$ (these majorants exist according to Theorem 2.2).

Consider the linear operator

$$K : (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \mapsto \sum_{\boldsymbol{\nu}\in\mathcal{F}} (I - \Pi_{w_{N;\boldsymbol{\nu}}}) v_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y}). \tag{2.20}$$

We observe that $K : \ell^2(\mathcal{F}, X) \to L^2(U, X)$ and $K : \ell^1(\mathcal{F}, X) \to L^\infty(U, X)$ are well-defined and bounded, with the sum in (2.20) converging unconditionally in both cases. The unconditional convergence in $L^\infty(U, X)$ follows by the assumed absolute summability of $\|u_{\boldsymbol{\nu}}\|_{X^{\mathrm{s}}} \|L_{\boldsymbol{\nu}}\|_{L^\infty(U)} = \|u_{\boldsymbol{\nu}}\|_{X^{\mathrm{s}}}$ w.r.t. $\boldsymbol{\nu} \in \mathcal{F}$. For $L^2(U, X)$ it follows from the use of the Parseval identity since for any enumeration $(\boldsymbol{\nu}_j)_{j\in\mathbb{N}}$ of $(\boldsymbol{\nu})_{\boldsymbol{\nu}\in\mathcal{F}}$ it holds

$$\|\sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}} - \sum_{j=1}^{N} u_{\boldsymbol{\nu}_j} L_{\boldsymbol{\nu}_j}\|^2_{L^2(U,X)} = \sum_{j>N} \|u_{\boldsymbol{\nu}_j}\|^2_X \|L_{\boldsymbol{\nu}_j}\|^2_{L^2(U)} \le \sum_{j>N} \|u_{\boldsymbol{\nu}_j}\|^2_X, \tag{2.21}$$

which tends to 0 as $N \to \infty$, because $\sum_{\boldsymbol{\nu}\in\mathcal{F}} \|u_{\boldsymbol{\nu}}\|^2_X < \infty$. In turn, the Parseval identity holds because X was assumed to be a Hilbert space and thus so is $L^2(U, X)$.

For $\boldsymbol{y} \in U$, denote in the following $v := \sum_{\boldsymbol{\nu}\in\mathcal{F}} v_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y})$. Under the provision that $(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^1(\mathcal{F}, X)$, we have for some C depending on α

$$\begin{aligned}\|K((v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}})\|_{L^\infty(U,X)} &\le \sum_{\boldsymbol{\nu}\in\mathcal{F}} \|L_{\boldsymbol{\nu}}\|_{L^\infty(U,\mathbb{R})} \|(I - \Pi_{w_{N;\boldsymbol{\nu}}}) v_{\boldsymbol{\nu}}\|_X \\ &\le C \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{\|v_{\boldsymbol{\nu}}\|_{X^{\mathrm{s}}} (w_{N;\boldsymbol{\nu}} + 1)^{-\alpha}, \|v_{\boldsymbol{\nu}}\|_X\}. \end{aligned}\tag{2.22}$$

Since $\|L_{\boldsymbol{\nu}}\|_{L^2(U)} \le 1$, we get for $(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^2(\mathcal{F}, X)$ by the Parseval identity

$$\begin{aligned}\|K((v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}})\|^2_{L^2(U,X)} &\le \sum_{\boldsymbol{\nu}\in\mathcal{F}} \|L_{\boldsymbol{\nu}}\|^2_{L^2(U,\mathbb{R})} \|(I - \Pi_{w_{N;\boldsymbol{\nu}}}) v_{\boldsymbol{\nu}}\|^2_X \\ &\le C \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{\|v_{\boldsymbol{\nu}}\|_{X^{\mathrm{s}}} (w_{N;\boldsymbol{\nu}} + 1)^{-\alpha}, \|v_{\boldsymbol{\nu}}\|_X\}^2. \end{aligned}\tag{2.23}$$

Let $\Lambda_0 := \{\boldsymbol{\nu} \in \mathcal{F} : w_{N;\boldsymbol{\nu}} = 0\}$ and $\Lambda_1 := \mathcal{F}\backslash\Lambda_0$. Set $r_{\boldsymbol{\nu}} := (w_{N;\boldsymbol{\nu}} + 1)^{-\alpha}$ for $\boldsymbol{\nu} \in \Lambda_1$. We define the ($\mathbf{w}_N$-dependent) product space $Y^q := \ell^q_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}) \times \ell^q(\Lambda_0, X)$, where here and in the following, for two Banach spaces B_1, B_2 the space $B_1 \times B_2$ is equipped with the norm $\|(x, y)\|_{B_1\times B_2} := \|x\|_{B_1} + \|y\|_{B_2}$. For $q^* \in (2, \infty)$, we define the number θ by $1/q^* = (1 - \theta)/2$, or equivalently, $1/q = (1 - \theta)/2 + \theta$. We use Theorem 2.10 and Corollary 2.11 to get

$$[L^2(U, X), L^\infty(U, X)]_\theta = L^{q^*}(U, X) \qquad \text{and} \qquad [\ell^2_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}), \ell^1_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}})]_\theta = \ell^q_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}), \tag{2.24}$$

where we employed $[\ell^2_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}), \ell^1_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}})]_\theta = [\ell^1_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}), \ell^2_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}})]_{1-\theta}$. The definition of the complex interpolation spaces (see for example [37, Section C.2]) and Theorem 2.10 applied to

$[\ell^2(\Lambda_0, X), \ell^1(\Lambda_0, X)]_\theta$ give

$$[Y^2, Y^1]_\theta = [\ell^2_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}) \times \ell^2(\Lambda_0, X), \ell^1_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}) \times \ell^1(\Lambda_0, X)]_\theta \\ = [\ell^2_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}), \ell^1_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}})]_\theta \times [\ell^2(\Lambda_0, X), \ell^1(\Lambda_0, X)]_\theta = Y^q, \tag{2.25}$$

with $\|\cdot\|_{[Y^2,Y^1]_\theta} = \|\cdot\|_{Y^q}$. We have shown in (2.22) the existence of a constant $C > 0$ such that

$$\|K((v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}})\|_{L^\infty(U,X)} \le C \left(\|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_1}\|_{\ell^1_{\mathbf{r}}(\Lambda_1,X^{\mathrm{s}})} + \|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_0}\|_{\ell^1(\Lambda_0,X)}\right) = C\|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{Y^1}. \tag{2.26}$$

Due to (2.23) we deduce the bound

$$\|K((v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}})\|_{L^2(U,X)} \le C \left(\|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_1}\|^2_{\ell^2_{\mathbf{r}}(\Lambda_1,X^{\mathrm{s}})} + \|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_0}\|^2_{\ell^2(\Lambda_0,X)}\right)^{1/2} \\ \le C \left(\|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_1}\|_{\ell^2_{\mathbf{r}}(\Lambda_1,X^{\mathrm{s}})} + \|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_0}\|_{\ell^2(\Lambda_0,X)}\right) = C\|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_0}\|_{Y^2}. \tag{2.27}$$

We conclude for $q^* \in [2, \infty]$ (cf., e.g, [37, Theorem C.2.6])

$$\|K((v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}})\|_{L^{q^*}(U,X)} \le C\|(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{Y^q} \qquad \forall\ (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in Y^q = \ell^q_{\mathbf{r}}(\Lambda_1, X^{\mathrm{s}}) \times \ell^q(\Lambda_0, X), \tag{2.28}$$

where C is independent of $\mathbf{w}_N$ (and thus independent of $\mathbf{r}$).

Finally, we observe that for u as in the statement of the theorem, there holds for a.e. $\boldsymbol{y} \in U$ with unconditional limits

$$K((u_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}) = \sum_{\boldsymbol{\nu}\in\mathcal{F}} (u_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y}) - \Pi_{w_{N;\boldsymbol{\nu}}} u_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y})) = u(\boldsymbol{y}) - \sum_{\boldsymbol{\nu}\in\mathcal{F}} \Pi_{w_{N;\boldsymbol{\nu}}} u_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y})\ .$$

By Theorem 2.2 $(u_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^1(\mathcal{F}, X)$. Moreover, $u_{\boldsymbol{\nu}} \in X^{\mathrm{s}}$ for all (finitely many) $\boldsymbol{\nu} \in \Lambda_1$ and thus $(u_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in Y^q$. This allows to conclude with (2.28), and the fact that $\|(u_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}\|^q_{Y^q}$ is bounded by the left-hand side of (2.17) due to our choice of the majorant sequences $\mathbf{t}_0$ and $\mathbf{t}_1$, that (2.19) is satisfied. ☐

2.2 Interpolation and quadrature

Next, we construct multilevel interpolation and quadrature operators. With respect to the $L^\infty(U, X)$ norm, the interpolant achieves the same convergence rate as the best N-term approximant in Theorem 2.14 (in terms of the total work, as is discussed in Section 2.3). Moreover, we employ recent results on high-dimensional Smolyak quadrature from [52], to prove convergence rates for the quadrature operator *better than known best N-term rates*. Indeed, utilizing the fact that certain multivariate monomials are in the kernel of both the integration and quadrature operator (cp. Lemma 2.15 (iii) below) allows to prove the convergence rate $2/p - 1$ for *single level* quadrature applied to $(\boldsymbol{b}, \varepsilon)$-holomorphic functions with $\boldsymbol{b} \in \ell^p$, $p \in (0, 1)$, see [52, Theorem 3.3]. This is an improvement over the previously known best N-term rate $1/p - 1$. The same argument will yield a similar improvement for multilevel quadrature.

Our present construction of the multilevel operators is similar to [22], however with less stringent assumptions. We also refer to [13] for a general discussion of multidimensional interpolation in our context, and to [35] and the references there, where, in particular, multilevel (spline) interpolation was combined with MC integration to approximate a parametric integral. The more recent works [33, 32] propose a multilevel quadrature algorithm for a parametric diffusion problem.

First, we briefly recall general Smolyak interpolants and Smolyak quadrature and also recapitulate a construction of the corresponding multilevel operators in Section 2.2.1. Subsequently, the precise convergence results will be given in Section 2.2.2.

2.2.1 Multilevel Smolyak interpolation/quadrature

The multilevel Smolyak algorithms will be based on sets of univariate abscissae in $[-1,1]$. For every $n \in \mathbb{N}_0$, let $(\chi_{n;j})_{j=0}^n \subset [-1,1]$ denote pairwise distinct points in $[-1,1]$ such that

$$\chi_{0;0} := 0\ . \tag{2.29}$$

The $(\chi_{n;j})_{j=0}^n$ represent the interpolation/quadrature points employed by the univariate interpolation and quadrature operators

$$I_n := \begin{cases} C^0([-1,1]) \to \mathbb{P}_n \\ f \mapsto \sum_{j=0}^n f(\chi_{n;j}) \prod_{\substack{i=0 \\ i\neq j}}^n \frac{y-\chi_{n;i}}{\chi_{n;j}-\chi_{n;i}} \end{cases} \qquad \text{and} \quad Q_n := \begin{cases} C^0([-1,1]) \to \mathbb{R} \\ f \mapsto \sum_{j=0}^n \omega_{n;j} f(\chi_{n;j}), \end{cases} \tag{2.30}$$

where $\mathbb{P}_n = \operatorname{span}\{y^j : j = 0,\dots,n\}$, and $(\omega_{n;j})_{j=0}^n \subset \mathbb{R}$ denote the unique quadrature weights such that $Q_n(f) = \int_{-1}^1 f(y)\mathrm{d}\mu(y)$ for all $f \in \mathbb{P}_n$. In particular $Q_n(f) = \int_{-1}^1 I_n f(y)\mathrm{d}\mu(y)$ for all $f \in C^0([-1,1])$. Additionally, we adhere to the convention $I_{-1} := 0 \in \mathbb{P}_0$, by which we mean the operator mapping continuous functions to the 0-polynomial, and similarly $Q_{-1} := 0 \in \mathbb{R}$. Throughout we assume that there exists $0 < \tau < \infty$ fixed such that the Lebesgue constant $\operatorname{Leb}((\chi_{n;j})_{j=0}^n)$ of $(\chi_{n;j})_{j=0}^n$ satisfies

$$\operatorname{Leb}((\chi_{n;j})_{j=0}^n) \le (n+1)^\tau \qquad \forall\, n \in \mathbb{N}, \tag{2.31}$$

where

$$\operatorname{Leb}((\chi_{n;j})_{j=0}^n) := \sup_{\|f\|_{L^\infty([-1,1])}\le 1} \|I_n(f)\|_{L^\infty([-1,1])}.$$

Next, for all $\boldsymbol{\nu} \in \mathcal{F}$ let $I_{\boldsymbol{\nu}} := \bigotimes_{j\in\mathbb{N}} I_{\nu_j}$, $Q_{\boldsymbol{\nu}} := \bigotimes_{j\in\mathbb{N}} Q_{\nu_j}$. For $\boldsymbol{\nu} \in \mathcal{F}$ we introduce the $\boldsymbol{\nu}$-*increments*

$$\Delta_{\boldsymbol{\nu}}^I := \bigotimes_{j\in\mathbb{N}} (I_{\nu_j} - I_{\nu_j-1}) = \sum_{\{\boldsymbol{e}\in\{0,1\}^{\mathbb{N}}\,:\,\boldsymbol{\nu}-\boldsymbol{e}\in\mathcal{F}\}} (-1)^{|\boldsymbol{e}|} I_{\boldsymbol{\nu}-\boldsymbol{e}}, \tag{2.32a}$$

$$\Delta_{\boldsymbol{\nu}}^Q := \bigotimes_{j\in\mathbb{N}} (Q_{\nu_j} - Q_{\nu_j-1}) = \sum_{\{\boldsymbol{e}\in\{0,1\}^{\mathbb{N}}\,:\,\boldsymbol{\nu}-\boldsymbol{e}\in\mathcal{F}\}} (-1)^{|\boldsymbol{e}|} Q_{\boldsymbol{\nu}-\boldsymbol{e}}, \tag{2.32b}$$

and for $\emptyset \neq \Lambda \subseteq \mathcal{F}$, downward closed with $|\Lambda| < \infty$

$$I_\Lambda := \sum_{\boldsymbol{\nu}\in\Lambda} \Delta_{\boldsymbol{\nu}}^I = \sum_{\boldsymbol{\nu}\in\Lambda} \Bigg(\sum_{\{\boldsymbol{e}\in\{0,1\}^{\mathbb{N}}\,:\,\boldsymbol{\nu}+\boldsymbol{e}\in\Lambda\}} (-1)^{|\boldsymbol{e}|} \Bigg) I_{\boldsymbol{\nu}}, \tag{2.33a}$$

$$Q_\Lambda := \sum_{\boldsymbol{\nu}\in\Lambda} \Delta_{\boldsymbol{\nu}}^I = \sum_{\boldsymbol{\nu}\in\Lambda} \Bigg(\sum_{\{\boldsymbol{e}\in\{0,1\}^{\mathbb{N}}\,:\,\boldsymbol{\nu}+\boldsymbol{e}\in\Lambda\}} (-1)^{|\boldsymbol{e}|} \Bigg) Q_{\boldsymbol{\nu}}. \tag{2.33b}$$

Additionally

$$I_\emptyset := 0,\ Q_\emptyset := 0 \qquad \text{and} \qquad I_{\mathcal{F}} := \mathrm{Id},\ Q_{\mathcal{F}} := \int_U \cdot\ \mathrm{d}\mu(\boldsymbol{y}). \tag{2.34}$$

We note that for every $f \in C^0(U,X)$ and for every index set $\Lambda \subseteq \mathcal{F}$ that is downward closed and finite, there holds

$$Q_\Lambda f = \int_U I_\Lambda f(\boldsymbol{y})\, \mathrm{d}\mu(\boldsymbol{y})\ . \tag{2.35}$$

This follows from the unisolvency of I_Λ for arbitrary downward closed, finite $\Lambda \subseteq \mathcal{F}$, and from the fact that the univariate interpolation and quadrature operators satisfy such an equation. The main properties of the Smolyak type operators I_Λ and Q_Λ are summarized in the next lemma, see [13] and [52, Lemma 3.2].

Lemma 2.15. *Let $\Lambda \subseteq \mathcal{F}$ be downward closed and finite, and let $\mathcal{F}_2$ as in Definition 1.1. Then, the Smolyak operators I_Λ, Q_Λ in* (2.33) *satisfy the following:*

(i) $I_\Lambda P = P$ *and* $Q_\Lambda P = \int_{[-1,1]^{\mathbb{N}}} P(\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{y})$ *for all* $P \in \mathrm{span}\{\boldsymbol{y}^{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \Lambda\}$.

(ii) If additionally (2.31) *holds, then there exists a constant $C > 0$ independent of Λ such that*

$$\|I_\Lambda P\|_{L^\infty(U,X)} + |Q_\Lambda P| \le C |\{\boldsymbol{\eta} : \boldsymbol{\eta} \in \Lambda,\ \boldsymbol{\eta} \le \boldsymbol{\nu}\}|^{\tau+1} \|P\|_{L^\infty(U)} \qquad \forall P \in \mathrm{span}\{\boldsymbol{y}^{\boldsymbol{\mu}} : \boldsymbol{\mu} \le \boldsymbol{\nu}\}.$$

(iii) If additionally $\chi_{0;0} = 0$, *then* $Q_\Lambda P = \int_U P(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} = 0$ *for all* $P \in \mathrm{span}\{\boldsymbol{y}^{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \mathcal{F} \backslash \mathcal{F}_2\}$.

In order to define multilevel approximants of a function $u : U \to X$, within the following assumption we introduce a framework for approximations of u at different levels. It presumes weaker requirements than Assumption 2.13 which was previously used (cf. Remark 2.17). Moreover, in contrast to the discussion of Section 2.1 we do not assume u_l to be a (linear) projection of u onto some subspace. The next assumption includes the use of nonlinear approximation operators (as occur, e.g., for finite element approximations of nonlinear equations).

Assumption 2.16 (Discretization)**.** *The function $u : U \to X$, is $(\boldsymbol{b}_0, \varepsilon, X)$-holomorphic, $\boldsymbol{b}_0 \in \ell^{p_0}$, $p_0 \in (0, 1)$. There exists $C_u < \infty$ and for every $l \in \mathfrak{W}$ with $\mathfrak{W}$ as in Assumption 2.4, there exists a $(\boldsymbol{b}_0, \varepsilon, X)$-holomorphic map $u_l : U \to X$ with the following properties: It holds $u_0 = 0$ and there exists $\boldsymbol{b}_1 \in \ell^{p_1}$, $p_0 \le p_1 < 1$, such that for all $l \in \mathfrak{W}$ (cf.* (2.2)*)*

$$\sup_{\boldsymbol{z} \in O_{\boldsymbol{b}_0,\varepsilon}} \|u_l(\boldsymbol{z})\|_{X_{\mathbb{C}}} \le C_u \qquad \textit{and} \qquad \sup_{\boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}} \|u(\boldsymbol{z}) - u_l(\boldsymbol{z})\|_{X_{\mathbb{C}}} \le C_u l^{-\alpha}. \tag{2.36}$$

Remark 2.17. *Let Assumption 2.13 be satisfied, and let u be $(\boldsymbol{b}_0, \varepsilon, X)$- and $(\boldsymbol{b}_1, \varepsilon, X^{\mathrm{s}})$-holomorphic. One verifies that Assumption 2.16 then holds with $u_l(\boldsymbol{z}) := \Pi_l \Re(u(\boldsymbol{z})) + \mathrm{i}\Pi_l \Im(u(\boldsymbol{z}))$, and C_u proportional to*

$$\sup_{\boldsymbol{z} \in O_{\boldsymbol{b}_0,\varepsilon}} \|u(\boldsymbol{z})\|_{X_{\mathbb{C}}} + \sup_{\boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}} \|u(\boldsymbol{z})\|_{X^{\mathrm{s}}_{\mathbb{C}}} \tag{2.37}$$

in case this quantity is finite. This is the view taken in [22]. We point out however, that such a Π_l is usually not available in practice. For example, if $u_l(\boldsymbol{y})$ is the Galerkin projection of the parametric PDE solution $u(\boldsymbol{y})$ by a stable (uniformly w.r.t. $\boldsymbol{y}$) Petrov-Galerkin Finite Element Method with l degrees of freedom, then typically the coefficients of the PDE, and thus the projector Π_l itself, will depend on the parameter $\boldsymbol{y}$. That is, we obtain parametric Petrov-Galerkin projectors $\Pi_l(\boldsymbol{y})$ with $u_l(\boldsymbol{y}) = \Pi_l(\boldsymbol{y}) u(\boldsymbol{y})$, where $\Pi_l(\boldsymbol{y})$ are stable uniformly w.r.t. l and $\boldsymbol{y}$. This, and the fact that it will allow us to treat nonlinear equations, is why we work with Assumption 2.16.

We are now in position to introduce the multilevel approximants. Let $\mathfrak{W}$ be the set in Assumption 2.4, and let $\mathbf{w} = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}} \in \mathfrak{W}^{\mathcal{F}}$ with $|\mathbf{w}| < \infty$ and the property

$$\boldsymbol{\nu} \le \boldsymbol{\mu} \Rightarrow w_{\boldsymbol{\nu}} \ge w_{\boldsymbol{\mu}}. \tag{2.38}$$

For now $w_{\boldsymbol{\nu}}$ can be interpreted as the amount of work invested in the approximation of the Taylor coefficient belonging to the multiindex $\boldsymbol{\nu}$ (we shall elucidate this in Section 2.3). For such $\mathbf{w}$ and u satisfying Assumption 2.16 define the *sparse interpolation and quadrature operators*

$$I_{\mathbf{w}}(u) := \sum_{\boldsymbol{\nu}\in\mathcal{F}} \Delta^I_{\boldsymbol{\nu}} u_{w_{\boldsymbol{\nu}}} \qquad \text{and} \qquad Q_{\mathbf{w}}(u) := \sum_{\boldsymbol{\nu}\in\mathcal{F}} \Delta^Q_{\boldsymbol{\nu}} u_{w_{\boldsymbol{\nu}}}. \tag{2.39}$$

Set for $l \in \mathbb{N}_0$

$$\Lambda_l = \Lambda_l(\mathbf{w}) := \{\boldsymbol{\nu} \in \mathcal{F} \,:\, w_{\boldsymbol{\nu}} \geq l\}. \tag{2.40}$$

From (2.38) we know that Λ_l is downward closed for all $l \in \mathbb{N}_0$ and the sets are nested, i.e.,

$$\mathcal{F} = \Lambda_0 \supseteq \Lambda_1 \supseteq \Lambda_2 \ldots. \tag{2.41}$$

Then, with (2.33) and using $u_{\mathfrak{w}_0} = u_0 = 0$, we may write

$$I_{\mathbf{w}}(u) = \sum_{j\in\mathbb{N}_0} \sum_{\boldsymbol{\nu}\in\Lambda_{\mathfrak{w}_j}\backslash\Lambda_{\mathfrak{w}_{j+1}}} \Delta^I_{\boldsymbol{\nu}} u_{\mathfrak{w}_j} = \sum_{j\in\mathbb{N}_0} (I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}}) u_{\mathfrak{w}_j} = \sum_{j\in\mathbb{N}} (I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}}) u_{\mathfrak{w}_j}, \tag{2.42a}$$

$$Q_{\mathbf{w}}(u) = \sum_{j\in\mathbb{N}_0} (Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}}) u_{\mathfrak{w}_j} = \sum_{j\in\mathbb{N}} (Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}}) u_{\mathfrak{w}_j}. \tag{2.42b}$$

Note that $|\mathbf{w}| < \infty$ implies $|\Lambda_l| < \infty$ for all $l \in \mathbb{N}$, as well as $\Lambda_l = \emptyset$ for all but finitely many $l \in \mathbb{N}$. Due to our conventions $I_\emptyset := 0$, $Q_\emptyset := 0$, the sums in (2.42) are actually finite.

2.2.2 Convergence

To state the main properties of $(\boldsymbol{b}, \varepsilon)$-holomorphic functions as needed in this section, first the set $\mathfrak{I}$ is introduced. As will become apparent in Section 2.3, our Smolyak interpolation and quadrature operators will only employ tensor operators corresponding to multiindices in a set like $\mathcal{F} \cap \mathfrak{I}^{\mathbb{N}}$, where $\mathfrak{I}$ denotes the set of admissible numbers of interpolation- resp. quadrature points in each coordinate, which is crucial in reducing the overall complexity.

Assumption 2.18. *The set $\mathfrak{I} = \{\mathfrak{i}_j \,:\, j \in \mathbb{N}_0\} \subseteq \mathbb{N}_0$ consists of the strictly monotonically growing, nonnegative sequence $(\mathfrak{i}_j)_{j\in\mathbb{N}_0}$, where $\mathfrak{i}_0 = 0$ and $\mathfrak{i}_1 = 1$. There exists a constant $1 \leq K_{\mathfrak{I}} < \infty$ such that*

$$\forall j \in \mathbb{N}: \qquad \mathfrak{i}_{j+1} - \mathfrak{i}_j \leq K_{\mathfrak{I}}(\mathfrak{i}_j - \mathfrak{i}_{j-1}). \tag{2.43}$$

For an arbitrary, finite index set $\Lambda \subseteq \mathcal{F}$ we introduce its *effective dimension* and its *maximal degree* by

$$d(\Lambda) := \max_{\boldsymbol{\nu}\in\Lambda} |\operatorname{supp} \boldsymbol{\nu}| \qquad \text{and} \qquad m(\Lambda) := \max_{\boldsymbol{\nu}\in\Lambda} \max_{j\in\mathbb{N}} \nu_j. \tag{2.44}$$

For a sequence $(m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ and for $0 < x \in \mathbb{R}$, define $\Lambda((m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}; x) := \{\boldsymbol{\nu} \in \mathcal{F} \,:\, m_{\boldsymbol{\nu}} \geq x\} \subseteq \mathcal{F}$. Monotonic sequences $(m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ satisfying

$$\begin{aligned} d\big(\Lambda((m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}; x)\big) &= o\big(\log(|\Lambda((m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}; x)|)\big) \qquad \text{as } x \to 0, \\ m\big(\Lambda((m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}; x)\big) &= O\big(\log(|\Lambda((m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}; x)|)\big) \qquad \text{as } x \to 0, \end{aligned} \tag{2.45}$$

will be of particular interest to us, as they produce index sets with slowly growing maximal degree and effective dimension (suggesting strong sparsity) which will prove advantageous when considering the computational effort in Section 2.3. The next theorem was shown in [52, Theorem 2.11, Lemma 3.11]. It is the foundation of the approximation results in this section.

Theorem 2.19. *Let $p \in (0,1)$, $0 \le \tau < \infty$ and consider the index set $\mathfrak{I}$ as defined in Assumption 2.18. Let further $u : U \to X$ be $(\boldsymbol{b}, \varepsilon)$-holomorphic, with $\boldsymbol{b} \in \ell^p$. Then*

(i) there exists C independent of u and $C_{\boldsymbol{b},u}$ in Definition 2.1, as well as a sequence $(t_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p_{\mathrm{m}}(\mathcal{F})$ such that for some $u_{\boldsymbol{\nu}} \in X$

$$u(\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}} \boldsymbol{y}^{\boldsymbol{\nu}} \tag{2.46}$$

in the sense of unconditional convergence in $L^\infty(U, X)$, and additionally (cp. (2.3))

$$\|u_{\boldsymbol{\nu}}\|_X \prod_{j\in\mathbb{N}} (\nu_j + 1)^\tau \le C C_{\boldsymbol{b},u} t_{\boldsymbol{\nu}} \qquad \forall\, \boldsymbol{\nu} \in \mathcal{F}. \tag{2.47}$$

(ii) Additionally, for fixed $k \in \mathbb{N}$ and for the index set $\mathcal{F}_k$ as introduced in Definition 1.1, there exists a monotonically decreasing majorant $(m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ of the extension by zero of $(t_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}_k}$ to $\mathcal{F}$ satisfying $(m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{p/k}(\mathcal{F})$. This sequence depends on $\boldsymbol{b}$, ε, τ, but is independent of u, $C_{\boldsymbol{b},u}$. Moreover,

$$m_{\boldsymbol{\nu}} = m_{\boldsymbol{\mu}} \quad \text{if and only if} \quad \lfloor \boldsymbol{\nu} \rfloor_{\mathfrak{I}} = \lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{I}}, \tag{2.48}$$

and $(m_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ satisfies (2.45).

The proof of our convergence rate bounds stated in Theorems 2.21, 2.22 ahead will require the following variant of Theorem 2.6. Its proof is again provided in the appendix.

Proposition 2.20. *Let $\mathfrak{W}$, $\mathfrak{I}$ as in Assumptions 2.4, 2.18. Let $\mathbf{m}_i = (m_{i;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{p_i}_{\mathrm{m}}(\mathcal{F}, \mathbb{R})$, $i \in \{0,1\}$, $p_0 \in (0,1)$, $p_0 \le p_1 \le \infty$, $\alpha > 0$ and assume that $\mathbf{m}_0$, $\mathbf{m}_1$ satisfy (2.45) as well as (2.48).*

Then, there exists a constant $C > 0$ and for every $N \in \mathbb{N}$, there exists $\mathbf{w}_N = (w_{N;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \mathfrak{W}^{\mathcal{F}}$ such that

(i) with $r = \alpha$ if $p_1 \le (\alpha+1)^{-1}$ and $r = \alpha(p_0^{-1} - 1)/(\alpha + p_0^{-1} - p_1^{-1})$ otherwise,

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{m_{0;\boldsymbol{\nu}}, m_{1;\boldsymbol{\nu}}(w_{N;\boldsymbol{\nu}} + 1)^{-\alpha}\} \le C N^{-r}, \tag{2.49}$$

(ii) $|\mathbf{w}_N| \le N$, $\mathbf{w}_N$ satisfies (2.38) and with $\Lambda_{l;N} := \{\boldsymbol{\nu} \in \mathcal{F} : w_{N;\boldsymbol{\nu}} \ge l\}$

$$d\big(\Lambda_{l;N}\big) = o\big(\log(|\Lambda_{l;N}|)\big), \qquad m\big(\Lambda_{l;N}\big) = O\big(\log(|\Lambda_{l;N}|)\big) \qquad \text{as } N \to \infty, \tag{2.50}$$

for all $l \in \mathbb{N}$, where the constants hidden in the Landau notation $o\big(\log(|\Lambda_{l;N}|)\big)$ and $O\big(\log(|\Lambda_{l;N}|)\big)$ do not depend on l. Furthermore, for all l, $N \in \mathbb{N}$

$$\boldsymbol{\nu} \in \Lambda_{l;N} \Leftrightarrow \lfloor \boldsymbol{\nu} \rfloor_{\mathfrak{I}} \in \Lambda_{l;N} \qquad \forall\, \boldsymbol{\nu} \in \mathcal{F}. \tag{2.51}$$

Finally, for every $N \in \mathbb{N}$, $l \in \mathbb{N}_0$, the index set $\Lambda_{l;N}$ is downward closed.

We are now in position to state our convergence rate results. All dimension-independent convergence rates are expressed in terms of the parameter N, which we emphasize does *not* signify the number of collocation points, but it rather is a measure of the work, to be formalized subsequently in Section 2.3. We assume $|\mathbf{w}_N| \le N$.

Theorem 2.21. *Let Assumptions 2.4 and 2.16 and* (2.31) *be satisfied. Then, there exists $C < \infty$ and for every $N \in \mathbb{N}$ there exists $\mathbf{w}_N \in \mathfrak{W}^{\mathcal{F}}$ with $|\mathbf{w}_N| \le N$ satisfying Proposition 2.20 (ii), such that*

$$\|u - I_{\mathbf{w}_N} u\|_{L^\infty(U,X)} \le C N^{-r}, \qquad r = \begin{cases} \alpha & \text{if } p_1 \le \frac{1}{\alpha+1}, \\ \alpha \frac{p_0^{-1}-1}{\alpha + p_0^{-1} - p_1^{-1}} & \text{otherwise.} \end{cases} \tag{2.52}$$

Proof. By Assumption 2.16, for every $l \in \mathfrak{W}$, the discretization error $u - u_l$ is $(\boldsymbol{b}_0, \varepsilon)$-holomorphic with modulus bound $C_{\boldsymbol{b}_0, u-u_l} = C_u + C_{\boldsymbol{b}_0,u}$ and it is $(\boldsymbol{b}_1, \varepsilon)$-holomorphic with modulus bound $C_{\boldsymbol{b}_1, u-u_l} = C_u (l+1)^{-\alpha}$ (cp. (2.3), (2.36)).

We may therefore apply Theorem 2.19 with $k = 1$ and $\tilde{\tau} := \tau + 1$, where $\tau > 0$ is as in (2.31). According to this theorem, for all $l \in \mathfrak{W}$ the $(\boldsymbol{b}_0, \varepsilon)$-holomorphic function $u - u_l$ allows the unconditionally convergent expansion

$$u - u_l = \sum_{\boldsymbol{\nu} \in \mathcal{F}} e_{l,\boldsymbol{\nu}} \boldsymbol{y}^{\boldsymbol{\nu}} \in L^\infty(U, X). \tag{2.53}$$

Furthermore, Theorem 2.19 gives that the Taylor coefficients $e_{l,\boldsymbol{\nu}} \in X$ satisfy for every $\boldsymbol{\nu} \in \mathcal{F}$

$$\begin{aligned} \|e_{l,\boldsymbol{\nu}}\|_X \prod_{j \in \mathbb{N}} (\nu_j + 1)^{\tau+1} &\le C C_{\boldsymbol{b}_0, u-u_l} m_{0;\boldsymbol{\nu}}, \\ \|e_{l,\boldsymbol{\nu}}\|_X \prod_{j \in \mathbb{N}} (\nu_j + 1)^{\tau+1} &\le C C_{\boldsymbol{b}_1, u-u_l} m_{1;\boldsymbol{\nu}} = C_u (l+1)^{-\alpha} m_{1;\boldsymbol{\nu}}, \end{aligned} \tag{2.54}$$

for two sequences $m_{i;\boldsymbol{\nu}} \in \ell^{p_i}_{\mathrm{m}}(\mathcal{F})$, $i \in \{0,1\}$, independent of l. Choose $\mathbf{w}_N$ as in Proposition 2.20. Then, in particular, $|\mathbf{w}_N| \le N$ and $\mathbf{w}_N$ satisfies Proposition 2.20 (ii).

For simplicity we fix N and write $\mathbf{w} := \mathbf{w}_N = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}}$. Observe that $\Lambda_l = \{\boldsymbol{\nu} \in \mathcal{F} \,:\, w_{\boldsymbol{\nu}} \ge l\} = \emptyset$ for all $l \in \mathfrak{W}$ such that $l > \max_{\boldsymbol{\nu} \in \mathcal{F}} w_{\boldsymbol{\nu}}$, so that $I_{\Lambda_l} = 0$ in this case, and consequently $I_{\Lambda_l} = 0$ for all but finitely many $l \in \mathfrak{W}$. Now, using continuity of $(I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}}) : C^0(U,X) \to C^0(U,X)$, the definition of $I_{\mathbf{w}}$ in (2.42) and (2.53) give

$$\begin{aligned} u - I_{\mathbf{w}} u &= \sum_{j \in \mathbb{N}_0} (I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}})(u - u_{\mathfrak{w}_j}) = \sum_{j \in \mathbb{N}_0} (I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}}) \sum_{\boldsymbol{\nu} \in \mathcal{F}} e_{\mathfrak{w}_j, \boldsymbol{\nu}} \boldsymbol{y}^{\boldsymbol{\nu}} \\ &= \sum_{\boldsymbol{\nu} \in \mathcal{F}} \sum_{j \in \mathbb{N}_0} e_{\mathfrak{w}_j, \boldsymbol{\nu}} (I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}}) \boldsymbol{y}^{\boldsymbol{\nu}}, \end{aligned} \tag{2.55}$$

in the sense of unconditional convergence in $L^\infty(U,X)$. Here we have employed $\sum_{j \in \mathbb{N}_0} (I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}}) = I_{\Lambda_0} = I_{\mathcal{F}} = \mathrm{Id}$, and it was allowed to exchange the sums because $\sum_{\boldsymbol{\nu} \in \mathcal{F}} \|e_{l,\boldsymbol{\nu}}\|_X < \infty$ for all $l \in \mathfrak{W}$ according to (2.54), because $(m_{0;\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}} \in \ell^{p_0}(\mathcal{F}) \hookrightarrow \ell^1(\mathcal{F})$ and because the difference $I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}}$ vanishes for all but finitely many j so that the sum in (2.55) is unconditionally convergent. Next we recall that each Λ_l is downward closed according to Proposition 2.20. Therefore, for all $l \in \mathfrak{W}$ (see, e.g., [13]),

$$I_{\Lambda_l} \boldsymbol{y}^{\boldsymbol{\nu}} = I_{\{\boldsymbol{\eta} \in \Lambda_l \,:\, \boldsymbol{\eta} \le \boldsymbol{\nu}\}} \boldsymbol{y}^{\boldsymbol{\nu}}. \tag{2.56}$$

This leads us to introduce for every $\boldsymbol{\nu} \in \mathcal{F}$ the set

$$A_{\boldsymbol{\nu}} := \{ j \in \mathbb{N}_0 \,:\, \{\boldsymbol{\eta} \in \Lambda_{\mathfrak{w}_j} \,:\, \boldsymbol{\eta} \le \boldsymbol{\nu}\} \ne \{\boldsymbol{\eta} \in \Lambda_{\mathfrak{w}_{j+1}} \,:\, \boldsymbol{\eta} \le \boldsymbol{\nu}\} \}. \tag{2.57}$$

Then, it holds $(I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}})\boldsymbol{y}^{\boldsymbol{\nu}} = 0$ whenever $j \notin A_{\boldsymbol{\nu}}$. For each $j \in A_{\boldsymbol{\nu}}$ there must exist some $\boldsymbol{\eta}_j \in \Lambda_{\mathfrak{w}_j}$ with $\boldsymbol{\eta}_j \le \boldsymbol{\nu}$ and $\boldsymbol{\eta}_j \notin \Lambda_{\mathfrak{w}_{j+1}}$. The nestedness (2.41) of the Λ_l implies that $\boldsymbol{\eta}_j \notin \Lambda_{\mathfrak{w}_m}$ for all $m > j$, and therefore $\boldsymbol{\eta}_j \neq \boldsymbol{\eta}_m$ for all $m > j$. Hence the $(\boldsymbol{\eta}_j)_{j\in A_{\boldsymbol{\nu}}}$ are pairwise distinct. Therefore $|A_{\boldsymbol{\nu}}| \le |\{\boldsymbol{\eta} : \boldsymbol{\eta} \le \boldsymbol{\nu}\}| = \prod_{\{i : \nu_i \neq 0\}}(\nu_i + 1)$. Additionally, since $\boldsymbol{\nu} \in \Lambda_{w_{\boldsymbol{\nu}}} \subseteq \Lambda_{w_{\boldsymbol{\nu}}-1} \ldots$, and thus $\{\boldsymbol{\eta} \in \Lambda_{\mathfrak{w}_j} : \boldsymbol{\eta} \le \boldsymbol{\nu}\} = \{\boldsymbol{\eta} : \boldsymbol{\eta} \le \boldsymbol{\nu}\}$ for all $j \in \mathbb{N}_0$ such that $\mathfrak{w}_j \le w_{\boldsymbol{\nu}}$, we have $\min_{j\in A_{\boldsymbol{\nu}}} \mathfrak{w}_j \ge w_{\boldsymbol{\nu}}$. Employing Lemma 2.15, items (i) and (ii) we arrive at

$$
\begin{aligned}
\|u - I_{\mathbf{w}}u\|_{L^\infty(U,X)} &\le \sum_{\boldsymbol{\nu}\in\mathcal{F}} \sum_{j\in A_{\boldsymbol{\nu}}} \|e_{\mathfrak{w}_j,\boldsymbol{\nu}}\|_X \|(I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}})\boldsymbol{y}^{\boldsymbol{\nu}}\|_{L^\infty(U,\mathbb{R})} \\
&\le \sum_{\boldsymbol{\nu}\in\mathcal{F}} \sup_{\{l\in\mathfrak{W} : l\ge w_{\boldsymbol{\nu}}\}} \|e_{l,\boldsymbol{\nu}}\|_X \sum_{j\in A_{\boldsymbol{\nu}}} (\|I_{\Lambda_{\mathfrak{w}_j}}\boldsymbol{y}^{\boldsymbol{\nu}}\| + \|I_{\Lambda_{\mathfrak{w}_{j+1}}}\boldsymbol{y}^{\boldsymbol{\nu}}\|) \\
&\le \sum_{\boldsymbol{\nu}\in\mathcal{F}} \sup_{\{l\in\mathfrak{W} : l\ge w_{\boldsymbol{\nu}}\}} \|e_{l,\boldsymbol{\nu}}\|_X \sum_{\{\boldsymbol{\eta} : \boldsymbol{\eta}\le\boldsymbol{\nu}\}} 2|\{\boldsymbol{\gamma} : \boldsymbol{\gamma} \le \boldsymbol{\nu}\}|^{\tau+1} \\
&= 2\sum_{\boldsymbol{\nu}\in\mathcal{F}} \sup_{\{l\in\mathfrak{W} : l\ge w_{\boldsymbol{\nu}}\}} \|e_{l,\boldsymbol{\nu}}\|_X \prod_{j\in\mathbb{N}}(\nu_j + 1)^{\tau+1} \\
&\le C(C_{\boldsymbol{b}_0,u} + C_u) \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{m_{0;\boldsymbol{\nu}}, (w_{\boldsymbol{\nu}} + 1)^{-\alpha} m_{1;\boldsymbol{\nu}}\},
\end{aligned}
\tag{2.58}
$$

where we have used (2.54) to estimate

$$
\sup_{\{l\in\mathfrak{W} : l\ge w_{\boldsymbol{\nu}}\}} \|e_{l,\boldsymbol{\nu}}\|_X \le \min\{m_{0;\boldsymbol{\nu}}, (w_{\boldsymbol{\nu}} + 1)^{-\alpha} m_{1;\boldsymbol{\nu}}\}. \tag{2.59}
$$

We conclude the proof by referring to (2.49). □

Theorem 2.22. *Let Assumptions 2.4 and 2.16 and* (2.29), (2.31) *be satisfied. Then, there exists* $C < \infty$ *and for every* $N \in \mathbb{N}$ *there exists* $\mathbf{w}_N \in \mathfrak{W}^{\mathcal{F}}$ *with* $|\mathbf{w}_N| \le N$ *satisfying Proposition 2.20 (ii), such that*

$$
\left\| \int_U u(\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y}) - Q_{\mathbf{w}_N}u \right\|_X \le CN^{-r}, \qquad r = \begin{cases} \alpha & \text{if } \frac{p_1}{2} \le \frac{1}{\alpha+1} \\ \alpha\frac{2p_0^{-1}-1}{\alpha+2p_0^{-1}-2p_1^{-1}} & \text{otherwise.} \end{cases} \tag{2.60}
$$

Proof. By Assumption 2.16, for every $l \in \mathfrak{W}$ the discretization error $u - u_l$ is $(\boldsymbol{b}_0, \varepsilon)$-holomorphic with uniform (with respect to the discretization parameter l) bound on the modulus $C_{\boldsymbol{b}_0,u-u_l} = C_u + C_{\boldsymbol{b}_0,u}$. Moreover, it is $(\boldsymbol{b}_1, \varepsilon)$-holomorphic with $C_{\boldsymbol{b}_1,u-u_l} = C_u(l + 1)^{-\alpha}$ (cp. (2.3), (2.36)). We now apply Theorem 2.19 with $k = 2$, and with $\tilde{\tau} := \tau + 1$, where $\tau > 0$ is as in (2.31). The rest of the proof parallels to the argument used in the proof of Theorem 2.21.

By Theorem 2.19, for every $l \in \mathfrak{W}$ we have the absolutely convergent expansion $u - u_l = \sum_{\boldsymbol{\nu}\in\mathcal{F}} e_{l,\boldsymbol{\nu}}\boldsymbol{y}^{\boldsymbol{\nu}} \in L^\infty(U, X)$, and similar as in (2.54) it holds for every $\boldsymbol{\nu} \in \mathcal{F}_2$ (because we applied Theorem 2.19 with $k = 2$)

$$
\|e_{l,\boldsymbol{\nu}}\|_X \prod_{j\in\mathbb{N}}(\nu_j + 1)^{\tau+1} \le CC_{\boldsymbol{b}_0,u-u_l} m_{0;\boldsymbol{\nu}}, \qquad \|e_{l,\boldsymbol{\nu}}\|_X \prod_{j\in\mathbb{N}}(\nu_j + 1)^{\tau+1} \le C_u(l + 1)^{-\alpha} m_{1;\boldsymbol{\nu}} \tag{2.61}
$$

for two sequences $(m_{i;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell_{\mathrm{m}}^{p_i/2}(\mathcal{F})$, $i \in \{0, 1\}$, independent of l. Let again $\mathbf{w}_N$ be as in Proposition 2.20 for these two sequences, fix N and write $\mathbf{w} := \mathbf{w}_N = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ in the following.

As in the proof of Theorem 2.21 it holds $Q_{\Lambda_l} = 0$ for all $l \in \mathfrak{W}$ such that $l > \max_{\boldsymbol{\nu}\in\mathcal{F}} w_{\boldsymbol{\nu}}$. Using continuity of $(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}}) : C^0(U,X) \to \mathbb{R}$, the definition of $Q_{\mathbf{w}}$ in (2.42) gives as in (2.55) in the sense of unconditional convergence

$$\int_U u(\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y}) - Q_{\mathbf{w}}u = \sum_{j\in\mathbb{N}_0}(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}})(u - u_{\mathfrak{w}_j}) = \sum_{\boldsymbol{\nu}\in\mathcal{F}}\sum_{j\in\mathbb{N}_0} e_{\mathfrak{w}_j,\boldsymbol{\nu}}(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}})\boldsymbol{y}^{\boldsymbol{\nu}} \quad (2.62)$$

where we have employed $\sum_{j\in\mathbb{N}_0}(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}}) = Q_{\Lambda_0} = Q_{\mathcal{F}} = \int_U \cdot \,\mathrm{d}\mu(\boldsymbol{y})$. Using Lemma 2.15 (iii) we conclude

$$\int_U u(\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y}) - Q_{\mathbf{w}}u = \sum_{\boldsymbol{\nu}\in\mathcal{F}_2}\sum_{j\in\mathbb{N}_0} e_{\mathfrak{w}_j,\boldsymbol{\nu}}(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}})\boldsymbol{y}^{\boldsymbol{\nu}}. \quad (2.63)$$

With

$$A_{\boldsymbol{\nu}} := \{j \in \mathbb{N}_0 \,:\, \{\boldsymbol{\eta} \in \Lambda_{\mathfrak{w}_j} \,:\, \boldsymbol{\eta} \le \boldsymbol{\nu}\} \neq \{\boldsymbol{\eta} \in \Lambda_{\mathfrak{w}_{j+1}} \,:\, \boldsymbol{\eta} \le \boldsymbol{\nu}\}\} \quad (2.64)$$

as in (2.57) we have with the there stated arguments that $(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}})\boldsymbol{y}^{\boldsymbol{\nu}} = 0$ whenever $j \notin A_{\boldsymbol{\nu}}$, and $\min_{j\in A_{\boldsymbol{\nu}}} \mathfrak{w}_j \ge w_{\boldsymbol{\nu}}$. Employing (2.63) and Lemma 2.15, items (i) and (ii), we arrive at

$$\begin{aligned}
\left\|\int_U u(\boldsymbol{y})\mathrm{d}\boldsymbol{y} - Q_{\mathbf{w}}u\right\|_X &\le \sum_{\boldsymbol{\nu}\in\mathcal{F}_2}\sum_{j\in A_{\boldsymbol{\nu}}} \|e_{\mathfrak{w}_j,\boldsymbol{\nu}}\|_X |(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}})\boldsymbol{y}^{\boldsymbol{\nu}}| \\
&\le \sum_{\boldsymbol{\nu}\in\mathcal{F}_2} \sup_{\{l\in\mathfrak{W}\,:\,l\ge w_{\boldsymbol{\nu}}\}} \|e_{l,\boldsymbol{\nu}}\|_X \sum_{j\in A_{\boldsymbol{\nu}}}(|Q_{\Lambda_{\mathfrak{w}_j}}\boldsymbol{y}^{\boldsymbol{\nu}}| + |Q_{\Lambda_{\mathfrak{w}_{j+1}}}\boldsymbol{y}^{\boldsymbol{\nu}}|) \\
&\le \sum_{\boldsymbol{\nu}\in\mathcal{F}_2} \sup_{\{l\in\mathfrak{W}\,:\,l\ge w_{\boldsymbol{\nu}}\}} \|e_{l,\boldsymbol{\nu}}\|_X \sum_{\{\boldsymbol{\eta}\,:\,\boldsymbol{\eta}\le\boldsymbol{\nu}\}} 2|\{\boldsymbol{\gamma} \,:\, \boldsymbol{\gamma} \le \boldsymbol{\nu}\}|^{\tau+1} \\
&= 2\sum_{\boldsymbol{\nu}\in\mathcal{F}_2} \sup_{\{l\in\mathfrak{W}\,:\,l\ge w_{\boldsymbol{\nu}}\}} \|e_{l,\boldsymbol{\nu}}\|_X \prod_{j\in\mathbb{N}}(\nu_j+1)^{\tau+1} \le C(C_{\boldsymbol{b}_0,u} + C_u)\sum_{\boldsymbol{\nu}\in\mathcal{F}_2}\min\{m_{0;\boldsymbol{\nu}}, (w_{\boldsymbol{\nu}}+1)^{-\alpha}m_{1;\boldsymbol{\nu}}\},
\end{aligned} \quad (2.65)$$

where we have used an estimate analogous to (2.59) for $\boldsymbol{\nu} \in \mathcal{F}_2$. We conclude the proof with pointing out that $\mathbf{m}_0 \in \ell^{p_0/2}(\mathcal{F})$, $\mathbf{m}_1 \in \ell^{p_1/2}(\mathcal{F})$ satisfy (2.49). $\square$

Remark 2.23. *Theorem 2.22 remains true if instead of* (2.31)*, the operators* Q_n *in* (2.30) *satisfy* $\|Q_n\| = \sup_{\|f\|_{C^0([-1,1])}\le 1}|Q_n(f)| \le (n+1)^\tau$ *for all* $n \in \mathbb{N}_0$*. In this case, the proof of the theorem proceeds verbatim. For example, if* Q_n *is the Gauss-Legendre quadrature on* $[-1,1]$*, then the weights* $\omega_{n;j}$ *in* (2.30) *are all positive, and one easily finds* $\|Q_n\| = 1$ *as is well-known (remember that we use* $1/2$ *times the Lebesgue measure on* $[-1,1]$*).*

2.3 Error vs. work analysis

Theorems 2.21 and 2.22 give convergence rates in terms of $N \ge |\mathbf{w}_N|$. In this section it is shown that, assuming one evaluation of u_l in Assumption 2.16 to be of cost $O(l)$, N is essentially an estimate of the overall complexity of evaluating the interpolant respectively computing the quadrature value. This leads to a convergence result in terms of the total work (with basically the same rates as in the mentioned theorems). The precise statement is formulated in Theorems 2.31, 2.32. To analyze the work of computing $I_{\mathbf{w}}$, $Q_{\mathbf{w}}$, with $\mathbf{w} = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ and $\mathbf{w} \in \mathfrak{W}^{\mathcal{F}}$ where $\mathfrak{W}$ is as in Assumption 2.4, we shall use the representations in (2.42). That is, with $\Lambda_l = \{\boldsymbol{\nu} \in \mathcal{F} : w_{\boldsymbol{\nu}} \geq l\}$

$$I_{\mathbf{w}}(u) = \sum_{j\in\mathbb{N}} \left(I_{\Lambda_{\mathfrak{w}_j}} - I_{\Lambda_{\mathfrak{w}_{j+1}}} \right) u_{\mathfrak{w}_j}, \qquad Q_{\mathbf{w}}(u) = \sum_{j\in\mathbb{N}} \left(Q_{\Lambda_{\mathfrak{w}_j}} - Q_{\Lambda_{\mathfrak{w}_{j+1}}} \right) u_{\mathfrak{w}_j}. \tag{2.66}$$

Evaluating $I_{\mathbf{w}}u$ at some $\boldsymbol{y} \in U$ thus involves

(i) computing $u_{\mathfrak{w}_j}(\boldsymbol{\chi})$ at all interpolation points $\boldsymbol{\chi} \in U$ required by $I_{\Lambda_{\mathfrak{w}_j}}$, $I_{\Lambda_{\mathfrak{w}_{j+1}}}$ for all $j \in \mathbb{N}$,

(ii) evaluating $I_{\Lambda_{\mathfrak{w}_j}}$, $I_{\Lambda_{\mathfrak{w}_{j+1}}}$ at $\boldsymbol{y}$, given the values of $u_{\mathfrak{w}_j}$, $u_{\mathfrak{w}_{j+1}}$ from (i), for all $j \in \mathbb{N}$.

In case of $Q_{\mathbf{w}}$, the computations can be structured analogously. We discuss these two items separately in Section 2.3.1, Section 2.3.2, before considering the error w.r.t. the overall complexity in Section 2.3.3. Furthermore, we will impose the following condition on the sets $\mathfrak{I}$, $\mathfrak{W}$, which is satisfied in case the elements of the set in which we choose the components of our multiindices ($\mathfrak{I}$) or work levels ($\mathfrak{W}$), grow exponentially.

Assumption 2.24. *Let Assumptions 2.4 and 2.18 be satisfied. For any $\gamma \geq 1$, there exist constants $K_{\mathfrak{W}}$, $K_{\mathfrak{I}}$, which in addition to the requirements of Assumptions 2.4, 2.18 satisfy*

$$\sum_{j=1}^{m} \mathfrak{w}_j^{\gamma} \leq K_{\mathfrak{W}} \mathfrak{w}_m^{\gamma} \qquad \textit{and} \qquad \sum_{j=1}^{m} (\mathfrak{i}_j + 1) \leq K_{\mathfrak{I}}(\mathfrak{i}_m + 1). \tag{2.67}$$

The following discussion concentrates on the situation of Proposition 2.20 and Theorems 2.21, 2.22.

2.3.1 Evaluation of function values

Consider a finite downward closed set $\Lambda \subseteq \mathcal{F}$. With (2.33) we have

$$I_{\Lambda}u = \sum_{\{\boldsymbol{\nu}\in\Lambda\,:\,c_{\Lambda;\boldsymbol{\nu}}\neq 0\}} c_{\Lambda;\boldsymbol{\nu}} I_{\boldsymbol{\nu}} u, \quad \text{where} \quad c_{\Lambda;\boldsymbol{\nu}} := \sum_{\{\boldsymbol{e}\in\{0,1\}^{\mathbb{N}}\,:\,\boldsymbol{\nu}+\boldsymbol{e}\in\Lambda\}} (-1)^{|\boldsymbol{e}|}. \tag{2.68}$$

An analogous representation holds for Q_{Λ}. In order to evaluate $I_{\Lambda}u$ or $Q_{\Lambda}u$ for some $u : U \to X$, we need to evaluate the tensor interpolants/quadrature operators $I_{\boldsymbol{\nu}}$ and $Q_{\boldsymbol{\nu}}$ introduced in Section 2.2.1. For these operators, $c_{\Lambda;\boldsymbol{\nu}} \neq 0$ in (2.68). Recalling the construction of those tensor operators, we therefore require the function value of u at each point in the set

$$\mathrm{pts}(\Lambda) := \bigcup_{\{\boldsymbol{\nu}\in\Lambda\,:\,c_{\Lambda;\boldsymbol{\nu}}\neq 0\}} \{(\chi_{\nu_1;j_1}, \chi_{\nu_2,j_2}, \dots) \in U \,:\, 0 \leq j_i \leq \nu_i \; \forall\, i \in \mathbb{N}\}. \tag{2.69}$$

Remark 2.25. *Suppose that the $\chi_{n;j}$ are nested, in the sense that there exists a sequence $(\chi_j)_{j\in\mathbb{N}}$ such that $\chi_{n;j} = \chi_j$ for all $n \in \mathbb{N}$ and all $0 \leq j \leq n$. Then, the number of points in the set (2.69) is bounded by the number of multiindices in Λ, since each multiindices $\boldsymbol{\nu} \in \Lambda$ corresponds to one point $(\chi_{\nu_j})_{j\in\mathbb{N}} \in U$. Thus $|\mathrm{pts}(\Lambda)| \leq |\Lambda|$. In the general case, where the point sets are not nested, this is not necessarily true.*

The number of points in (2.69) can be estimated by

$$\sum_{\{\boldsymbol{\nu}\in\Lambda\,:\,c_{\Lambda;\boldsymbol{\nu}}\neq 0\}}\prod_{j\in\mathbb{N}}(\nu_j+1). \tag{2.70}$$

To control this quantity, we now make use of our assumptions on the set $\mathfrak{I}$.

Under Assumption 2.24, suppose that Λ has the property (cp. (2.6))

$$\boldsymbol{\nu}\in\Lambda \Longrightarrow \lceil\boldsymbol{\nu}\rceil_{\mathfrak{I}}\in\Lambda. \tag{2.71}$$

For $\Lambda\subseteq\mathcal{F}$, set

$$\Lambda|_{\mathfrak{I}} := \{\boldsymbol{\nu}\in\Lambda\;:\;\nu_j\in\mathfrak{I}\;\forall\, j\in\mathbb{N}\}. \tag{2.72}$$

Lemma 2.26 ([52, Lemma 3.8])**.** *Let $\Lambda\subset\mathcal{F}$ be downward closed with the property* (2.71)*. Then for all $\boldsymbol{\nu}\in\Lambda\backslash\Lambda|_{\mathfrak{I}}$*

$$c_{\Lambda;\boldsymbol{\nu}} := \sum_{\{\boldsymbol{e}\in\{0,1\}^{\mathbb{N}}\,:\,\boldsymbol{\nu}+\boldsymbol{e}\in\Lambda\}}(-1)^{|\boldsymbol{e}|}=0. \tag{2.73}$$

Lemma 2.27 ([52, Lemma 3.9])**.** *Let $\mathfrak{I}$ be as in Assumption 2.24. Let $\Lambda\subseteq\mathcal{F}$ be downward closed and $|\Lambda|<\infty$. Then, with $d(\Lambda)$, $\Lambda|_{\mathfrak{I}}$ as defined in* (2.44), (2.72)*, there holds*

$$\sum_{\boldsymbol{\nu}\in\Lambda|_{\mathfrak{I}}}\prod_{j\in\mathbb{N}}(\nu_j+1)\le K_{\mathfrak{I}}^{d(\Lambda)}|\Lambda|. \tag{2.74}$$

Under the assumptions of Proposition 2.20, the asymptotic behaviour of $N\mapsto d(\Lambda_{N;l})$ as $N\to\infty$ is known. Its growth is, in fact, very slow (cp. (2.50)). Consequently, the number of points grows almost linearly in the number of multiindices.

Lemma 2.28. *Let $\Lambda_{l;N}$ as in Proposition 2.20, l, $N\in\mathbb{N}$ and let $\delta>0$. Then there exists a constant $C>0$ (depending on δ) such that for all l, $N\in\mathbb{N}$ it holds with* (2.69) *that* $|\mathrm{pts}(\Lambda_{l;N})|\le C|\Lambda_{l;N}|^{1+\delta}$.

Proof. From (2.51) one easily deduces with $\tilde{\mathfrak{I}} := \{0\}\cup\{a-1\;:\;0<a\in\mathfrak{I}\}$ that

$$\boldsymbol{\nu}\in\Lambda_{l;N}\Rightarrow\lceil\boldsymbol{\nu}\rceil_{\tilde{\mathfrak{I}}}\in\Lambda_{l;N}, \tag{2.75}$$

i.e. (2.71) holds with $\tilde{\mathfrak{I}}$. Hence Lemmata 2.26 and 2.27 allow to bound (2.70) by $C|\Lambda_{l;N}|^{1+\delta}$ due to $d(\Lambda_{l;N})=o(\log(|\Lambda_{l;N}|))$ as $|\Lambda_{l;N}|\to\infty$, which is (2.50). □

2.3.2 Cost of evaluation of I_Λ, Q_Λ

Let again $\Lambda\subseteq\mathcal{F}$ be downward closed and finite. To analyze the cost of evaluating $I_\Lambda u: U\to X$ at some $\boldsymbol{y}\in U$ respectively computing $Q_\Lambda u\in X$, we proceed as in [52] and assume that the values of u at the interpolation/quadrature points are already known. We restrict ourselves to the evaluation of $I_\Lambda u$ for $\boldsymbol{y}\in U$ for the moment. Recalling the definition of $m(\Lambda)$, $d(\Lambda)$, we can consider the following quantity introduced in [52]

$$\mathrm{cost}(I_\Lambda) := \underbrace{m(\Lambda)^3}_{\text{precomp. of interp. coeffs.}} + \underbrace{2^{d(\Lambda)}|\Lambda|}_{\text{comp. of }(c_{\Lambda;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}} + \sum_{\{\boldsymbol{\nu}\in\Lambda\,:\,c_{\Lambda;\boldsymbol{\nu}}\neq 0\}}\underbrace{\prod_{j\in\mathbb{N}}(\nu_j+1)}_{\text{evaluation of }I_{\boldsymbol{\nu}}}, \tag{2.76a}$$

to be a an upper bound for the cost of computing $I_\Lambda u(\boldsymbol{y})$ in (2.68): computation of $(c_{\Lambda;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ requires a computational cost bounded by $2^{d(\Lambda)}|\Lambda|$, see [52, Lemma 3.6]. Next consider the computational cost of evaluating the tensor interpolant $I_{\boldsymbol{\nu}} = \bigotimes_{j\in\mathrm{supp}\,\boldsymbol{\nu}} I_{\nu_j}$. The univariate interpolant I_n with $n+1$ distinct points $\{\chi_{n;j}\}_{j=0}^n \subset [-1,1]$, can be evaluated at $y \in [-1,1]$ with complexity $O(n)$, assuming certain coefficients (not depending on y) have been precomputed with complexity $O(n^2)$. One such algorithm is the so-called *barycentric interpolation formula* (see, e.g., [4]). The computation of said coefficients amounts to $\sum_{j=1}^{m(\Lambda)} j^2 = O(m(\Lambda)^3)$ operations. Once completed, each $I_{\boldsymbol{\nu}}$ in (2.68) adds $\prod_j(\nu_j+1)$ to the total complexity of evaluating (2.68) at some $\boldsymbol{y} \in [-1,1]^{\mathbb{N}}$. Finally, the summation over all $\boldsymbol{\nu} \in \Lambda$ in (2.68) can be absorbed in the last term of (2.76a). In case of Q_Λ the only difference is that instead of the interpolation coefficients, we need to compute the quadrature weights in (2.30), which requires the solution of a linear system of dimension j for each $j = 1,\dots,m(\Lambda)$. This adds $\sum_{j=1}^{m(\Lambda)} j^3 = O(m(\Lambda)^4)$ to the total complexity. We therefore define the *quadrature cost* associated with a downward closed set Λ as

$$\mathrm{cost}(Q_\Lambda) := \mathrm{cost}(I_\Lambda) - m(\Lambda)^3 + m(\Lambda)^4. \tag{2.76b}$$

Lemma 2.29. *Let $\Lambda_{l;N}$ as in Proposition 2.20, l, $N \in \mathbb{N}$ and let $\delta > 0$. Then there exists a constant $C > 0$ (depending on δ) such that for all l, $N \in \mathbb{N}$ it holds with* (2.76)

$$\mathrm{cost}(I_{\Lambda_{l;N}}) \le C|\Lambda_{l;N}|^{1+\delta} \qquad \text{and} \qquad \mathrm{cost}(Q_{\Lambda_{l;N}}) \le C|\Lambda_{l;N}|^{1+\delta}. \tag{2.77}$$

Proof. First consider $I_{\Lambda_{l;N}}$. As in the proof of Lemma 2.28, equation (2.71) holds with $\tilde{\mathfrak{I}} = \{0\} \cup \{a-1 \,:\, 0 < a \in \mathfrak{I}\}$. Hence Lemmata 2.26 and 2.27 allow to bound the last term in (2.76a) by $C|\Lambda_{l;N}|^{1+\delta}$ due to $d(\Lambda_{l;N}) = o(\log(|\Lambda_{l;N}|))$ as $|\Lambda_{l;N}| \to \infty$, which is true by (2.50). The first two terms in (2.76a) are bounded, additionally employing $m(\Lambda_{l;N}) = O(\log(|\Lambda_{l;N}|))$ as $|\Lambda_{l;N}| \to \infty$. For the cost of the Smolyak quadrature $Q_{\Lambda_{l;N}}$, by (2.76b) instead of $m(\Lambda_{l;N})^3$ one has the term $m(\Lambda_{l;N})^4$. The claim is clearly fulfilled also in this case. □

2.3.3 Complexity Bound

We are now in position to obtain bounds on the algorithmic complexity, i.e. of error vs. work. To this end, we adopt the following work model.

Assumption 2.30. *Let u satisfy Assumption 2.13, let $\Lambda \subseteq \mathcal{F}$ be finite and downward closed and let $\gamma \ge 1$.*

(i) For $\boldsymbol{\chi} \in U$ arbitrary and every $l \in \mathbb{N}$, the function u_l in Assumption 2.16 can be evaluated at $\boldsymbol{\chi}$ with computational work proportional to l^γ.

(ii) The work load to evaluate $I_\Lambda u$, $Q_\Lambda u$ given the function values at the interpolation/quadrature points, can be bounded up to a constant by the right-hand sides of (2.76)*.*

Theorem 2.31. *Let* (2.31) *and Assumption 2.30 for some $\gamma \ge 1$ be satisfied. Let $u : U \to X$ satisfy Assumption 2.16. Let $\delta > 0$, $\boldsymbol{y} \in U$ and let $\mathfrak{W}$ be as in Assumption 2.24. Then there exists a constant $C = C(\delta) > 0$ and for every $N \in \mathbb{N}$ there exists $\mathbf{w}_N \in \mathfrak{W}^{\mathbb{N}}$ such that the total work to compute $I_{\mathbf{w}_N}(\boldsymbol{y})$ is bounded by CN, and such that*

$$\|u - I_{\mathbf{w}_N} u\|_{L^\infty(U,X)} \le CN^{-\frac{r}{\gamma(1+\delta)}} \qquad r = \begin{cases} \alpha & \text{if } p_1 \le \frac{1}{\alpha+1}, \\ \alpha\frac{p_0^{-1}-1}{\alpha+p_0^{-1}-p_1^{-1}} & \text{otherwise.} \end{cases} \tag{2.78}$$

Proof. Choose $\mathfrak{I}$ satisfying Assumption 2.24 (e.g. $\mathfrak{I} = \{0\} \cup \{2^j : j \in \mathbb{N}_0\}$). Let $\tilde{\mathbf{w}}_N = (\tilde{w}_{N;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ as in Theorem 2.21. To sum up all work contributions required to compute $I_{\tilde{\mathbf{w}}_N}$, we start with the evaluation of the u_l at the quadrature points. As in Proposition 2.20 let $\Lambda_{l;N} = \{\boldsymbol{\nu} \in \mathcal{F} : \tilde{w}_{N;\boldsymbol{\nu}} \geq l\}$. For each level $l \in \mathfrak{W}$, by (2.66) and with (2.69) we need to evaluate u_l at all points in $\mathrm{pts}(\Lambda_{l;N}) \cup \mathrm{pts}(\Lambda_{l+1;N})$. Since one evaluation is of cost $O(l^\gamma)$ by Assumption 2.30 (i) and using Lemma 2.28, this part sums up to

$$\begin{aligned}
&\sum_{0<l\in\mathfrak{W}} l^\gamma(|\mathrm{pts}(\Lambda_{l;N})| + |\mathrm{pts}(\Lambda_{l+1;N})|) \leq 2 \sum_{0<l\in\mathfrak{W}} (l+1)^\gamma |\mathrm{pts}(\Lambda_{l;N})| \leq C \sum_{0<l\in\mathfrak{W}} l^\gamma |\mathrm{pts}(\Lambda_{l;N})| \\
&\quad\leq C \sum_{0<l\in\mathfrak{W}} l^\gamma |\Lambda_{l;N}|^{1+\delta} \leq C \left(\sum_{0<l\in\mathfrak{W}} l^\gamma |\Lambda_{l;N}| \right)^{1+\delta} = C \left(\sum_{\{\boldsymbol{\nu}\in\mathcal{F} : \tilde{w}_{N;\boldsymbol{\nu}}\neq 0\}} \sum_{\{0<l\in\mathfrak{W} : l\leq\tilde{w}_{N;\boldsymbol{\nu}}\}} l^\gamma \right)^{1+\delta} \\
&\quad\leq C \left(\sum_{\{\boldsymbol{\nu}\in\mathcal{F} : \tilde{w}_{N;\boldsymbol{\nu}}\neq 0\}} \tilde{w}_{N;\boldsymbol{\nu}}^\gamma \right)^{1+\delta} \leq C|\tilde{\mathbf{w}}_N|^{\gamma(1+\delta)} \leq CN^{\gamma(1+\delta)},
\end{aligned} \tag{2.79}$$

where we have employed Assumption 2.24 to bound $\sum_{\{0<l\in\mathfrak{W} : l\leq\tilde{w}_{N;\boldsymbol{\nu}}\}} l^\gamma \leq C\tilde{w}_{N;\boldsymbol{\nu}}^\gamma$.

Given those function values, it remains to evaluate the interpolants, i.e. compute $\sum_{0<l\in\mathfrak{W}}(I_{\Lambda_l} - I_{\Lambda_{l+1}})u_l$. We use Assumption 2.30 (ii) to bound the cost of this part using Lemma 2.29 by

$$\begin{aligned}
&\sum_{0<l\in\mathfrak{W}} \left(\mathrm{cost}(I_{\Lambda_{l;N}}) + \mathrm{cost}(I_{\Lambda_{l+1;N}})\right) \leq C \sum_{0<l\in\mathfrak{W}} \left(|\Lambda_{l;N}|^{1+\delta} + |\Lambda_{l+1;N}|^{1+\delta}\right) \leq 2C \sum_{0<l\in\mathfrak{W}} |\Lambda_{l;N}|^{1+\delta} \\
&\quad\leq 2C \left(\sum_{0<l\in\mathfrak{W}} |\Lambda_{l;N}| \right)^{1+\delta} \leq 2C \left(\sum_{\{\boldsymbol{\nu}\in\mathcal{F} : \tilde{w}_{N;\boldsymbol{\nu}}\neq 0\}} \sum_{\{0<l\in\mathfrak{W} : l\leq\tilde{w}_{N;\boldsymbol{\nu}}\}} 1 \right)^{1+\delta} \leq 2C|\tilde{\mathbf{w}}_N|^{1+\delta}.
\end{aligned} \tag{2.80}$$

Adding (2.79) and (2.80) we get the total work amount of $CN^{\gamma(1+\delta)}$ for $\tilde{\mathbf{w}}_N$. By Theorem 2.21, the error $\|u - I_{\tilde{\mathbf{w}}_N}\|_{L^\infty(U,X)}$ satisfies equation 2.52. Hence (2.78) holds with $\mathbf{w}_N := \tilde{\mathbf{w}}_{\lfloor N^{1/(\gamma(1+\delta))}\rfloor}$. □

In the same fashion we can prove the following result.

Theorem 2.32. *Let* (2.29), (2.31) *and Assumption 2.30 for some* $\gamma \geq 1$ *be satisfied. Let the map* $u : U \to X$ *satisfy Assumption 2.16 and let* $\mathfrak{W}$ *be as in Assumption 2.24. For every* $\delta > 0$ *there exists a constant* $C = C(\delta) > 0$ *and for every* $N \in \mathbb{N}$ *there exists* $\mathbf{w}_N \in \mathfrak{W}^{\mathbb{N}}$ *such that the total work to compute* $Q_{\mathbf{w}_N} u \in X$ *is bounded by* CN*, and such that*

$$\left\| \int_U u(\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{y}) - Q_{\mathbf{w}_N} u \right\|_X \leq CN^{-\frac{r}{\gamma(1+\delta)}}, \qquad r = \begin{cases} \alpha & \text{if } \frac{p_1}{2} \leq \frac{1}{\alpha+1} \\ \alpha \frac{2p_0^{-1}-1}{\alpha+2p_0^{-1}-2p_1^{-1}} & \text{otherwise.} \end{cases} \tag{2.81}$$

Remark 2.33. *Assume nestedness of the interpolation/quadrature points as stated in Remark 2.25. Then, as observed in this remark,* $\mathrm{pts}(\Lambda) \leq |\Lambda|$ *for any downward closed index set* $\Lambda \subseteq \mathcal{F}$*. Notice that this is the only requirement involving* $\mathfrak{I}$*, that is employed in the first part of the proof of Theorem 2.31, which bounds the work associated to item (i) at the beginning of Section 2.3. Thus, if* $\mathfrak{I}$ *does not satisfy Assumptions 2.18, 2.24, the statements of Theorems 2.31, 2.32 remain true in terms of the work quantified by item (i).*

Remark 2.34. *In the setting of Theorems 2.31, 2.32, suppose that $G : X_{\mathbb{C}} \to \mathbb{C}$ is some bounded linear functional (often termed a "quantity of interest" in applications). Then with u, u_l as in Assumption 2.16, the functions $\tilde{u} := G(u) : U \to \mathbb{C}$ and $\tilde{u}_l := G(u_l) : U \to \mathbb{C}$ also satisfy the necessary requirements of Assumption 2.16 w.r.t. to the Banach space $\tilde{X} := \mathbb{C}$. Hence Theorems 2.31, 2.32, immediately apply to such quantities of interest as well.*

Remark 2.35. *Assume that our goal is to find a surrogate of some function $u : U \to X$, by computing its interpolant as in and under the assumptions of Theorem 2.31. Let at first $\gamma = 1$ in Assumption 2.30. Consider the proof of Theorem 2.31: After having taken care of the computation of all function values needed by the interpolant with cost $O(N^{1+\delta})$ (cp. (2.79)), one evaluation of $I_{\mathbf{w}_N}(\boldsymbol{y})$ at $\boldsymbol{y} \in U$ has cost $O(N^{1+\delta})$ with $\delta > 0$ arbitrary small (cp. (2.80)). As $N \to \infty$, the error is of size $O(N^{-r})$, with $r \le \alpha$ as in Theorem 2.31. On the other hand, by Assumption 2.30 it is possible to evaluate $u_N(\boldsymbol{y})$ with cost $O(N)$ and the error $\sup_{\boldsymbol{y}\in U} \|u(\boldsymbol{y}) - u_N(\boldsymbol{y})\|_X$ is of size $O(N^{-\alpha})$. Hence the computation of $I_{\mathbf{w}_N}$ has not brought an improvement in terms of the convergence rate.*

This can be improved by the following reinterpretation of α: For $\gamma > 1$ let $\tilde{u}_l := u_{\lceil l^\gamma \rceil}$. Then the "rate" in Assumption 2.16 becomes $\tilde{\alpha} = \gamma\alpha$. The computation of all function values needed by the interpolant now requires a cost of $O(N^{\gamma(1+\delta)})$ (cp. (2.79)). However, once this is done, an evaluation of $I_{\mathbf{w}_N}(\boldsymbol{y})$ is still of cost $O(N^{1+\delta})$ delivering an approximation with error of size $O(N^{-\tilde{r}})$ where $\tilde{r} \le \tilde{\alpha} = \gamma\alpha$ is now given by

$$\tilde{r} = \begin{cases} k\alpha & \text{if } p_1 \le (1+\gamma\alpha)^{-1}, \\ \gamma\alpha \frac{p_0^{-1}-1}{\gamma\alpha + p_0^{-1} - p_1^{-1}} & \text{otherwise.} \end{cases} \tag{2.82}$$

*A direct evaluation of $u_{\lceil N^\gamma \rceil}(\boldsymbol{y})$ on the other hand, has a complexity of $O(N^\gamma)$ and gives the error $O(N^{-\gamma\alpha})$. In particular for $0 < p_1 \le (1+\gamma\alpha)^{-1}$, we have improved the computational cost of finding an approximation with error $O(N^{-\gamma\alpha})$ from $O(N^\gamma)$ to $O(N^{1+\delta})$ (*after *having performed the precomputational step of determining all required function values with complexity $O(N^\gamma)$).*

3 Holomorphic extensions of Petrov-Galerkin approximations

So far, we analyzed multilevel gpc approximation of countably-parametric maps taking values in Banach spaces. The application we mainly have in mind are solution manifolds of countably parametric PDEs. The functions u_l in Assumption 2.16 then stem from a numerical approximation of PDE solutions. A particularly important class of numerical solvers comprise Petrov-Galerkin methods. The aim of this section is to verify Assumption 2.16 when u_l denotes the Petrov-Galerkin approximation in an l-dimensional subspace to some implicit equation (e.g. a PDE) with solution u. This will yield relevance of our result to a very broad class of problems. We proceed along the lines of the proofs of related results in previous works [15, 16, 21, 38]. However, the statements obtained here are more general and target the framework of holomorphic data-to-solution maps in Section 2. They are based on the implicit function theorem and on results from [45] on Petrov-Galerkin discretizations of nonlinear operator equations.

Let us consider an implicit equation of the type

$$\mathcal{N}(u, \xi) = 0, \tag{3.1}$$

where $\mathcal{N} : X \times \Xi \to Y'$, with X, Y being reflexive Banach spaces over $\mathbb{K}$ with $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ fixed, and additionally Ξ is either a real or a complex Banach space. As before, we denote their complexifications by $X_{\mathbb{C}}$, $Y_{\mathbb{C}}$ and $\Xi_{\mathbb{C}}$ (in particular, they coincide with X, Y, Ξ in case they are already complex), and mention that the dual space $(Y_{\mathbb{C}})'$ with the dual norm is isomorphic to $(Y')_{\mathbb{C}}$ with the complexification norm, so that we will not distinguish between the two in the following (see, e.g., [37, Proposition B.4.2]).

We shall use the following notation: for the residual map $\mathcal{N}$ as in (3.1), we write $\partial_1\mathcal{N}(u,\xi) \in L(X,Y')$ to refer to its first partial derivative w.r.t. u, and, similarly, $\partial_2\mathcal{N}(u,\xi) \in L(X,\Xi')$ to refer to its first partial derivative w.r.t. ξ. Here, $L(X,Y')$ and $L(X,\Xi')$ stand for the spaces of all linear mappings from X to Y' and from X to Ξ', respectively. For $k \in \mathbb{N}_0$, the k-th order differential of some function $F : X \to Y$ at $x \in X$ will be denoted by $d^kF(x) \in L(X, L(X, \dots L(X,Y)\dots))$. In case $k = 1$ we also write $dF(x) \in L(X,Y)$ instead. For $k \in \mathbb{N}_0$ and $O \subseteq X$, we set

$$\|F\|_{C^k(O)} := \sum_{j=0}^{k} \sup_{x\in O} \|d^jF(x)\|_{L(X,\dots L(X,Y)\dots)}. \tag{3.2}$$

Furthermore, for a ball with radius r and center $x \in X$, we shall write $B_r(x) \subseteq X$.

Assumption 3.1. *(i)* Existence of solutions: *The set $K \subseteq \Xi$ is compact and convex. For every parameter $\xi \in K$, there exists $u(\xi) \in X$ continuously depending on ξ such that $\mathcal{N}(u(\xi),\xi) = 0$.*

(ii) Holomorphy of $\mathcal{N}$: *The map $\mathcal{N}$ allows a holomorphic extension onto some open superset $O_{\mathcal{N}}$ of $\{(u(\xi),\xi) : \xi \in K\} \subseteq X_{\mathbb{C}} \times \Xi_{\mathbb{C}}$ and mapping to $Y'_{\mathbb{C}}$. Furthermore $\partial_1\mathcal{N}(u(\xi),\xi) \in L(X,Y')$ is an isomorphism for all $\xi \in K$.*

The holomorphic extension in the previous assumption will also be denoted by $\mathcal{N}$.

Lemma 3.2. *Let V, W be vector spaces over $\mathbb{R}$ and $F_{\mathbb{C}} : V_{\mathbb{C}} \to W_{\mathbb{C}}$ holomorphic such that $F|_V =: F_{\mathbb{R}} : V \to W$. Then for $x \in V$, $dF_{\mathbb{C}}(x) \in L(V_{\mathbb{C}}, W_{\mathbb{C}})$ is an isomorphism iff $dF_{\mathbb{R}}(x) \in L(V,W)$ is, and $\|dF_{\mathbb{C}}(x)^{-1}\|_{L(W_{\mathbb{C}},V_{\mathbb{C}})} \le 2\|dF_{\mathbb{R}}(x)^{-1}\|_{L(W,V)}$.*

Proof. We split $F_{\mathbb{C}}$ into its real and imaginary part via $F_{\mathbb{C}} = F_1 + \mathrm{i}F_2$, and for x_1, $x_2 \in V$ we consider $F_j(x_1 + \mathrm{i}x_2)$, $j \in \{1,2\}$. Fix $x \in V$. Since $F_{\mathbb{C}}(V) \subseteq W$, $\partial_{x_1}F_2(x) = 0$. By the Cauchy Riemann equations $\partial_{x_2}F_1(x) = -\partial_{x_1}F_2(x) = 0$, and furthermore $\partial_{x_1}F_1(x) = \partial_{x_2}F_2(x)$. From this we can conclude,

$$dF_{\mathbb{C}}(x)(h + \mathrm{i}g) = \partial_{x_1}F_1(x)(h) + \mathrm{i}\partial_{x_1}F_1(x)(g). \tag{3.3}$$

Clearly $\partial_{x_1}F_1(x) = dF_{\mathbb{R}}(x) \in L(V,W)$. We have shown $dF_{\mathbb{C}}(x)(h+\mathrm{i}g) = dF_{\mathbb{R}}(x)(h) + \mathrm{i}dF_{\mathbb{R}}(x)(g)$, which gives the first part of the claim. To estimate the norm of the inverse, let $a + \mathrm{i}b$ be arbitrary in $W_{\mathbb{C}}$. Then

$$\begin{aligned}\|dF_{\mathbb{C}}^{-1}(x)(a+\mathrm{i}b)\|_{V_{\mathbb{C}}} &= \|dF_{\mathbb{R}}^{-1}(x)(a) + \mathrm{i}dF_{\mathbb{R}}^{-1}(x)(b)\|_{V_{\mathbb{C}}} \le \|dF_{\mathbb{R}}^{-1}(x)(a)\|_V + \|dF_{\mathbb{R}}^{-1}(x)(b)\|_V \\ &\le \|dF_{\mathbb{R}}^{-1}(x)\|_{L(W,V)}(\|a\|_W + \|b\|_W) \le 2\|dF_{\mathbb{R}}^{-1}(x)\|_{L(W,V)}(\|a+\mathrm{i}b\|_{W_{\mathbb{C}}}). \quad \square\end{aligned}$$

Proposition 3.3. *Under Assumption 3.1 there exists an open set $O \subseteq \Xi_{\mathbb{C}}$ containing K, onto which u allows a unique continuous extension satisfying $\mathcal{N}(u(\xi),\xi) = 0$ for all $\xi \in O$. Furthermore u is holomorphic on O and it holds $\sup_{\xi\in O}\|u(\xi)\|_{X_{\mathbb{C}}} \le C_{O;u} < \infty$. Here, O and $C_{O;u}$ only depend on u through*

$$O_{\mathcal{N}}, \quad \|\mathcal{N}\|_{C^2(O_{\mathcal{N}},Y'_{\mathbb{C}})}, \quad \sup_{\xi\in K}\|\partial_1\mathcal{N}^{-1}(u(\xi),\xi)\|_{L(Y',X)} \quad \textit{and} \quad \sup_{\xi\in K}\|u(\xi)\|_X. \tag{3.4}$$

Proof. According to Lemma 3.2, for every $\xi \in K$, there holds $\partial_1 \mathcal{N}(u(\xi), \xi) \in L(X_\mathbb{C}, Y'_\mathbb{C})$ and this partial derivative is an isomorphism. Let $A := \partial_1 \mathcal{N}(u(\xi), \xi) \in L(X_\mathbb{C}, Y'_\mathbb{C})$. Notice that any operator $B \in L(X_\mathbb{C}, Y'_\mathbb{C})$ with $\|A - B\|_{L(X_\mathbb{C}, Y'_\mathbb{C})} < \|A^{-1}\|^{-1}_{L(Y'_\mathbb{C}, X_\mathbb{C})}$ is boundedly invertible with

$$\|B^{-1}\|_{L(Y'_\mathbb{C}, X_\mathbb{C})} \leq \|A^{-1}\|_{L(Y'_\mathbb{C}, X_\mathbb{C})}(1 - \|A - B\|_{L(X_\mathbb{C}, Y'_\mathbb{C})}\|A^{-1}\|_{L(Y'_\mathbb{C}, X_\mathbb{C})})^{-1}.$$

Thus, we can find $\delta_\xi > 0$ solely depending on ξ and on the quantities in (3.4), such that $B_{\delta_\xi}(u(\xi)) \times B_{\delta_\xi}(\xi) \subseteq O_\mathcal{N}$ as well as

$$\|\partial_1 \mathcal{N}^{-1}(v, \zeta)\|_{L(Y'_\mathbb{C}, X_\mathbb{C})} \leq 2\|\partial_1 \mathcal{N}^{-1}(u(\xi), \xi)\|_{L^\infty(K, L(Y'_\mathbb{C}, X_\mathbb{C}))} \qquad \forall\, v, \zeta \in B_{\delta_\xi}(u(\xi)) \times B_{\delta_\xi}(\xi) \subseteq X_\mathbb{C} \times \Xi_\mathbb{C} \tag{3.5}$$

is ensured.

Next, following the notation in [23, Theorem 15.1], for $\xi \in K$ fixed define the map $S : X_\mathbb{C} \times \Xi_\mathbb{C} \to X_\mathbb{C}$ via

$$S(v, \zeta) := \partial_1 \mathcal{N}(u(\xi), \xi)^{-1}(\mathcal{N}(v, \zeta)) - v. \tag{3.6}$$

It holds that $\partial_1 S(u(\xi), \xi) = 0$ in $L(X_\mathbb{C}, X_\mathbb{C})$ and $S(u(\xi), \xi) = 0$ in $X_\mathbb{C}$. Therefore, we can further decrease δ_ξ, and find $0 < r_\xi \leq \delta_\xi$ such that

$$\|\partial_1 S(v, \zeta)\|_{L(X_\mathbb{C}, X_\mathbb{C})} < \frac{1}{2} \quad \forall\, v, \zeta \in B_{\delta_\xi}(u(\xi)) \times B_{\delta_\xi}(\xi) \qquad \text{and} \qquad \|S(u(\xi), \cdot)\|_{L^\infty(B_{r_\xi}(\xi), Y'_\mathbb{C})} < \frac{\delta}{2}. \tag{3.7}$$

We point out that δ_ξ, r_ξ can again be chosen in sole dependence on ξ and on the four quantities in (3.4).

Set $N_\xi := B_{r_\xi}(\xi) \subseteq \Xi_\mathbb{C}$ and $M_\xi := B_{\delta_\xi}(u(\xi)) \subseteq X_\mathbb{C}$. From the proof of the implicit function theorem as provided in [23, Theorem 15.1], properties (3.7) ensure the existence of a unique map $\tilde{u}$ from N_ξ to M_ξ satisfying $\tilde{u}(\xi) = u(\xi)$ and $\mathcal{N}(\tilde{u}(\xi'), \xi') = 0$ for all $\xi' \in N_\xi$ (essentially because $S(\cdot, \zeta)$ is a contraction, so that the Banach fixed point theorem can be applied to it). Additionally, this extension is holomorphic [23, Proposition 15.2, Theorem 15.3].

By compactness, we can cover K with finitely many balls $N_{\xi_1}, \ldots, N_{\xi_n}$. The union O of these balls still only depends on (3.4). To show that the resulting extension defined on each N_{ξ_i} via $\tilde{u}_{\xi_i}$ is well-defined, assume wlog $N := N_{\xi_1} \cap N_{\xi_2} \neq \emptyset$. We need to verify $\tilde{u}_{\xi_1}|_N \equiv \tilde{u}_{\xi_2}|_N$. Using convexity of K and the fact that N_{ξ_i} is a ball around $\xi_i \in K$ by construction, it holds $K \cap N \neq \emptyset$. Fix $x \in K \cap N$. Then $\tilde{u}_{\xi_1}(x) = u(x) = \tilde{u}_{\xi_2}(x)$ and thus $u(x) \in M_{\xi_1} \cap M_{\xi_2} \neq \emptyset$. Using continuity of $u_{\xi_i} : N_{\xi_i} \to M_{\xi_i}$, we can find open sets N_x, M_x such that $x \in N_x \subseteq N_{\xi_1} \cap N_{\xi_2}$, $u(x) \in M_x \subseteq M_{\xi_1} \cap M_{\xi_2}$ and such that $\tilde{u}_{\xi_i} : N_x \to M_x$, $i \in \{1, 2\}$. By the stated uniqueness of $\tilde{u}_{\xi_i} : N_{\xi_i} \to M_{\xi_i}$, $i \in \{1, 2\}$, those two functions coincide on the open subset N_x of $N_{\xi_1} \cap N_{\xi_2}$. Since they are holomorphic, they coincide on all of $N_{\xi_1} \cap N_{\xi_2}$ by the identity principle (see for example [40, Proposition 5.7] for the statement in Banach spaces).

Next, differentiating $\mathcal{N}(u(\zeta), \zeta) = 0$ w.r.t. ζ, we find

$$u'(\zeta) = -\partial_1 \mathcal{N}(u(\zeta), \zeta) \circ \partial_2 \mathcal{N}(u(\zeta), \zeta) \in L(\Xi_\mathbb{C}, X_\mathbb{C}), \tag{3.8}$$

and thus

$$\|u'\|_{C^0(N_\xi, L(\Xi_\mathbb{C}, X_\mathbb{C}))} \leq \|\partial_1 \mathcal{N}\|_{C^0(M_\xi \times N_\xi, L(X_\mathbb{C}, Y'_\mathbb{C}))}\|\partial_2 \mathcal{N}\|_{C^0(M_\xi \times N_\xi, L(\Xi_\mathbb{C}, Y'_\mathbb{C}))} \leq \|\mathcal{N}\|^2_{C^1(O_\mathcal{N}; Y'_\mathbb{C})}. \tag{3.9}$$

Finally, since each $\zeta \in O$ is in some ball $B_{r_\xi}(\xi) = N_\xi$, we have

$$\|u(\zeta)\|_{X_\mathbb{C}} = \|u(\xi) + \int_0^1 u'(\xi + t(\zeta - \xi))(\zeta - \xi)\,\mathrm{d}t\|_{X_\mathbb{C}} \leq \|u(\xi)\|_{X_\mathbb{C}} + \|u'\|_{L^\infty(N_\xi, L(\Xi_\mathbb{C}, X_\mathbb{C}))} r_\xi. \quad (3.10)$$

This together with (3.10) shows that $C_{O;u}$ as in the statement of the proposition can be found. □

Assumption 3.4 (Convergence of Petrov-Galerkin scheme)**.** *The spaces X, Y are reflexive and separable, and there are sequences $\{X_l\}_{l\in\mathbb{N}} \subset X$ and $\{Y_l\}_{l\in\mathbb{N}} \subset Y$ of subspaces of dimension $l = \dim(X_l) = \dim(Y_l)$ such that $\bigcup_{l\in\mathbb{N}} X_l$ and $\bigcup_{l\in\mathbb{N}} Y_l$ are dense in X and Y, respectively, and $\mathcal{N}_l := \mathcal{N}|_{X_l} : X_l \times \Xi \to Y_l'$. Moreover, with the additional convention $X_\infty := X$, $Y_\infty := Y$, for some $\beta > 0$, every $l \in \mathbb{N} \cup \{\infty\}$ and every $\xi \in K$ it holds that*

$$\partial_1 \mathcal{N}_l(u(\xi), \xi) \in L(X_l, Y_l') \text{ is an isomorphism and } \|\partial_1 \mathcal{N}_l(u(\xi), \xi)^{-1}\|_{L(Y_l', X_l)} \leq \beta^{-1}. \quad (3.11)$$

With the notation of Assumption 3.4, (in case it exists) we write $u_l(\xi) \in (X_l)_\mathbb{C}$ for the solution of

$$\langle \mathcal{N}(u_l(\xi), \xi), v_l\rangle = \langle \mathcal{N}_l(u_l(\xi), \xi), v_l\rangle = 0 \quad \text{for all } v_l \in (Y_l)_\mathbb{C}. \quad (3.12)$$

Proposition 3.5. *Let Assumptions 3.1 and 3.4 hold. Then, there exists an open superset $O \subseteq \Xi_\mathbb{C}$ of K, and constants $0 < C < \infty$, $\eta < \infty$, and $l_0 \in \mathbb{N}_0$ such that for each $l \geq l_0$, and for all $\xi \in O$ there exists a unique $u_l(\xi) \in X_\mathbb{C}$ satisfying* (3.12) *and the error bound*

$$\|u(\xi) - u_l(\xi)\|_{X_\mathbb{C}} \leq \eta\,. \quad (3.13)$$

Furthermore, the parametric solution maps $u : O \to X_\mathbb{C}$ and $u_l : O \to (X_l)_\mathbb{C}$ are holomorphic, uniformly bounded w.r.t. the discretization parameter l and, moreover, there holds

$$\|u(\xi) - u_l(\xi)\|_{X_\mathbb{C}} \leq C \min_{x_l \in X_l} \|u(\xi) - x_l\|_{X_\mathbb{C}} \qquad \forall\, \xi \in O_\varepsilon,\ l \in \mathbb{N}\,. \quad (3.14)$$

Proof. By Proposition 3.3 we can holomorphically extend $u : O_1 \to X_\mathbb{C}$ to a uniformly bounded function on some $O_1 \subseteq \Xi_\mathbb{C}$.

In the next step, we prove the existence of u_l. We start with the following preliminary observations exploiting the compactness of K:

(i) By Lemma 3.2 and (3.11),

$$\|\partial_1 \mathcal{N}_l(u(\xi), \xi)^{-1}\|_{L((Y_l)'_\mathbb{C}, (X_l)_\mathbb{C})} \leq 2\beta^{-1} \quad (3.15)$$

for all $\xi \in K$ and all $l \in \mathbb{N}_0$. Let $O_\mathcal{N}$ be as in Assumption 3.1. Then $\mathcal{N}_l$ is holomorphic on $O_{\mathcal{N}_l} := O_\mathcal{N} \cap (X_l)_\mathbb{C} \times \Xi_\mathbb{C}$ for every $l \in \mathbb{N}$. Using compactness of K and possibly shrinking $O_\mathcal{N}$ if necessary, we can assume due to the fact that $\mathcal{N}$ is holomorphic and with a similar argument as in the beginning of the proof of Proposition 3.3, that there exists $M < \infty$ independent of l with

$$\|\partial_1 \mathcal{N}_l^{-1}\|_{C^0(O_{\mathcal{N}_l}, L((X_l)_\mathbb{C}, (Y_l)'_\mathbb{C}))} \leq M \qquad \text{and} \qquad \|\mathcal{N}_l\|_{C^2(O_{\mathcal{N}_l})} \leq M. \quad (3.16)$$

(ii) Again by compactness of K, we obtain the existence of $O_2 \subseteq O_1 \subseteq \Xi_\mathbb{C}$ open and containing K as well as $\varepsilon_0 > 0$ such that $B_{\varepsilon_0}(u(\xi)) \times B_{\varepsilon_0}(\xi) \subseteq O_\mathcal{N}$ for all $\xi \in O_2$. Thus, (3.16) implies in particular that $\partial_1 \mathcal{N}(\cdot, \xi)$ is a Lipschitz mapping on $B_{\varepsilon_0}(u(\xi))$ for $\xi \in O_2$ with a Lipschitz constant independent of ξ.

(iii) Using density of $\bigcup_{l\in\mathbb{N}} X_l$ in X, compactness of K and continuity of $O_1 \ni \xi \mapsto u(\xi)$, for $\eta > 0$ arbitrary we can find $\tilde{l}_0 \in \mathbb{N}$ and $O_3 \subseteq O_2 \subseteq \Xi_{\mathbb{C}}$ open and containing K with

$$\sup_{\xi\in O_3} \min_{x_l\in X_l} \|u(\xi) - x_l\|_{X_{\mathbb{C}}} < \eta \tag{3.17}$$

for all $l \geq \tilde{l}_0$.

We now assume $\eta(\varepsilon_0, M) > 0$ in (3.17) to be small enough, the exact dependence will be explained in the following. Fix $\xi \in O_3$. Due to Assumption 3.4 and (ii), by [45, Theorem 4] (which also holds in the case of complex Banach spaces, as can be checked) there exists $l_0 \in \mathbb{N}$ and $\delta_0 > 0$ such that there is a unique $u_l(\xi) \in (X_l)_{\mathbb{C}}$ satisfying

$$\langle \mathcal{N}(u_l(\xi), \xi), y_l \rangle = 0, \quad \forall\, y_l \in (Y_l)_{\mathbb{C}} \qquad \text{and} \qquad \|u(\xi) - u_l(\xi)\|_{X_{\mathbb{C}}} \leq \delta_0 \,. \tag{3.18}$$

Retracing the steps of the proof in [45], one observes that l_0 and δ_0 solely depend on ε_0, M and η, as long as η is chosen small enough in dependence of ε_0 and M (the crucial equations determining this dependence are [45, (2.11) and (3.26)]). With these choices, due to (ii), $u_l(\xi)$ is well-defined for all $\xi \in O_3$ and $l \geq l_0$. Moreover, for $\xi \in O_3$ there holds the apriori estimate [45, (3.13)]

$$\|u(\xi) - u_l(\xi)\|_{X_{\mathbb{C}}} \leq C \min_{x_l\in(X_l)_{\mathbb{C}}} \|u(\xi) - x_l\|_{X_{\mathbb{C}}} \tag{3.19}$$

with $C = C(M)$ independent of ξ.

It remains to show holomorphy of $\xi \mapsto u_l(\xi)$. We start by showing continuity. We observe that by Proposition 3.3, for any $\xi \in O_3$, there exists a unique continuous function $\tilde{u}_l$ defined on some ball $B_{r_\xi}(\xi) \subseteq \Xi_{\mathbb{C}}$ mapping to $(X_l)_{\mathbb{C}}$ which is an extension, i.e. $\tilde{u}_l(\xi) = u_l(\xi)$ and $\mathcal{N}_l(\tilde{u}_l(\zeta), \zeta) = 0 \in (Y_l)'_{\mathbb{C}}$ for all $\zeta \in B_{r_\xi}(\xi)$. Since $\xi \mapsto u(\xi)$ is Lipschitz continuous (in some neighbourhood of K), we have

$$\|u_l(\xi)-u_l(\zeta)\|_{X_{\mathbb{C}}} \leq \|u_l(\xi)-u(\xi)\|_{X_{\mathbb{C}}}+\|u(\xi)-u(\zeta)\|_{X_{\mathbb{C}}}+\|u(\zeta)-u_l(\zeta)\|_{X_{\mathbb{C}}} \leq C(\eta+\|\xi-\zeta\|_{X_{\mathbb{C}}}) \tag{3.20}$$

for some $C < \infty$. Decreasing η and choosing ξ, ζ close enough, the uniqueness of $\tilde{u}_l$, as a mapping to a neighbourhood of $u_l(\xi)$, then ensures $u_l(\zeta) = \tilde{u}_l(\xi)$ on some neighbourhood of ξ. This shows (local) continuity of u_l. Similar as before, one checks that all constants and radii can be chosen independent of ξ, so that η can be chosen small enough independent of ξ. Hence, with this choice, u_l is continuous. Finally, employing once more Proposition 3.3, we infer that $u_l : O \to (X_l)_{\mathbb{C}}$ must be holomorphic with $O := O_1 \cap O_2 \cap O_3$. $\square$

4 Applications

In order for the previous results to apply to a large range of parametric partial differential and operator equations, up to this point, we kept the presentation general. As an illustrative example we shall now concentrate on second order linear elliptic Dirichlet problems on polygonal domains D. Due to the presence of corners in D, solutions of the parametric problem exhibit, in general, corner singularities. We shall require analytic regularity in weighted Sobolev spaces, from [8]. For an example of a nonlinear equation, we refer to [21], where shape holomorphy for the stationary Navier-Stokes equations has been shown. Holomorphic dependence of solutions in parametric domains for the Maxwell equations was established in [38]. These results can be adapted to establish the necessary requirements of Theorems 2.31 and 2.32 also for the scattering problems considered in [38].

4.1 Linear elliptic advection-reaction-diffusion equations in polygons

Let $\mathrm{D} \subseteq \mathbb{R}^2$ be a polygonal Lipschitz domain. For $\boldsymbol{y} \in U$ consider

$$-\operatorname{div}(a(\boldsymbol{y})\nabla\tilde{u}(\boldsymbol{y})) + c(\boldsymbol{y}) \cdot \nabla\tilde{u}(\boldsymbol{y}) + e(\boldsymbol{y})\tilde{u}(\boldsymbol{y}) = f(\boldsymbol{y}) \qquad \text{in D,} \tag{4.1a}$$

$$\tilde{u}(\boldsymbol{y}) = g(\boldsymbol{y}) \qquad \text{on } \partial\mathrm{D}. \tag{4.1b}$$

Here $\tilde{u}(\boldsymbol{y}) : \mathrm{D} \to \mathbb{C}$, and ∇ and div act on the spatial variable $x \in \mathrm{D}$. We denote the parametric differential operator on the left-hand side of (4.1a) by $\mathcal{A}(u(\boldsymbol{y}), \boldsymbol{y})$ and consider it as an operator $\mathcal{A} : X \times U \to X'$, where $X := H_0^1(\mathrm{D})$ and where $\mathcal{A}$ is linear in the first argument. The following restrictions are imposed on the data:

Assumption 4.1. *(i)* Holomorphic parameter dependence: *there is a function $\tilde{g} : U \to H^1(\mathrm{D}, \mathbb{C})$ such that for every $\boldsymbol{y} \in U$ it holds that $\tilde{g}|_{\partial\mathrm{D}}(\boldsymbol{y}) = g(\boldsymbol{y}) \in H^{1/2}(\partial\mathrm{D})$; moreover, the functions $a : U \to L^\infty(\mathrm{D}, \mathbb{C}^{2\times 2})$, $c : U \to L^\infty(\mathrm{D}, \mathbb{C}^2)$, $e : U \to L^\infty(\mathrm{D}, \mathbb{C})$, $f : U \to H^{-1}(\mathrm{D}, \mathbb{C})$ and $\tilde{g} : U \to H^1(\mathrm{D}, \mathbb{C})$ admit $(\boldsymbol{b}_0, \varepsilon)$-holomorphic extensions to a complex neighbourhood $O_{\boldsymbol{b}_0,\varepsilon}$ of U as in* (2.2) *with $\boldsymbol{b}_0 \in \ell^{p_0}$, $0 < p_0 < 1$, $\varepsilon > 0$.*

(ii) Uniform ellipticity: *with $\Re(\cdot)$ denoting the real part and $\overline{v}$ the conjugation of v, we have*

$$\inf_{\boldsymbol{z} \in O_{\boldsymbol{b}_0,\varepsilon}} \inf_{0 \neq v \in X} \frac{\Re(\mathcal{A}(v, \boldsymbol{z})(\overline{v}))}{\|v\|_X^2} > 0. \tag{4.2}$$

We consider the parametric PDE (4.1) in the following weak form: Given $\boldsymbol{y} \in U$, find $u(\boldsymbol{y}) \in X$ s.t. with $\tilde{f} := f + \operatorname{div} a(\nabla\tilde{g}) - c \cdot \tilde{g} - e\tilde{g} \in X'$

$$\int_{\mathrm{D}} (a(\boldsymbol{y})\nabla u(\boldsymbol{y}) \cdot \nabla v + (c(\boldsymbol{y}) \cdot \nabla u(\boldsymbol{y}))v + e(\boldsymbol{y})u(\boldsymbol{y})v)\,\mathrm{d}x = {}_{X'}\langle \tilde{f}(\boldsymbol{y}), v\rangle_X \qquad \forall v \in X\ . \tag{4.3}$$

The relation of $u(\boldsymbol{y})$ to the parametric solution $\tilde{u}(\boldsymbol{y})$ in (4.1) is given by $u(\boldsymbol{y}) = \tilde{u}(\boldsymbol{y}) - \tilde{g}(\boldsymbol{y})$.

4.1.1 Polygonal domains

Following [8], for $s \in \mathbb{N}_0$, $\zeta \in \mathbb{R}$ we introduce the Kondratiev type Sobolev spaces

$$\mathcal{K}_\zeta^s(\mathrm{D}) := \{u : \mathrm{D} \to \mathbb{C} \,:\, r_{\mathrm{D}}^{|\alpha|-\zeta}\partial^\alpha u \in L^2(\mathrm{D}), |\alpha| \le s\}. \tag{4.4}$$

Here $r_{\mathrm{D}} : \mathrm{D} \to \mathbb{R}$ is a smooth function which, in a vicinity of a corner, equals the distance to this corner. Furthermore, we define

$$\mathcal{W}^{s,\infty}(\mathrm{D}) := \{u : \mathrm{D} \to \mathbb{C} \,:\, r_{\mathrm{D}}^{|\alpha|}\partial^\alpha u \in L^\infty(\mathrm{D}), |\alpha| \le s\}, \tag{4.5}$$

where the norms of the spaces in (4.4), (4.5) are the obvious ones. For the details on these spaces we refer to [8]. In addition to Assumption 4.1, we work with *weighted* regularity where $\zeta \in \mathbb{R}$ and $s \in \mathbb{N}$ are to be specified later.

Assumption 4.2 (Uniform parametric regularity in weighted spaces)**.** *Under Assumption 4.1 and for some $\boldsymbol{b}_1 \in \ell^{p_1}$, $p_0 \le p_1 < 1$ with $\boldsymbol{b}_1 \ge \boldsymbol{b}_0$ componentwise, it holds for $O_{\boldsymbol{b}_1,\varepsilon} \subseteq O_{\boldsymbol{b}_0,\varepsilon}$ as defined in* (2.2)

$$\sup_{\boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}} (\|a(\boldsymbol{z})\|_{\mathcal{W}^{s,\infty}(\mathrm{D})} + \|r_{\mathrm{D}}c(\boldsymbol{z})\|_{\mathcal{W}^{s,\infty}(\mathrm{D})} + \|r_{\mathrm{D}}^2 e(\boldsymbol{z})\|_{\mathcal{W}^{s,\infty}(\mathrm{D})}) < \infty, \tag{4.6a}$$

$$\sup_{\boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}} \|f(\boldsymbol{z}) + \operatorname{div}(\tilde{g}(\boldsymbol{z})) - b \cdot \nabla\tilde{g}(\boldsymbol{z}) + c\tilde{g}(\boldsymbol{z})\|_{\mathcal{K}_{\zeta-1}^{s-1}(\mathrm{D})} < \infty. \tag{4.6b}$$

Example 4.3 (Affine dependence)**.** *Consider* $s_0 = 0$*,* $s_1 > 0$ *and* $\overline{a} \in \mathcal{W}^{s_1,\infty}(\mathrm{D}, \mathbb{R}^{2\times 2})$ *symmetric such that for some* $c_0 > 0$*,* $\xi^\top \overline{a}\xi \geq c_0|\xi|^2$ *for all* $\xi \in \mathbb{C}^2$ *uniformly on* D*. Furthermore, let* $(\psi_{a;j})_{j\in\mathbb{N}} \subseteq \mathcal{W}^{s_1,\infty}(\mathrm{D}, \mathbb{C}^{2\times 2})$ *with* $\sum_{j\in\mathbb{N}} \|(\psi_{a;j})_{lk}\|_{L^\infty(\mathrm{D})} < \infty$ *for all* l*,* $k \in \{1, 2\}$*. Set* $b_{a;i;j} := \max_{l,k} \|(\psi_{a;j})_{lk}\|_{\mathcal{W}^{s_i,\infty}}$*, and assume* $\boldsymbol{b}_{a;i} := (b_{a;i;j})_{j\in\mathbb{N}} \in \ell^{p_i}(\mathbb{N})$*,* $i \in \{0, 1\}$ *for some* $0 < p_0 \leq p_1 < 1$*. Then we define for* $\theta > 0$ *(to be chosen later)*

$$a(\boldsymbol{z}) := \overline{a} + \theta \sum_{j\in\mathbb{N}} z_j \psi_{a;j}, \tag{4.7}$$

which for $\varepsilon > 0$ *fixed and* $\boldsymbol{z} \in O_{\boldsymbol{b}_{a;i},\varepsilon}$ *(cp. (2.2)) arbitrary,* $i \in \{0, 1\}$*, satisfies for some* $\boldsymbol{\rho}$ *as in* (2.1)

$$\|a(\boldsymbol{z}_i)\|_{\mathcal{W}^{s_i,\infty}} \leq \theta \sum_{j\in\mathbb{N}} b_{a;i;j}\rho_j = \sum_{j\in\mathbb{N}} b_{a;i;j}(\rho_j - 1) + \sum_{j\in\mathbb{N}} b_{a;i;j} \leq \varepsilon + \|(b_{a;i;j})_{j\in\mathbb{N}}\|_{\ell^1} < \infty. \tag{4.8}$$

Thus $a(\boldsymbol{z}) \in \mathcal{W}^{s_i,\infty}$ *is well-defined for all* $\boldsymbol{z} \in O_{\boldsymbol{b}_{a;i};\varepsilon}$*. Here we have used* (2.1) *and* $\boldsymbol{b}_{a;i} \in \ell^{p_i} \hookrightarrow \ell^1$*. The fact that* U *is compact and* $a(\boldsymbol{z})$ *continuously depending on* $\boldsymbol{z} \in O_{\boldsymbol{b}_{a;i};\varepsilon}$ *in* $\mathcal{W}^{s_i,\infty}$*, which can be deduced similar as in* (4.8)*, also shows uniform boundedness of* $\|a(\boldsymbol{z})\|_{\mathcal{W}^{s_i,\infty}}$ *as in* (4.6a)*,* $i \in \{0, 1\}$*, and hence Assumption 4.2 with* $s = s_1$ *for* $a(\boldsymbol{z})$*. Holomorphy of* $a(\boldsymbol{z}) \in \mathcal{W}^{s_0,\infty} = L^\infty$ *in each* z_j *is trivial, since the dependence is affine and continuous. Thus* $(\boldsymbol{b}_{a;0}, \varepsilon, L^\infty)$*-holomorphy of* $\boldsymbol{z} \mapsto a(\boldsymbol{z})$ *is verified, which implies Assumption 4.1 (i) for* a*. Choosing* $\theta > 0$ *sufficiently small ensures* $\overline{\xi}^\top a(\boldsymbol{z})\xi \geq |\xi|^2 c_0/2$ *for all* $\xi \in \mathbb{C}^2$ *and all* $\boldsymbol{z} \in O_{\boldsymbol{b}_{a;0}}$*; here* c_0 *is the ellipticity constant of* a*.*

Similarly we can affinely expand c*,* e*,* f *and* $\tilde{g}$ *in the appropriate spaces from Assumptions 4.1, 4.2 (for some fixed* ζ *of our choice), with some functions* $\psi_{*;j}$*,* $* \in \{c, e, f, g\}$*, whose norms are in* ℓ^{p_i}*, w.r.t. the spaces utilizing the smoothness parameter* s_i*,* $i \in \{0, 1\}$*. This will ensure Assumption 4.2 with* $s = s_1$*. Imposing appropriate smallness conditions on the expansion functions corresponding to* $c(\boldsymbol{z})$ *and* $e(\boldsymbol{z})$ *(as we did with* $\theta > 0$ *above) will guarantee Assumption 4.1 (ii). Item (i) of this assumption follows in the same manner as before and with* $b_{i;j} := \max_{*\in\{a,c,e,f,g\}} b_{*;i;j}$*,* $i \in \{0, 1\}$*.*

Next, we consider an example of uncertain inputs which are non-affine. These inputs arise, for example, in connection with domain UQ.

Example 4.4 (Domain uncertainty quantification)**.** *We introduce domain transformations analogous to [21, 38]. Let* $\mathrm{D} = \mathrm{D}_{\boldsymbol{0}} \subseteq \mathbb{R}^2$*, and assume that* $(\psi_j)_{j\in\mathbb{N}}$ *is a sequence of* $W^{1,\infty}(\mathbb{R}^2, \mathbb{R}^2)$ *functions such that with* $s_0 := 0$*, and with some fixed* $s_1 \geq 0$ *and* $b_{i;j} := \|\psi_j\|_{W^{1,\infty}(\mathbb{R}^2)} + \|d\psi_j\|_{\mathcal{W}^{s_i,\infty}(\mathbb{R}^2)}$*,* $i \in \{0, 1\}$*,*

$$\sum_{j\in\mathbb{N}} \|\psi_j\|_{W^{1,\infty}(\mathbb{R}^2)} < 1 \qquad \text{and} \qquad \boldsymbol{b} = (b_{i;j})_{j\in\mathbb{N}} \in \ell^{p_i}(\mathbb{N}) \tag{4.9}$$

for some $0 < p_0 \leq p_1 < 1$*. For* $\boldsymbol{y} \in U$*, setting for some fixed* $\theta \in (0, 1]$

$$T(\boldsymbol{y})(x) := x + \theta \sum_{j\in\mathbb{N}} y_j \psi_j(x), \tag{4.10}$$

i.e. $T(\boldsymbol{y}) : \mathbb{R}^2 \to \mathbb{R}^2$*, one checks that* $\boldsymbol{y} \to T(\boldsymbol{y})$ *is a continuous mapping from* U *to* $W^{1,\infty}(\mathbb{R}^2, \mathbb{R}^2)$ *with* $T(\boldsymbol{y})$ *and its inverse being bijective and uniformly Lipschitz independent of* $\boldsymbol{y}$*. Assume* $T(\boldsymbol{y})$ *to be such that* $\mathrm{D}_{\boldsymbol{y}} := T(\boldsymbol{y})(\mathrm{D}_{\boldsymbol{0}})$ *is a Lipschitz domain for all* $\boldsymbol{y} \in U$*, and consider* (4.1) *on the domain*

$\mathrm{D}_{\boldsymbol{y}}$. From the data we require $a : \mathrm{D}_H \to \mathbb{C}^{2\times 2}$, $c : \mathrm{D}_H \to \mathbb{C}^2$ and e, f, $g : \mathrm{D}_H \to \mathbb{C}$ to be analytic on some hold-all D_H with $\mathrm{D}_{\boldsymbol{y}} + B_\varepsilon(0) \subseteq \mathrm{D}_H \subseteq \mathbb{R}^2$ for some fixed $\varepsilon > 0$ and all $\boldsymbol{y} \in U$. Furthermore, we shall assume (4.1) to be uniformly well-posed, in the sense that the differential operator on the left-hand side of (4.1) is bounded and boundedly invertible as a mapping from $H_0^1(\mathrm{D}_{\boldsymbol{y}})$ to $H^{-1}(\mathrm{D}_{\boldsymbol{y}})$, with bounds independent of $\boldsymbol{y} \in U$.

By the transformation $T(\boldsymbol{y})(x)$, the model problem (4.3) on the parametric physical domain $\mathrm{D} = \mathrm{D}_{\boldsymbol{y}}$ is equivalent to the following parametric model in the fixed, nominal domain $\mathrm{D}_{\mathbf{0}}$*: For $\boldsymbol{y} \in U$ and with $X = H_0^1(\mathrm{D}_{\mathbf{0}})$, find $u(\boldsymbol{y}) \in X$ s.t. with $\tilde{f} := f + \mathrm{div}(a\nabla g) - c \cdot \nabla g - eg$*

$$\int_{\mathrm{D}_{\mathbf{0}}} \Big(\nabla v^\top dT(\boldsymbol{y})^{-1}(a \circ T(\boldsymbol{y}))dT(\boldsymbol{y})^{-\top}\nabla u(\boldsymbol{y}) + c \circ T(\boldsymbol{y}) \cdot dT(\boldsymbol{y})\nabla u(\boldsymbol{y})v \\ + e \circ T(\boldsymbol{y})u(\boldsymbol{y})v \Big) \det dT(\boldsymbol{y})\mathrm{d}x = \int_{\mathrm{D}_{\mathbf{0}}} \tilde{f} \circ T(\boldsymbol{y}) \det dT(\boldsymbol{y})v \qquad \forall\ v \in X, \tag{4.11}$$

where $dT(\boldsymbol{y}) : \mathrm{D}_{\mathbf{0}} \to \mathbb{R}^{2\times 2}$ denotes the Jacobian. The solution $u(\boldsymbol{y})$ of (4.11) is then related to the solution $\tilde{u}(\boldsymbol{y})$ of (4.3) via $\tilde{u}(\boldsymbol{y}) = u(\boldsymbol{y}) \circ T(\boldsymbol{y})^{-1}$.

Let us motivate Assumptions 4.1, 4.2 for this setting. It is easily checked that due to the affine dependence on y_j, the map $\boldsymbol{y} \mapsto T(\boldsymbol{y})$ is $(\boldsymbol{b}_0, \varepsilon, W^{1,\infty}(\mathrm{D}))$ holomorphic. This and holomorphy of a, c, e, f, g, yield Assumption 4.1 (i) (with "$a(\boldsymbol{y})$" being $dT(\boldsymbol{y})^{-1}(a \circ T(\boldsymbol{y}))dT(\boldsymbol{y})^{-\top} \det dT(\boldsymbol{y})$ etc.), for the details, see [21, Section 5.2]. Next, from (4.9) and similar as in Example 4.3 (cf. (4.8)) we can deduce $\sup_{\boldsymbol{z}\in O_{\boldsymbol{b}_1;\varepsilon}} \|dT(\boldsymbol{y})\|_{\mathcal{W}^{s_1,\infty}(\mathrm{D})} < \infty$. This can be used to verify Assumption 4.2 for $\zeta = 0$ and $s = s_1 \geq 1$: for example $\tilde{f} \circ T(\boldsymbol{y}) \in \mathcal{W}^{s_1}(\mathrm{D})$ and $\det dT_{\boldsymbol{y}} \in \mathcal{W}^{s_1}(\mathrm{D})$ since $\tilde{f}$ is analytic and $dT(\boldsymbol{y}) \in \mathcal{W}^{s_1}(\mathrm{D})$. Thus $(\tilde{f} \circ T_{\boldsymbol{y}}) \det dT_{\boldsymbol{y}} \in \mathcal{W}^{s_1}(\mathrm{D}) \hookrightarrow \mathcal{K}_{-1}^{s_1-1}(\mathrm{D})$ with its norm bounded by $C(\tilde{f})\|T(\boldsymbol{y})\|_{\mathcal{W}^{s_1}(\mathrm{D})}$ for some $\tilde{f}$-dependent constant. Similarly one can treat the other terms. Finally, assuming ellipticity of the coefficient a and appropriate smallness of $\theta > 0$, similar as in Example 4.3 we can verify Assumption 4.1 (ii).

For some regular (triangular) mesh $\mathcal{T}$ on D and $n \in \mathbb{N}$ we define the space $\mathcal{S}^n(\mathcal{T}) \subseteq X$ of continuous piecewise polynomials of degree n on $\mathcal{T}$. It is well-known, that the use of graded meshes allows approximation of functions in the spaces $\mathcal{K}_\zeta^s$ with the optimal rate (see [39, Proposition 5.9]):

Theorem 4.5. *Let $n \in \mathbb{N}$, $n \leq s$. There exist $C > 0$ such that for all $l \in \mathfrak{W} := \{2^j : j \in \mathbb{N}_0\}$ there is a triangular mesh $\mathcal{T}_l$ on D such that the meshwidth behaves as $O(l^{-1/2})$ and $\dim \mathcal{S}^n(\mathcal{T}_l) \leq Cl$ for all $l \in \mathfrak{W}$. Additionally, for each $u \in \mathcal{K}_\zeta^{s+1}(\mathrm{D})$, $l \in \mathfrak{W}$*

$$\inf_{x_l \in \mathcal{S}^n(\mathcal{T}_l)} \|u - x_l\|_{H^1} \leq Cl^{-\frac{n}{2}} \|u\|_{\mathcal{K}_\zeta^{s+1}(\mathrm{D})}. \tag{4.12}$$

With $\mathcal{T}_l$ as in Theorem 4.5, we now fix $n \leq s$ and set $X_l := S^n(\mathcal{T}_l)$. The next corollary is a direct consequence of Proposition 3.5. It states that the FEM solutions satisfy Assumption 2.16. Therefore, the multilevel interpolant/quadrature described in Section 2.2 utilising the finite element solutions, will achieve the convergence rates stated in Theorems 2.31, 2.32.

For the proof, which is based on Taylor gpc expansions of the parametric solution, we require an additional small data assumption. This could be avoided by using Legendre instead of Taylor expansions as we do here, cp. [38, Lemma 5.1] and also Remark 2.3. Moreover, the following assumption is not necessary when considering affine parameter dependence as in Example 4.3, however it becomes relevant when considering the domain transformation model presented in Example

4.4. For simplicity, we do not dwell on this and simply assume the following with $\delta > 0$ to be specified later:

Assumption 4.6. *Let a, b, c, e, f, g be as in Assumption 4.1. It holds*

$$\sup_{\boldsymbol{z}\in O_{\boldsymbol{b}_0,\varepsilon}} \inf_{\boldsymbol{y}\in U} \|a(\boldsymbol{z}) - a(\boldsymbol{y})\|_{L^\infty(D)} < \delta. \tag{4.13}$$

Analogous bounds are satisfied by b, c, e, f, g w.r.t. the spaces of Assumption 4.1.

Corollary 4.7. *Let Assumption 4.1 be satisfied. Then, there exist $\tilde{\varepsilon} > 0$, $C > 0$, $\eta > 0$, $\delta > 0$ and $l_0 > 0$ such that if Assumption 4.2 holds for some $|\zeta| < \eta$ and if Assumption 4.6 is satisfied, we have the following: There is a unique $(\boldsymbol{b}_0, \tilde{\varepsilon}, X)$-holomorphic map $u : U \to X$ and $u(\boldsymbol{y})$ solves (4.3) for all $\boldsymbol{y} \in U$. Furthermore, with $\mathfrak{W}$ as in Theorem 4.5, for all $l \in \mathfrak{W}$, $l \geq l_0$, there is a unique $(\boldsymbol{b}_0, \tilde{\varepsilon}, X)$-holomorphic function $u_l : U \to X$ such that $u_l(\boldsymbol{y}) \in X_l$ solves (4.3) with X replaced by X_l and the extensions satisfy*

$$\sup_{\boldsymbol{z}\in O_{\boldsymbol{b}_0,\tilde{\varepsilon}}} \|u(\boldsymbol{z}) - u_l(\boldsymbol{z})\|_{X_{\mathbb{C}}} \leq C \qquad \text{and} \qquad \sup_{\boldsymbol{z}\in O_{\boldsymbol{b}_1,\varepsilon}} \|u(\boldsymbol{z}) - u_l(\boldsymbol{z})\|_{X_{\mathbb{C}}} \leq C l^{-\alpha} \tag{4.14}$$

with $\alpha := n/2$ and for some C independent of l.

Proof. Set $\xi := (a, c, e, f, \tilde{g}) \in \Xi$ where

$$\Xi := L^\infty(\mathrm{D}, \mathbb{R}^{2\times 2}) \times L^\infty(\mathrm{D}, \mathbb{R}^2) \times L^\infty(\mathrm{D}, \mathbb{R}) \times H^{-1}(\mathrm{D}) \times H^1(\mathrm{D}). \tag{4.15}$$

According to Assumption 4.1, $\xi(\boldsymbol{y})$ is $(\boldsymbol{b}_0, \varepsilon, \Xi)$-holomorphic. Moreover, with $\xi \in \Xi$ as above, we let $\mathcal{A} : X \times \Xi \to X'$ and $\mathcal{N} : X \times \Xi \to X'$ via (cf. (4.3))

$$\mathcal{A}(u, \xi)(v) := \int_{\mathrm{D}} a\nabla u \cdot \nabla v + (c \cdot \nabla u(\boldsymbol{y}))v + e\tilde{u}(\boldsymbol{y})v \quad \text{and} \quad \mathcal{N}(u, \xi)(v) = \mathcal{A}(u, \xi) - \int_{\mathrm{D}} \tilde{f}(\xi)v \tag{4.16}$$

for $v \in X$ and where $\tilde{f}$ is as defined above (4.3). Note that $\partial_1 \mathcal{N}(u, \xi) = \mathcal{A}(\cdot, \xi) \in L(X, X')$.

Set $\tilde{K} := \{\xi(\boldsymbol{y}) : \boldsymbol{y} \in U\} \subseteq \Xi$ and let $K \subseteq \Xi$ be the closed convex hull of $\tilde{K}$. With this notation, we verify the assumptions of Proposition 3.5 item by item:

- Assumption 3.1 (i): The set $\tilde{K}$ is compact as the image of the compact set U under the continuous map $\boldsymbol{y} \mapsto \xi(\boldsymbol{y})$ (continuity holds by definition for $(\boldsymbol{b}_0, \varepsilon)$-holomorphic maps). This implies, that K, as the closed convex hull of $\tilde{K}$ in the Banach space Ξ, is compact as well, see [1, Theorem 5.35]. By Assumption 4.1, the operator $\mathcal{A}(\cdot, \xi) \in L(X, X')$ is elliptic and bounded for all $\xi \in \tilde{K}$ with some ellipticity and continuity constants C_0, C_1. Note that $\xi \mapsto \mathcal{A}(\cdot, \xi)$ is linear. Hence, for any finite convex combination $\xi := \sum_{i=1}^n \lambda_i \xi_i$ with $\xi_i \in \tilde{K}$, it holds that $\mathcal{A}(\cdot, \xi) = \sum_{i=1}^n \lambda_i \mathcal{A}(\cdot, \xi_i)$ is elliptic and bounded with the same constants C_0, C_1. Since K consists of the closure of the set of all convex combinations, the statement is also true for any $\xi \in K$. By the same reason, we have that $\tilde{f}(\xi) \in H^{-1}$ in (4.16) is uniformly bounded for all $\xi \in K$. Thus, for each K the solution $u(\xi)$ of $\mathcal{N}(u(\xi), \xi) = 0$, which by definition is the solution $u(\boldsymbol{y})$ of (4.3), exists and is unique. Continuous dependence of $u(\xi)$ on the data ξ is classical.
- Assumption 3.1 (ii): The map $\mathcal{N} : X \times \Xi \to X'$ is well-defined, affine and bounded in both arguments. Therefore holomorphy, and the existence of a holomorphic extension to $X \times \Xi$ is trivial.

- Assumption 3.4: The space $X = H_0^1(\mathrm{D})$ is reflexive and separable. For $\xi \in K$ fixed, we have $\partial_1 \mathcal{N}(u(\xi), \xi)(\cdot) = \mathcal{A}(\cdot, \xi) \in L(X, X')$. As argued above, this map is uniformly bounded and uniformly boundedly invertible, due to the stated uniform coercivity and boundedness. The same holds for $\mathcal{A} \in L(X_l, Y_l')$ considered on any (finite element) subspace $X_l = Y_l \subseteq X$. Density of $\bigcup_{l \in \mathbb{N}} X^n(\mathcal{T}_l)$ in $X = H_0^1(\mathrm{D})$ is again well known, since the meshwidth of the mesh $\mathcal{T}_l$ tends to zero as $l \to \infty$.

Applying Proposition 3.5, gives an open set $O \subseteq \Xi_{\mathbb{C}}$ containing $\tilde{K}$ as well as holomorphic extensions of u, u_l on O for $l \geq l_0$. Choosing $\delta > 0$ in Assumption 4.6 small enough, we can ensure $\xi(\boldsymbol{z})$ to be in O for all $\boldsymbol{z} \in O_{\boldsymbol{b}_0,\varepsilon} \supseteq O_{\boldsymbol{b}_1,\varepsilon}$, with $O_{\boldsymbol{b}_i,\varepsilon}$ defined in (2.2).

Setting $u(\boldsymbol{z}) := u(\xi(\boldsymbol{z}))$ and $u_l(\boldsymbol{z}) := u_l(\xi(\boldsymbol{z}))$ gives uniformly bounded extensions in $X_{\mathbb{C}}$. Since $\xi(\boldsymbol{y})$ is $(\boldsymbol{b}_0, \varepsilon)$-holomorphic, u and u_l are $(\boldsymbol{b}_0, \varepsilon, X)$-holomorphic, due to the fact that the composition of holomorphic functions is again holomorphic. Thus $\|u(\boldsymbol{z})\|_X$ is uniformly bounded for $\boldsymbol{z} \in O_{\boldsymbol{b}_0,\varepsilon}$, and by (3.14) the same holds for $\|u_l(\boldsymbol{z})\|_X$ with a constant independent of $l \geq l_0$, showing the first inequality in (4.14).

It remains to choose ζ and show the second inequality in (4.14). In the following we will use that by Assumption 4.2 all $\xi \in \{\xi(\boldsymbol{z}) : \boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}\}$ (which is a superset of $\tilde{K}$) lie in the smoother (as compared to Ξ) spaces of Assumption 4.2, and their norms are uniformly bounded according to (4.6). Hence, for every $\xi \in \tilde{K}$ according to [8, Theorem 1.1], there exists $\eta_\xi > 0$, such that for all $|\zeta| \leq \eta_\xi$ the map $\mathcal{N}(\cdot, \xi) : \mathcal{K}^{s+1}_{\zeta+1}(\mathrm{D}) \to \mathcal{K}^{s-1}_{\zeta-1}(\mathrm{D})$ is an isomorphism. Inspection of the proof reveals that η_ξ depends continuously on the data $\xi \in \Xi$ (see dependence of γ_1, γ_2 on β in [8, Remark 4.3]). Using compactness of K and once more a covering argument, by decreasing $\delta > 0$ if necessary, we can find $\eta > 0$ such that for all $|\zeta| < \eta$ it holds that $\mathcal{A}(\cdot, \xi) \in L(\mathcal{K}^{s+1}_{\zeta+1}(\mathrm{D}), \mathcal{K}^{s-1}_{\zeta-1}(\mathrm{D}))$ is an isomorphism and such that the norm of $\mathcal{A}(\cdot, \xi)^{-1} \in L(\mathcal{K}^{s-1}_{\zeta-1}(\mathrm{D}), \mathcal{K}^{s+1}_{\zeta+1}(\mathrm{D}))$ is uniformly bounded for all $\xi \in \{\xi(\boldsymbol{z}) : \boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}\}$. For such ζ, we can now conclude that, due to the uniform boundedness of the data in the sense of Assumption 4.2 (in particular $\tilde{f}(\xi) \in \mathcal{K}^{s-1}_{\zeta-1}$), the parametric solution $u(\xi) = \mathcal{A}(\cdot, \xi)^{-1}(\tilde{f}(\xi))$ belongs to the weighted space $\mathcal{K}^{s+1}_{\zeta+1}(\mathrm{D})$, and its norm is uniformly bounded for all elements in $\{\xi(\boldsymbol{z}) : \boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}\}$. Thus (3.14) and Theorem 4.5 give the second inequality in (4.14). □

Remark 4.8. *The results in [8] do not merely show the solution to be in the Kondratiev space. In particular, they also prove holomorphic dependence of the solution on the data in the Kondratiev spaces. Whereas closely connected to our analysis, we did not employ this holomorphy, as it is not necessary for our line of arguments.*

4.1.2 Cartesian product domains

It is well-known that sparse-grid spaces allow to approximate functions of mixed Sobolev regularity on a d-dimensional cube with an algebraic rate independent of d (up to logarithms), see for instance [53, 50, 9, 29]. In a similar spirit as to what has been done in [33, 32], utilising such FEM spaces further decreases the asymptotic complexity of the algorithm. To illustrate this, for $s \in \mathbb{N}$ and γ, $\delta > 0$ as well as $\mathrm{D} := [-1, 1]^2$ set

$$\|u\|^2_{H^{s,s}_{\gamma,\delta}(\mathrm{D})} := \sum_{i,j \leq s} \int_{\mathrm{D}} |(1 - x_1^2)^\gamma (1 - x_2^2)^\delta \partial_1^i \partial_2^j u(x_1, x_2)|^2 \mathrm{d}x, \tag{4.17}$$

where $x = (x_1, x_2)$ and ∂_i^j refers to the jth derivative of the ith coordinate, $j \in \mathbb{N}_0$, $i \in \{1, 2\}$. We then introduce the weighted anisotropic Sobolev spaces

$$H^{s,s}_{\gamma,\delta}(\mathrm{D}) := \{u \in L^2(\mathrm{D}) \,:\, \|u\|_{H^{s,s}_{\gamma,\delta}(\mathrm{D})} < \infty\}. \tag{4.18}$$

Theorem 4.9 ([42, Theorem 1]). *Let n, $s \in \mathbb{N}$, $n \le s$ and γ, $\delta > 0$. For $l \in \mathfrak{W} := \{2^{nj} \,:\, j \in \mathbb{N}_0\}$, there exist spaces $X_l \subseteq X$ with $\dim(X_l) \le l$ and a constant C such that for all $u \in H^{s+1,s+1}_{\gamma,\delta}(\mathrm{D})$*

$$\inf_{x_l \in X_l} \|u - x_l\|_{H^1} \le C l^{-n} \log(l)^{3/2} \|u\|_{H^{s+1,s+1}_{\gamma,\delta}(\mathrm{D})}. \tag{4.19}$$

We do not recall the precise definition of X_l in Theorem 4.9, but merely mention that they are sparse-grid wavelet spaces. Note that for $\gamma > \min\{0, (\tilde{s} + 1 - \zeta)/2\}$ there holds the continuous embedding

$$\mathcal{K}^{2(\tilde{s}+1)}_{\zeta}(\mathrm{D}) \hookrightarrow H^{\tilde{s}+1,\tilde{s}+1}_{\gamma,\gamma}(\mathrm{D}). \tag{4.20}$$

We emphasize that the space on the left-hand side has much more regularity. In this sense, the next result does not aim at utmost generality concerning the smoothness of the PDE coefficients. The proof exploits Theorem 4.9 and (4.20) instead of Theorem 4.5, but is apart from that in complete analogy to the one of Corollary 4.7.

Corollary 4.10. *Let n, $\tilde{s} \in \mathbb{N}$, $n \le \tilde{s}$ and $\mathfrak{W}$ be as in Theorem 4.9. Let Assumption 4.1 be satisfied. Then, there exist $\tilde{\varepsilon} > 0$, $C > 0$, $\eta > 0$, $\delta > 0$ and $l_0 > 0$ such that if Assumption 4.2 holds with $s := 2\tilde{s} + 1$ and $|\zeta| \le \eta$, and if Assumption 4.6 is satisfied with δ, we have the following: For any $\epsilon > 0$, there is a unique $(\boldsymbol{b}_0, \tilde{\varepsilon}, X)$-holomorphic map $u : U \to X$ and $u(\boldsymbol{y})$ solves (4.3) for all $\boldsymbol{y} \in U$. Furthermore, for all $l \in \mathfrak{W}$, $l \ge l_0$, there is a unique $(\boldsymbol{b}_0, \tilde{\varepsilon}, X)$-holomorphic function $u_l : U \to X$ such that $u_l(\boldsymbol{y}) \in X_l$ solves (4.3) with X replaced by X_l and the extensions satisfy (4.14) with $\alpha := n - \epsilon$ and for some C independent of l.*

5 Numerical Experiments

We now report on numerical experiments and observed convergence rates for the presented multilevel algorithm. In Section 5.2 multilevel convergence is compared to single level convergence for a real valued test function. In Sections 5.3, 5.4 we test the algorithm for a diffusion problem with parametric diffusion coefficient in one dimension, and for a diffusion problem in two dimensions on a varying domain, respectively.

5.1 Implementation

Before presenting our results we comment on a few aspects of the implementation.

5.1.1 Choice of discretization levels

Let $\mathbf{m}_0 = (m_{0;\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}} \in \ell^{p_0}_{\mathrm{m}}(\mathcal{F})$, $\mathbf{m}_1 = (m_{1;\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}} \in \ell^{p_1}_{\mathrm{m}}(\mathcal{F})$ be two sequences satisfying (2.48) and with the property that for all $\boldsymbol{\nu} \in \mathcal{F}$

$$\|\partial^{\boldsymbol{\nu}} u(\boldsymbol{y})/\boldsymbol{\nu}!|_{\boldsymbol{y}=\mathbf{0}}\|_X \le C m_{0;\boldsymbol{\nu}} \qquad \text{and} \qquad \|\partial^{\boldsymbol{\nu}} (u(\boldsymbol{y}) - u_l(\boldsymbol{y}))/\boldsymbol{\nu}!|_{\boldsymbol{y}=\mathbf{0}}\|_X \le C(l+1)^{-\alpha} m_{1;\boldsymbol{\nu}} \tag{5.1}$$

where u, u_l are as in Assumption 2.16.

From Theorem 2.6 and the proofs of Theorems 2.21, 2.22, we know that the levels $w_{\boldsymbol{\nu}}$ should reflect the behaviour $w_{\boldsymbol{\nu}} \sim m_{1;\boldsymbol{\nu}}^{1/(\alpha+1)}$. Whereas Theorem 2.6 gives an explicit constant C_ε such that $w_{\boldsymbol{\nu}} = \lceil C_\varepsilon m_{1;\boldsymbol{\nu}}^{1/(\alpha+1)} - 1\rceil_{\mathfrak{W}}$ (cp. (2.8) with $q = 1$), the computation of C_ε requires the computation of the N^β largest $(m_{1;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ (cf. Theorem 2.6), but does not guarantee $w_{\boldsymbol{\nu}} > 0$ for these multiindices. In practice, if we determine the N^β largest values of the estimator $(m_{1;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$, we also wish to use all of the corresponding multiindices in the construction of our interpolation or quadrature operator. In other words, for each estimator $m_{1;\boldsymbol{\nu}}$ which we have found to be among the largest, we want to set its discretization level $w_{\boldsymbol{\nu}}$ to a positive number, so as not to have computed $m_{1;\boldsymbol{\nu}}$ in vain. The purpose of the next lemma is to circumvent this and construct an allocation of positive discretization levels which will result in optimal rates. The proof is given in the appendix.

Lemma 5.1. *Let* $\mathbf{m}_i = (m_{i;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$, $i \in \{0, 1\}$ *be two monotonically decreasing sequences, and let* $\mathfrak{W}$ *be as in Assumption 2.4. Denote by* $(m_{i;j})_{j\in\mathbb{N}}$, $i \in \{0, 1\}$, *two decreasing rearrangements of* $\mathbf{m}_i$ *over* $\mathbb{N}$. *Let* $0 < r_1 \le r_0$ *and assume that for every* $\delta > 0$ *there exist constants* C_1, C_2 *s.t. for* $i \in \{0, 1\}$

$$C_1 j^{-r_i} \le m_{i;j} \le C_2 j^{-r_i+\delta} \qquad \forall\ j \in \mathbb{N}. \tag{5.2}$$

Set $\tilde{m}_{1;\boldsymbol{\nu}} := \max\{m_{0;\boldsymbol{\nu}}^{r_1/r_0}, m_{1;\boldsymbol{\nu}}\}$, $\tilde{m}_{0;\boldsymbol{\nu}} := \max\{m_{0;\boldsymbol{\nu}}, m_{1;\boldsymbol{\nu}}^{r_0/r_1}\}$. *For every* $\varepsilon > 0$ *define* $\Lambda(\varepsilon) := \{\boldsymbol{\nu} \in \mathcal{F} : \tilde{m}_{1;\boldsymbol{\nu}} \ge \varepsilon\}$. *Define* $\mathbf{w}_\varepsilon = (w_{\varepsilon;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ *by* $w_{\varepsilon;\boldsymbol{\nu}} = 0$ *if* $\boldsymbol{\nu} \in \Lambda(\varepsilon)^c$ *and set for* $\boldsymbol{\nu} \in \Lambda(\varepsilon)$

$$w_{\varepsilon;\boldsymbol{\nu}} := \left\lceil C_\varepsilon \tilde{m}_{1;\boldsymbol{\nu}}^{\frac{1}{\alpha+1}} \right\rceil_{\mathfrak{W}} \qquad \text{where} \qquad C_\varepsilon := \begin{cases} \max_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} \tilde{m}_{1;\boldsymbol{\nu}}^{\frac{1}{\alpha}-\frac{1}{\alpha+1}} / \tilde{m}_{0;\boldsymbol{\nu}}^{1/\alpha} & \text{if } r_1 - 1 < \alpha, \\ \max_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} \tilde{m}_{1;\boldsymbol{\nu}}^{\frac{-1}{\alpha+1}} & \text{otherwise.} \end{cases} \tag{5.3}$$

Fix $\delta > 0$ *arbitrarily small. There then exists a constant* C *such that for all* $\varepsilon > 0$

$$\sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} (w_{\varepsilon;\boldsymbol{\nu}} + 1)^{-\alpha} m_{1;\boldsymbol{\nu}} + \sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)^c} m_{0;\boldsymbol{\nu}} \le C|\mathbf{w}_\varepsilon|^{-r}, \tag{5.4}$$

where $r = (1 - r_0)\beta - \delta$, $\beta = \alpha/(\alpha + r_0 - r_1)$, *if* $r_1 - 1 \le \alpha$, *and* $r = \alpha - \delta$ *otherwise.*

Remark 5.2. *The choice* $C_\varepsilon = \max_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} \tilde{m}_{1;\boldsymbol{\nu}}^{\frac{1}{\alpha}-\frac{1}{\alpha+1}} / \tilde{m}_{0;\boldsymbol{\nu}}^{1/\alpha} = \max_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} \tilde{m}_{1;\boldsymbol{\nu}}^{\frac{1}{\alpha}-\frac{1}{\alpha+1}-\frac{r_0}{\alpha r_1}}$ *in the first case is exactly such that* $w_{\varepsilon;\boldsymbol{\nu}}^{-\alpha}\tilde{m}_{1;\boldsymbol{\nu}} \le \tilde{m}_{0;\boldsymbol{\nu}}$ *for all* $\boldsymbol{\nu} \in \Lambda(\varepsilon)$. *It is intuitively clear that this must be satisfied, since* $\mathbf{w}_\varepsilon$ *should ideally minimize the left-hand side of* (5.4) *while also minimizing* $|\mathbf{w}_\varepsilon|$. *In the second case* C_ε *is such that* $w_{\varepsilon;\boldsymbol{\nu}} \ge 1$ *for all* $\boldsymbol{\nu} \in \Lambda(\varepsilon)$.

For our algorithm, we proceed as suggested by the lemma: First, a downward closed set $\Lambda = \{\boldsymbol{\nu} \in \mathcal{F} : \tilde{m}_{1;\boldsymbol{\nu}} \ge \varepsilon\}$ is determined for some fixed $\varepsilon > 0$, and then we choose the levels for $\boldsymbol{\nu} \in \Lambda$ according to (5.3). The occurring set $\mathfrak{W}$ in the lemma depends on the problem and the numerical solver (e.g., each element of $\mathfrak{W}$ corresponds to the number of degrees of freedom of a FEM solution on some available mesh).

In the case of quadrature we additionally have to take into account item (iii) of Lemma 2.15. To this end the above construction is adjusted by replacing $\mathbf{m}_i$ with $\tilde{\mathbf{m}}_i$ defined as

$$\tilde{m}_{i;\boldsymbol{\nu}} = m_{i;\lceil\boldsymbol{\nu}\rceil} \qquad \text{where} \qquad \lceil\boldsymbol{\nu}\rceil_j = \begin{cases} 0 & \text{if } \nu_j = 0, \\ \max\{2, \nu_j\} & \text{otherwise,} \end{cases} \tag{5.5}$$

for $i = \{0, 1\}$. We refer to [52] for more details on this.

Finally, one needs to determine sequences $\mathbf{m}_0$, $\mathbf{m}_1$ as stated above. With $\mathfrak{I}$ as in Assumptions 2.18, 2.24, assume $2 \in \mathfrak{I}$ in the following, so that the ensuing definition will not interfere with (5.5). Using estimates as presented in [18, 19] and [52] gives $m_{i;\boldsymbol{\nu}} = C(|\boldsymbol{\nu}_{\mathfrak{I}}|!/\boldsymbol{\nu}_{\mathfrak{I}}!)\boldsymbol{\rho}_i^{-\boldsymbol{\nu}_{\mathfrak{I}}}$ with $\boldsymbol{\nu}_{\mathfrak{I}} := \lfloor \boldsymbol{\nu} \rfloor_{\mathfrak{I}}$ for some suitable sequence $\boldsymbol{\rho}_i = (\rho_{i;j})_{j\in\mathbb{N}}$ with $\rho_{i;j} > 1$ for all $j \in \mathbb{N}$ and $\rho_{i;j}^{-1} \in \ell^{p_i}$ for $i \in \{0, 1\}$ (note that such $(m_{i;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ is not monotone, but employing estimates from [15] this could be avoided by using a slightly more involved formula). From the mentioned proofs one obtains $\rho_j^{-1} \sim b_j$. In practice, setting $m_{i;\boldsymbol{\nu}} = \boldsymbol{\rho}_i^{-\boldsymbol{\nu}_{\mathfrak{I}}}$ appears to perform better however. For this reason all of the following computations were done based on such sequences. Ultimately, let us note that the requirement (5.2) might seem restrictive. However, if $\rho_j \sim j^r$ for some $r > 1$, then the above sequences satisfy this assumption: For example, by [19, Lemma 7.1] it holds $(\boldsymbol{\rho}^{-\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^{1/r+\delta}$, so that Lemma A.5 implies decay $j^{-r+\delta}$ of a monotonically decreasing rearrangement $(t_j)_{j\in\mathbb{N}}$ of $(\boldsymbol{\rho}^{-\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$. On the other hand $\rho_j^{-1} \sim j^{-r}$ for $j \in \mathbb{N}$ is a subsequence of $(\boldsymbol{\rho}^{-\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$, so that the lower bound in (5.2) is satisfied. The same argument can be applied to $|\boldsymbol{\nu}|!/\boldsymbol{\nu}!\boldsymbol{\rho}^{-\boldsymbol{\nu}}$. Lemma 5.1 is thus precisely targeted at the kind of problems we are interested in.

5.1.2 Work measure

Let $\mathbf{w} = (w_j)_{j\in\mathbb{N}} \in \mathfrak{W}^{\mathcal{F}}$ and let $\Lambda_{\mathfrak{w}_i}$ be as in (2.40) depending on $\mathbf{w}$ and where $\mathfrak{W} = \{\mathfrak{w}_i : i \in \mathbb{N}_0\}$ with $\mathfrak{w}_0 = 0$. As a measure of the work, we then use the quantity

$$\mathrm{work}(\mathbf{w}) := \sum_{i\in\mathbb{N}} \mathfrak{w}_i \cdot \Big(\text{number of interpolation points employed by } (I_{\Lambda_{\mathfrak{w}_i}} - I_{\Lambda_{\mathfrak{w}_{i+1}}})\Big), \qquad (5.6)$$

for interpolation, and with $(I_{\Lambda_{\mathfrak{w}_i}} - I_{\Lambda_{\mathfrak{w}_{i+1}}})$ replaced by $(Q_{\Lambda_{\mathfrak{w}_i}} - Q_{\Lambda_{\mathfrak{w}_{i+1}}})$ for quadrature (cp. (2.42)).

This is a simplification of the work model employed in Section 2.3, in that it only takes into account item (i) but not item (ii) described at the beginning of Section 2.3: we merely consider the complexity of evaluating the function for all levels at each required interpolation/quadrature point, but not the complexity of evaluating $I_{\mathbf{w}}$, $Q_{\mathbf{w}}$ given the function values at the interpolation/quadrature points. We do so, since (ii) can be considered negligible for a moderate number of interpolation/quadrature points if each evaluation of the integrand is costly (as is the case if the integrand is the solution to some PDE).

For the examples in Sections 5.3 and 5.4, $\mathfrak{w}_i$ will be the number of degrees of freedom of a FEM solution to a diffusion problem. If the complexity of computing this solution is proportional to $\mathfrak{w}_i$, then (5.6) measures the work to determine all FEM approximations. Due to the sparsity of the obtained stiffness matrices, this is a reasonable assumption. In other words, the work quantity then amounts to the total number of degrees of freedom of all required FEM solutions.

5.1.3 Interpolation/Quadrature points

Using the terminology of [14], an $\Re$-*Leja sequence* is the projection of a Leja sequence defined on the unit circle onto $[-1, 1]$. For all experiments, as interpolation points we employ sections of such a sequence as described in [14]: for an $\Re$-Leja sequence $(\chi_j)_{j\in\mathbb{N}_0} \subseteq [-1, 1]$, the points $(\chi_{n;j})_{j=0}^n$ introduced at the beginning of Section 2.2.1 are chosen as $\chi_{n;j} = \chi_j$ for all $j = 0, \ldots, n$ and for all $n \in \mathbb{N}_0$. In particular, they are nested in the sense that $(\chi_{n;j})_{j=0}^n \subseteq (\chi_{m;j})_{j=0}^m$ for every $m \geq n$. It is known that these points satisfy (2.31) with $\tau = 3$, see [11, 10, 14]. The set $\mathfrak{I}$ in Assumptions 2.18, 2.24 is then chosen as $\mathfrak{I} = \mathbb{N}$, and hence does *not satisfy* the stated assumptions. This does not affect the convergence rate in terms of the work measure in Section 5.1.2 though, cp. Remark 2.33.

For the quadrature algorithm we additionally use (non-nested) Gauss-Legendre quadrature with $\mathfrak{I}$ in Assumption 2.18, 2.24 defined as $\mathfrak{I} = \{0\} \cup \{2^j : j \in \mathbb{N}\}$, cp. Remark 2.23. The choice of points is indicated in the plots by the label "leja" or "gauss".

5.2 Scalar parametric test function

In this section let $X = \mathbb{R}$. For $\theta > 0$, $r > 1$ consider $u : U \to X$ via

$$u(\boldsymbol{y}) := \frac{1}{1 + \theta \sum_{j\in\mathbb{N}} y_j j^{-r}}, \tag{5.7}$$

which is well-defined for all $\boldsymbol{y} \in U$ in case θ is so small that $\theta \sum_{j\in\mathbb{N}} j^{-r} < 1$. One can show that this function is $(\boldsymbol{b}, \varepsilon, \mathbb{C})$-holomorphic for some $\boldsymbol{b} \in \ell^p$ and any $p > 1/r$, see [52, Example 4.1]. For $l \in \mathbb{N}$ we now define

$$u_l(\boldsymbol{y}) := u(\boldsymbol{y}) + R_l \cdot l^{-\alpha} \tag{5.8}$$

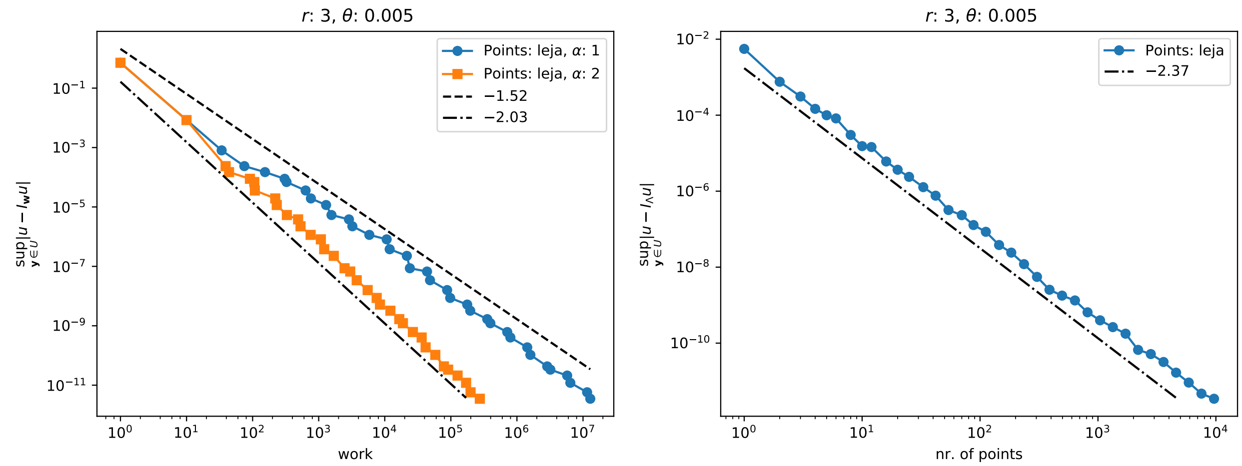
where $R_l \in [-2, -1] \cup [1, 2]$ denotes a randomly chosen (w.r.t. the uniform distribution) number independent of $\boldsymbol{y}$. Equation (5.8) artificially introduces the error $R_l \cdot l^{-\alpha}$, so that u_l can be considered as an approximation to u at level l. In particular u_l converges to u at rate $\alpha > 0$ (uniformly on U), which provides us with a simple test setting for the multilevel algorithm. Clearly, $(\boldsymbol{b}, \varepsilon)$-holomorphy of u implies the same for u_l and consequently Assumption 2.16 is satisfied with $p_0 = p_1 = p$ for any $p > 1/r$. In view of Theorems 2.31, 2.32 we thus expect the convergence rates

$$\min\{\alpha, r-1\} - \delta \qquad \text{and} \qquad \min\{\alpha, 2r-1\} - \delta \qquad \text{with } \delta > 0 \text{ arbitrarily small} \tag{5.9}$$

for multilevel interpolation/quadrature respectively w.r.t. the work quantity defined in (5.6). Throughout what follows, *if we speak of (proven) convergence rates, they are always understood up to some arbitrarily small $\delta > 0$, and we will not mention this anymore.* We point out that in practice the observed rates for such examples may depend strongly on θ due to the presence of a large preasymptotic range. For more details we refer to [52, Section 4.2].

In Figure 1 the convergence of the multilevel interpolant for $r = 3$, $\theta = 0.005$ and $\alpha \in \{2, 3\}$ is compared with *single-level interpolation*, by which we mean Smolyak interpolation based on the exact function values of u (this corresponds to the case "$\alpha = \infty$"), see for example [13, 52]. For single-level interpolation we plot the error vs. the number of interpolation points, whereas in the multilevel case, the x-axis shows the work measure defined in (5.6). The single-level interpolant in Fig. 1 (b) exceeds the predicted rate $r - 1 = 2$ (see, e.g. [13]). For $\alpha = 2$, the multilevel interpolant achieves the rate 2 stated in (5.9). Letting $\alpha = 1$, as expected the observed rate decreases, but it still exceeds the proven rate of 1 in this example. For both $\alpha \in \{1, 2\}$, the observed convergence rate of the single-level approximation (in terms of nr. of points) is better than the multilevel rate (in terms of the work). This is to some extent expected, since the multilevel algorithm additionally has to take care of the error introduced by approximating u with u_l.

Figure 2 shows the same for the quadrature algorithm with $r = 2$, $\theta = 0.05$ and employing either Gauss-Legendre quadrature or $\Re$-Leja quadrature points, cp. Section 5.1.3. The convergence rate of the single level quadrature is less than $2r - 1 = 3$. This can be attributed to the aforementioned preasymptotic behaviour for "large" values of $\theta > 0$. Subfigure (a) shows that the multilevel quadrature performs almost equally well for $\alpha = 3$. Note that the rate of the single-level method



(a) multilevel interpolation (b) single-level interpolation

Figure 1: Multilevel vs. single-level interpolation for the test function in (5.7) with $r = 3$, $\theta = 0.005$. The "worst case" error $\sup_{\boldsymbol{y} \in U} |u(\boldsymbol{y}) - I_{\mathbf{w}} u(\boldsymbol{y})|$ is numerically estimated by taking the maximum of $|u(\boldsymbol{y}) - I_{\mathbf{w}} u(\boldsymbol{y})|$ at 144 random points $\boldsymbol{y} \in U$. The proven rate is $\min\{3-1, \alpha\} = \alpha$ for the multilevel algorithm and $3 - 1 = 2$ for the single-level algorithm.

can be considered to be an upper bound of what the multilevel method can achieve. For $\alpha = 2$ we observe again the predicted rate 2. As for interpolation, in this example, the single-level rate is better than the multilevel rate. In all cases the difference between Gauss-Legendre quadrature and $\Re$-Leja quadrature is marginal.

5.3 1D model diffusion problem

Set $\mathrm{D} := [-1, 1]$ and let $u(\boldsymbol{y}) \in H_0^1(\mathrm{D})$ be the solution to

$$-\frac{d}{dx}\left(a(\boldsymbol{y})\frac{d}{dx}u(\boldsymbol{y})\right) = 1 \qquad \text{in D}, \tag{5.10a}$$

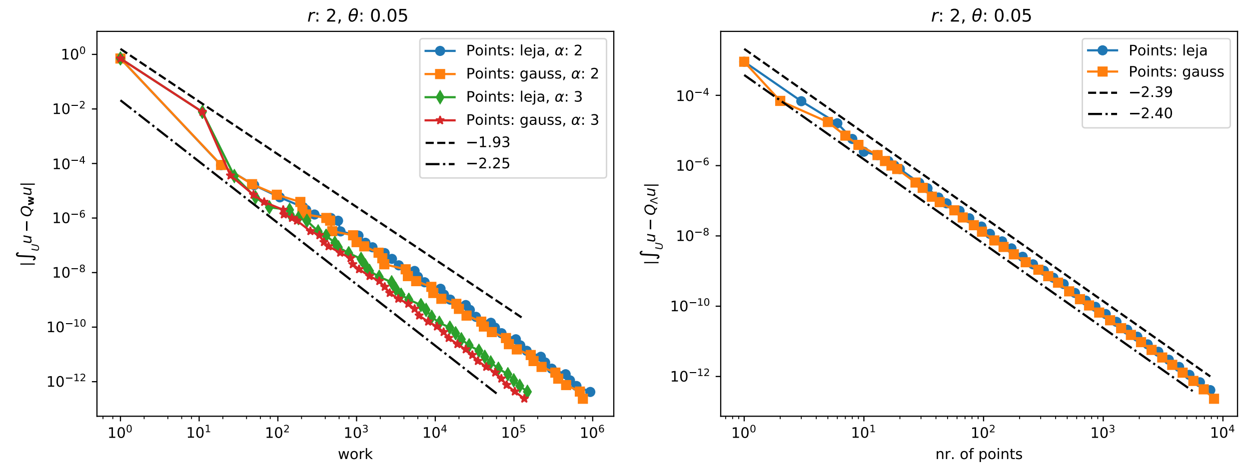
$$u = 0 \qquad \text{on } \partial\mathrm{D}, \tag{5.10b}$$

with the diffusion coefficient $a(\boldsymbol{y})(x) := 1 + \theta \sum_{j \in \mathbb{N}} y_j j^{-r} \sin(j\pi x)$ for $x \in \mathrm{D}$ and some $r > 1$, $\theta > 0$.

5.3.1 Theoretically predicted convergence rates

We approximate the solution $u(\boldsymbol{y})$ to (5.10) with continuous, piecewise linear finite elements: for each $l \in \mathfrak{W} := \{2^j + 1 : j \in \mathbb{N}\}$ denote by $u_l \in H_0^1(\mathrm{D})$ the fem solution w.r.t. a uniform mesh exhibiting l equidistant nodes on $[-1, 1]$. Let us sketch the verification of Assumption 2.16.

Let $\theta > 0$, $\varepsilon > 0$ so small that $\varepsilon + \theta \sum_{j \in \mathbb{N}} j^{-r} < 1$. Then with $b_{0;j} := \theta j^{-r}$, if $\sum_{j \in \mathbb{N}} (\rho_j - 1) b_{0;j} < \varepsilon$ and $\boldsymbol{z} \in O_\varepsilon$ s.t. $|z_j| \le \rho_j$, we have $\Re(a(\boldsymbol{z})) \ge 1 - \sum_{j \in \mathbb{N}} b_j \rho_j > 1 - \varepsilon > 0$. Thus $u(\boldsymbol{z}) \in H_0^1(\mathrm{D}, \mathbb{C})$ is well-defined, and one easily obtains $(\boldsymbol{b}_0, \varepsilon)$-holomorphy of $u(\boldsymbol{y})$ as well as of $u_l(\boldsymbol{y})$, with similar arguments as in Section 4. Next, for $0 < s < r - 1$, $s \in \mathbb{N}$, set $b_{1;j} := \theta j^{-r+s}$, let $\boldsymbol{\rho}$ such that



(a) multilevel quadrature (b) single-level quadrature

Figure 2: Multilevel vs. single-level quadrature for the test function in (5.7) with $r = 2$ and $\theta = 0.05$. The proven rate is $\min\{4-1, \alpha\} = \alpha$ for the multilevel algorithm and $4 - 1 = 3$ for the single-level algorithm.

$\sum_{j\in\mathbb{N}}(\rho_j - 1)b_{1;j} < 1$ and let $\boldsymbol{z}$ such that $|z_j| \le \rho_j$ for all $j \in \mathbb{N}$. Then

$$\|a(\boldsymbol{z})\|_{W^{s,\infty}(\mathrm{D})} \le \sum_{j\in\mathbb{N}} \rho_j \|\theta j^{-r}\sin(j\pi x)\|_{W^{s,\infty}(\mathrm{D})}|z_j| \le C\sum_{j\in\mathbb{N}} j^{-r+s}\rho_j \le C b_{1;j}\rho_j \le C\sum_{j\in\mathbb{N}} j^{-r+s} + 1. \tag{5.11}$$

Hence we have a uniform bound on $\|a(\boldsymbol{z})\|_{W^{s,\infty}(\mathrm{D})}$ for all $\boldsymbol{z} \in O_{\boldsymbol{b}_{1};\varepsilon}$ (cp. (2.2)). By standard regularity theory, this yields a uniform bound on $\|u(\boldsymbol{z})\|_{H^{1+s}}$ for all $\boldsymbol{z} \in O_{\boldsymbol{b}_1;\varepsilon}$. By standard finite element theory [7], for $s \ge 1$ we obtain

$$\sup_{\boldsymbol{z}\in O_{\boldsymbol{b}_1,\varepsilon}} \|u(\boldsymbol{z}) - u_l(\boldsymbol{z})\|_{H^1(\mathrm{D})} \le Cl^{-1} \qquad \text{and} \qquad \sup_{\boldsymbol{z}\in O_{\boldsymbol{b}_1,\varepsilon}} \|u(\boldsymbol{z}) - u_l(\boldsymbol{z})\|_{L^2(\mathrm{D})} \le Cl^{-2}, \tag{5.12}$$

where the second bound is obtained using the so-called "Aubin-Nitsche duality argument". Since $(\boldsymbol{b}, \varepsilon, H^1)$-holomorphy evidently implies $(\boldsymbol{b}, \varepsilon, L^2)$-holomorphy, we have verified Assumption 2.16 with $\alpha = 2$ in the case of $X = L^2(\mathrm{D})$ and $\alpha = 1$ for $X = H^1(\mathrm{D})$.

We will measure the error in $L^2(\mathrm{D})$, i.e. with $\alpha = 2$. Since we use linear finite elements, it suffices to choose $s = 1$, which yields $\boldsymbol{b}_0 \in \ell^{1/r+\delta}$ and $\boldsymbol{b}_1 \in \ell^{1/(r-1)+\delta}$ for any $\delta > 0$. With $p_0 = 1/r + \delta$ and $p_1 = 1/(r-1) + \delta$ Theorems 2.31, 2.32 therefore suggest the convergence rates (up to some arbitrarily small $\delta > 0$)

$$\begin{cases} 2 & \text{if } 4 \le r, \\ 2\frac{r-1}{3} & \text{otherwise}, \end{cases} \qquad \text{and} \qquad \begin{cases} 2 & \text{if } 5 \le 2r, \\ 2\frac{2r-1}{4} & \text{otherwise}, \end{cases} \tag{5.13}$$

for multilevel interpolation/quadrature in $L^2(\mathrm{D})$.

Remark 5.3. *More general than above, with* $\mathrm{D} = (-1, 1)^d$, $d \in \mathbb{N}$, *consider the diffusion problem* $-\operatorname{div}(a(\boldsymbol{y})u(\boldsymbol{y})) = 1$ *with boundary condition* $u|_{\partial\mathrm{D}} = 0$. *Suppose further that* $a(\boldsymbol{y}) = \overline{a} + \sum_{j\in\mathbb{N}} y_j\psi_j$

is such that $(\|\psi_j\|_{L^\infty(\mathrm{D})})_{j\in\mathbb{N}} \in \ell^{p_0}$ *and* $(\|\psi_j\|_{W^{s,\infty}(\mathrm{D})})_{j\in\mathbb{N}} \in \ell^{p_1}$ *with* $p_1 = 1/(1/p_0 - s)$ *(as is the case if e.g.* $\psi_j(x) = \sin(j\pi x_1)$*). Additionally we assume uniform ellipticity of* a *for all* $\boldsymbol{z} \in O_{\boldsymbol{b}_0;\varepsilon}$ *for some* $\varepsilon > 0$ *and where* $b_{0;j} \sim \|\psi_j\|_{L^\infty(\mathrm{D})}$*. With* $b_{1;j} \sim \|\psi_j\|_{W^{s,\infty}(\mathrm{D})}$ *and for* $\boldsymbol{z} \in O_{\boldsymbol{b}_1,\varepsilon}$*, since* $a(\boldsymbol{z}) \in W^{s,\infty}$*, the solution* $u(\boldsymbol{z})$ *is in a weighted Sobolev space with smoothness index* $s+1$*. Using (a result like) Theorem 4.5, finite elements of polynomial degree* n *on appropriately graded meshes achieve the FEM convergence rate* $\alpha = \min\{n, s+1-l\}/d$ *w.r.t.* $\|\cdot\|_{H^l(\mathrm{D})}$*,* $l \in \mathbb{N}_0$*, and where* $s = 1/p_0 - 1/p_1$ *such that* $s + 1 - l \geq 1$*.*

The $H^l(\mathrm{D})$ *convergence rate for the interpolant in Theorem 2.31 becomes*

$$\min\left\{\alpha, \frac{\alpha}{\alpha + (p_0^{-1} - (p_0^{-1} - s))}(p_0^{-1} - 1)\right\} = \min\left\{\alpha, \frac{\alpha}{\alpha + s}(p_0^{-1} - 1)\right\}, \tag{5.14}$$

which holds because

$$\alpha \frac{p_0^{-1} - 1}{\alpha + p_0^{-1} - p_1^{-1}} \leq \alpha \Leftrightarrow p_1^{-1} - 1 \leq \alpha. \tag{5.15}$$

To maximize (5.14)*, (unsurprisingly)* α *should be possibly large. We choose the polynomial degree* $n = s + 1 - l$*, which yields the rate*

$$\min\left\{\frac{s+1-l}{d}, \frac{s+1-l}{s+1-l+ds}(p_0^{-1} - 1)\right\}. \tag{5.16}$$

For $l = 1$ *this is* $\min\{s/d, (p_0^{-1} - 1)/(1+d)\}$*. The optimum rate the multilevel interpolation can achieve w.r.t. the* H^1*-norm is thus only* $1/(1+d)$ *times the rate achieved by single-level interpolation, which is* $p_0^{-1} - 1$ *(i.e. by interpolation assuming that* $u \in H^1(\mathrm{D})$ *can be evaluated exactly with complexity* $O(1)$*; the rate* $p_0^{-1} - 1$ *is then due to* $(\boldsymbol{b}_0, \varepsilon, H_0^1)$*-holomorphy of* u*). On the other hand, if* p_0 *is small enough, the multilevel interpolant reaches the convergence rate* α *of the FEM approximation. For multilevel quadrature, the above calculation gives the rate* $\min\{s/d, (2p_0^{-1}-1)/(1+2d)\}$ *w.r.t. the* $H^1(\mathrm{D})$ *norm, i.e. in the best case only* $1/(1+2d)$ *times the single-level quadrature rate.*

Finally, note that (5.13) *does not contradict the last statements, since the convergence rates in* (5.13) *are given w.r.t.* $L^2(\mathrm{D})$ *(instead of* $H^1(\mathrm{D})$*) for which analogous observations (yielding better factors) can be made.*

5.3.2 Observed rates

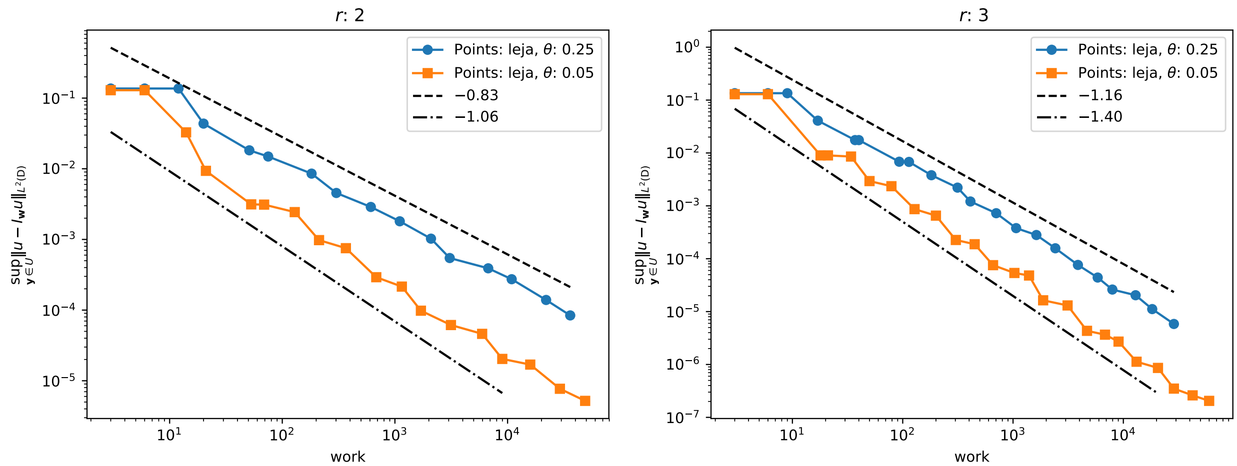
Fig. 3 shows the interpolation error for $r \in \{2, 3\}$ and $\theta \in \{0.25, 0.05\}$ measured in $L^2(\mathrm{D})$. The observed rates roughly coincide with the predicted ones, or exceed them. As before, the parameter θ has a noticeable influence on the convergence. Similar observations hold true for the quadrature error depicted in Fig. 4. As a reference value for $\int_U u(\boldsymbol{y}) \mathrm{d}\mu(\boldsymbol{y}) \in L^2(\mathrm{D})$, we use the last computed approximation.

5.4 2D domain uncertainty quantification

As a second example we consider again a diffusion problem, but this time on an uncertain domain which is assumed to be given as parametric family:

$$-\Delta \tilde{u}(\boldsymbol{y}) = 1 \qquad \text{in } \mathrm{D}_{\boldsymbol{y}}, \tag{5.17a}$$

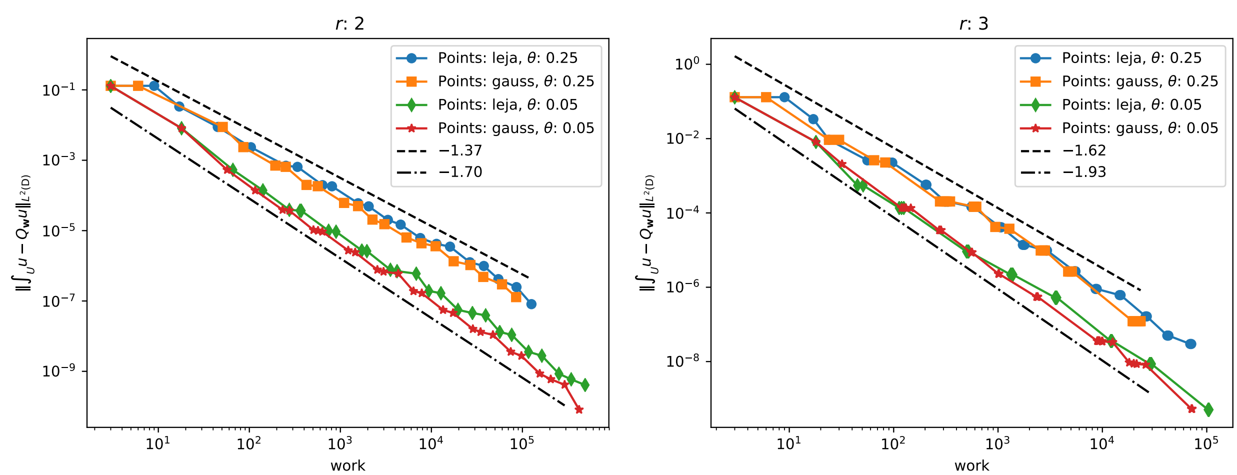
$$\tilde{u} = 0 \qquad \text{on } \partial\mathrm{D}_{\boldsymbol{y}}. \tag{5.17b}$$



(a) $r = 2$, proven rate $2/3$

(b) $r = 3$, proven rate $4/3$

Figure 3: Multilevel interpolation error for the one dimensional diffusion problem in (5.10). We consider the error $\sup_{\boldsymbol{y}\in U}\|u - I_{\mathbf{w}}u\|_{L^2(\mathrm{D})}$, i.e. w.r.t. to the L^2-norm in space. Since we use continuous, piecewise linear finite elements, the FEM rate is $\alpha = 2$. The supremum over $\boldsymbol{y} \in U$ is numerically estimated by taking the maximum of the error at 144 random points in U. The work measure is defined in (5.6).



(a) $r = 2$, proven rate $3/2$

(b) $r = 3$, proven rate 2

Figure 4: Multilevel quadrature error $\|\int_U u(\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y}) - Q_{\mathbf{w}}u\|_{L^2(\mathrm{D})}$ for the one dimensional diffusion problem in (5.10). Since we use linear finite elements, the FEM convergence rate is $\alpha = 2$ with respect to the work measure defined in (5.6).

As detailed in Section 4, this problem is equivalent to the weak formulation

$$\int_{D_0} \nabla u(\boldsymbol{y})^\top A(\boldsymbol{y}) \nabla v = \int_{D_0} v \det DT(\boldsymbol{y}) \qquad \forall\, v \in H_0^1(D_0), \tag{5.18a}$$

on the reference domain D_0 via $u(\boldsymbol{y}) = \tilde{u}(\boldsymbol{y}) \circ T(\boldsymbol{y})$, where $T_{\boldsymbol{y}} : D_0 \to D_{\boldsymbol{y}}$ and

$$A(\boldsymbol{y}) = DT(\boldsymbol{y})^{-\top} DT(\boldsymbol{y})^{-1} \det DT(\boldsymbol{y}) \in L^\infty(D_0, \mathbb{R}^{2\times 2}). \tag{5.18b}$$

Here we set $D_0 := [-1,1]^2$ and define, for $(x_1, x_2) \in D_0$,

$$T(\boldsymbol{y})(x_1, x_2) := (x_1, x_2) + \left(0, \frac{(x_2+1)}{2} \theta \sum_{j\in\mathbb{N}} y_j j^{-r-1} \sin(j\pi x_1)\right). \tag{5.19}$$

Then for every $\boldsymbol{y} \in U$, $D_{\boldsymbol{y}} \subseteq \mathbb{R}^2$ is a well-defined Lipschitz domain as long as $\theta > 0$ is small enough.

5.4.1 Theoretically predicted convergence rates

All computations will be done on the reference domain D_0. The function $u(\boldsymbol{y}) \in H_0^1(D_0)$ is approximated with a finite element method based on a quadrilateral mesh and piecewise polynomials of degree $n = 2$ in both coordinates. More precisely, for fixed $N \in \{2^m : m \in \mathbb{N}_0\}$, and with

$$x_j := \begin{cases} -1 + \left(\frac{j}{N}\right)^\beta & \text{if } 0 \le j \le N, \\ 1 - \left(\frac{j-2N}{N}\right)^\beta & \text{if } 1 \le j \le 2N, \end{cases} \tag{5.20}$$

our quadrilateral mesh on $[-1,1]^2$ has the nodes (x_i, x_j), i, $j = 0, \ldots, 2N$. With the grading factor $\beta \ge 1$ large enough, we expect to retain the optimal FEM rate $\alpha = n/d = 2/2 = 1$ (in H^1) stated in Theorem 4.5, as long as the transformation $T(\boldsymbol{y})$ (and thus the diffusion coefficient $A(\boldsymbol{y})$) is smooth enough.

Let us verify Assumption 2.16. With (5.19), we can write $T(\boldsymbol{y}) = \mathrm{Id} + \sum_{j\in\mathbb{N}} y_j \psi_j$, where $\psi_j(x_1, x_2) = (0, \theta((x_2+1)/2) j^{-(r+1)} \sin(j\pi x_1))$. With $d\psi_j \in L^\infty(D_0, \mathbb{R}^{2\times 2})$ denoting the Jacobian, clearly $(\|d\psi_j\|_{L^\infty})_{j\in\mathbb{N}} \in \ell^{1/r+\delta}$ for any $\delta > 0$. Furthermore, for $1 \le s < r$ we have $(\|d\psi_j\|_{W^{s,\infty}})_{j\in\mathbb{N}} \in \ell^{1/(r-s)+\delta}$. Assuming $r > 3$ (i.e. $A(\boldsymbol{y}) \in W^{3,\infty}$) and setting $s = 2$, by employing Example 4.4 and Corollary 4.7 we obtain that the FEM solutions u_l satisfy Assumption 2.16 for $\alpha = 1$ w.r.t. $X = H^1(D_0)$. Hence with $p_0 = 1/r + \delta$, $p_1 = 1/(r-s) + \delta$ and $s = 2$ the convergence rates of Theorems 2.31, 2.32 are (up to some arbitrarily small $\delta > 0$)

$$\begin{cases} 1 & \text{if } 4 \le r, \\ \frac{r-1}{3} & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \begin{cases} 1 & \text{if } 3 \le r, \\ \frac{2r-1}{5} & \text{otherwise,} \end{cases} \tag{5.21}$$

for multilevel interpolation/quadrature respectively. We observe that the multilevel algorithm achieves at most one third (interpolation) respectively one fifth (quadrature) of the single-level rates, cp. Remark 5.3.

Remark 5.4. *As discussed in Section 4.1.2, in order to keep the computational effort to a minimum, on cartesian product domains sparse-grid finite elements could be used, which are realized* *by the closely connected combination technique, see [30]. Also, in curved geometries, so-called* iso-geometric FEM *as introduced in [36] naturally afford separation of variables and allow for separate collocation of physical variables affording computational efficiency gains. Using the combination technique FEM as a numerical PDE solver with the presently proposed collocation essentially yields the so-called multiindex stochastic collocation method (MISC) proposed in [32].*

Roughly speaking, the combination technique FEM combines solutions on certain tensorized FEM spaces in an appropriate way to retain the optimal convergence rate (up to logarithmic factors) independent of the dimension d of the domain D*. Such results hold if the solution is smooth enough: in space dimension $d = 2$, physical domain $\mathrm{D} = (-1,1)^2$ and with the exact solution $u \in H^4(\mathrm{D})$: this implies mixed $H^{2,2}(\mathrm{D})$ regularity, and the $H^1(\mathrm{D})$-error behaves as $O(N^{-1}\log(N))$ w.r.t. the number of degrees of freedom N. For our example, this does not bring an improvement in terms of the convergence rate: Apart from the fact that according convergence results in weighted Sobolev spaces (which spaces are required for elliptic problems in nonsmooth domains) do not yet seem to be available for the combination technique, using tensorized, continuous, piecewise quadratic finite elements on graded meshes achieves the $H^1(\mathrm{D})$ convergence rate N^{-1}. As explained in Sec. 4.1.2, optimal d-independent approximation rates in weighted Sobolev spaces are also known to be achieved by sparse-grid FEM [42] (up to logarithmic factors). The combination technique has specific algorithmic benefits however, such as allowing for simple and efficient parallelization.*

5.4.2 Observed rates

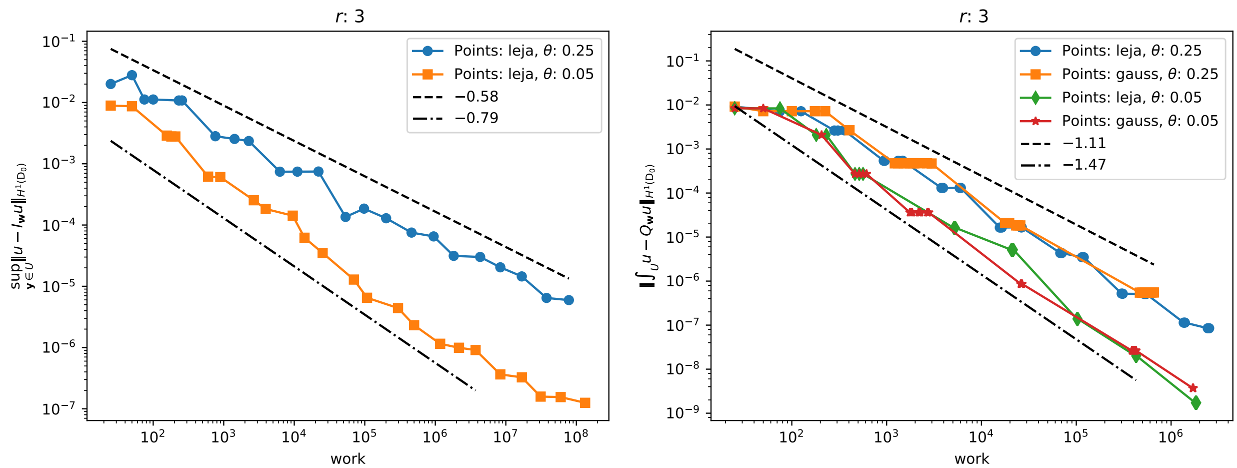
In Figure 5 we plot the interpolation and quadrature error for $r = 3$ in the $H^1(\mathrm{D}_{\mathbf{0}})$-norm. In both cases the proven rates are obtained or exceeded. Once more, the observed rates increase as θ decreases. As a reference value for $\int_U u(\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y}) \in H_0^1(\mathrm{D}_{\mathbf{0}})$ we use the last computed value.

6 Conclusions

We proposed and analyzed convergence rates of sparse-grid multilevel discretizations of well-posed, holomorphic-parametric operator equations which are possibly nonlinear, subject to possibly infinitely many parameters. Such problems typically arise in the context of operator equations with *distributed uncertain input data* from function spaces, when instances of these data are represented in terms of an unconditional basis of these spaces. Then, the mentioned parameter sequences are the coefficients in the expansion of the data with respect to the basis.

We showed that well-posedness and a suitable form of holomorphic parametric dependence of the operator and the data (quantified in the notion of $(\boldsymbol{b},\varepsilon)$-holomorphy) will imply corresponding holomorphy of the parameter-to-solution maps. We used this result to propose and analyze a sparse-grid collocation approximation of the parametric maps, with convergence rates determined only by the p-summability of the sequence $\boldsymbol{b}$ quantifying the holomorphy of the operator equation, generalizing earlier results in [17, 43, 33] and the references there to rather general, nonlinear operator equations. Importantly, we proposed a new, apriori approach for identifying near-optimal, unisolvent sparse-grids in high-dimensional parameter space in near-linear complexity in terms of the number N of collocation points.

We combined this collocation approximation with a sequence of hierarchic approximations of the operator equation. Notably, we only required abstract stability and consistency of these approximations, accommodating a very wide range of specific approximation schemes, such as Petrov-Galerkin



(a) multilevel interpolation, proven rate 2/3

(b) multilevel quadrature, proven rate 1

Figure 5: Multilevel interpolation error $\sup_{\boldsymbol{y}\in U}\|u - I_{\mathbf{w}}u\|_{H^1(\mathrm{D}_0)}$, and multilevel quadrature error $\|\int_U u(\boldsymbol{y})\mathrm{d}\mu(\boldsymbol{y}) - Q_{\mathbf{w}}u\|_{H^1(\mathrm{D}_0)}$ for the two dimensional diffusion problem in (5.18) with $r = 3$. The supremum over $\boldsymbol{y} \in U$ is numerically estimated by taking the maximum of the error at 70 random points in U. The FEM convergence rate is $\alpha = 1$ with respect to the work measure defined in (5.6).

projections of the solution, collocation approximations of the solution, spline-based approximations of uncertain geometry of the equation, etc., thereby constituting a "fully discrete, multilevel" version of [15]. Our analysis accounted for reduced summability of gpc expansion coefficients in stronger norms as is typically encountered when differentiating Karhunen-Loève expansions of distributed random inputs.

The present results on sparse-grid collocation approximation of nonlinear, holomorphic-parametric forward problems entail corresponding results for computational Bayesian inversion, due to corresponding holomorphic-parametric dependence of the parametric Bayesian posterior, with anisotropic Smolyak quadratures as proposed in [47, 48], or with higher order Quasi-Monte Carlo quadrature, as in [24]. They also provide benchmark rates for other high-dimensional approximation techniques, such as compressed sensing (see, e.g., [46] and the references there, or least-squares (see, e.g., [20] and the references there).

A Proof of Theorem 2.6

We now give a proof of Theorem 2.6. To this end we will work with sequences of the following type.

Assumption A.1. *The sequence* $\mathbf{t} = (t_j)_{j\in\mathbb{N}}$ *is monotonically decreasing and* $t_j \geq 0$ *for all* $j \in \mathbb{N}$.

Lemma A.2. *Let* $q \in (0, \infty)$, $\varepsilon > 0$ *and let* $\mathbf{t}$ *be as in Assumption A.1. Let* $\mathbf{l} = (l_j)_{j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$ *be componentwise minimal such that* $((l_j + 1)^{-\alpha q} - (l_j + 2)^{-\alpha q})t_j^q \leq \varepsilon$ *for all* $j \in \mathbb{N}$. *Then* $|\mathbf{l}| < \infty$ *and* $\sum_{j\in\mathbb{N}} t_j^q(l_j + 1)^{-q\alpha} \leq \sum_{j\in\mathbb{N}} t_j^q(m_j + 1)^{-q\alpha}$ *for all multiindices* $\mathbf{m} = (m_j)_{j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$ *with* $|\mathbf{m}| \leq |\mathbf{l}|$.

Proof. Throughout, always let $j \in \mathbb{N}$ and $w \in \mathbb{N}_0$. Set $s(j, w) := ((w + 1)^{-\alpha q} - (w + 2)^{-\alpha q})t_j^q \geq 0$ and $A := \{(j, w) : s(j, w) > \varepsilon\}$. The set A is finite since $t_j^q \to 0$ as $j \to \infty$, and thus $s(j, 0) \leq \varepsilon$ for all j large enough. Notice that A contains the $|A|$ pairs $(j,w) \in \mathbb{N} \times \mathbb{N}_0$ with the largest $s(j,w)$. Hence, for any subset $B \subset \mathbb{N} \times \mathbb{N}_0$ such that $|B| \le |A|$, there holds the inequality

$$\sum_{(j,w)\in B} s(j,w) \le \sum_{(j,w)\in A} s(j,w). \tag{A.1}$$

Let $\mathbf{m} \in \mathbb{N}_0^{\mathbb{N}}$ be arbitrary with $|\mathbf{m}| \le |A|$. Then $B := \{(j,w) : w < m_j\}$ satisfies $|B| = |\mathbf{m}| \le |A|$ and so by (A.1) we have

$$\sum_{j\in\mathbb{N}} (m_j+1)^{-\alpha q} t_j^q = \sum_{(j,w)\in\mathbb{N}\times\mathbb{N}_0} s(j,w) - \sum_{\{(j,w)\,:\,w<m_j\}} s(j,w) \ge \sum_{(j,w)\in\mathbb{N}\times\mathbb{N}_0} s(j,w) - \sum_{\{(j,w)\,:\,s(j,w)>\varepsilon\}} s(j,w). \tag{A.2}$$

Note that $s(j,w)$ is monotonically decaying in w. Therefore, if for each j, l_j is chosen minimal s.t. $s(j,l_j) \le \varepsilon$, we get $\{(j,w) : s(j,w) \le \varepsilon\} = \{(j,w) : w \ge l_j\}$ and $A = \{(j,w) : w < l_j\}$. Therefore $|A| = |\mathbf{l}|$, and the right-hand side of (A.2) equals

$$\sum_{\{(j,w)\,:\,s(j,w)\le\varepsilon\}} s(j,w) = \sum_{\{(j,w)\,:\,w\ge l_j\}} s(j,w) = \sum_{j\in\mathbb{N}} t_j^{\alpha q} \sum_{w\ge l_j} ((w+1)^{-\alpha q} - (w+2)^{-\alpha q}) = \sum_{j\in\mathbb{N}} (l_j+1)^{-\alpha q} t_j^q. \tag{A.3}$$

Since $\mathbf{m}$ in (A.2) was arbitrary with $|\mathbf{m}| \le |A| = |\mathbf{l}|$, the proof is concluded. □

We also point out, that another way to arrive at (roughly) this result is by allowing $l_j \in \mathbb{R}$, $l_j \ge 0$ and minimizing employing a Lagrange multiplier.

Proposition A.3. *Let* $\mathbf{t}$ *be as in Assumption A.1,* $\mathfrak{W}$ *as in Assumption 2.4,* $0 < q < \infty$ *and* $M, N \in \mathbb{N}$, $M \le N$. *Let* $\mathbf{w}_N = (w_{N;j})_{j\in\mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$ *be s.t.* $w_{N;j} = 0$ *for* $j > M$ *and*

$$w_{N;j} \in \mathfrak{W} \text{ is minimal s.t. } (w_{N;j}+1)^{-(\alpha q+1)} t_j^q \le \sigma_{N,M} \quad \forall\, j \in \{1,\dots,M\} \tag{A.4}$$

where

$$\sigma_{N,M} := N^{-(\alpha q+1)} \left(K_{\mathfrak{W}} \sum_{j=1}^{M} t_j^{\frac{q}{\alpha q+1}} \right)^{\alpha q+1}. \tag{A.5}$$

Then, for every $N \in \mathbb{N}$ *it holds*

(i) $K_{\mathfrak{W}}^{-1} N - M \le |\mathbf{w}_N| \le N$,

(ii)

$$\sum_{j=1}^{M} t_j^q (w_{N;j}+1)^{-\alpha q} \le 2N^{-\alpha q} \left(K_{\mathfrak{W}} \sum_{j=1}^{M} t_j^{\frac{q}{\alpha q+1}} \right)^{\alpha q+1}, \tag{A.6}$$

(iii) if $\mathbf{l} \in \mathbb{N}_0^{\mathbb{N}}$ *is arbitrary with* $|\mathbf{l}| \le |\mathbf{w}_N|$ *and* $\operatorname{supp}\mathbf{l} \subseteq \{1,\dots,M\}$, *then*

$$\sum_{j=1}^{M} t_j^q (l_j+1)^{-\alpha q} \ge (5K_{\mathfrak{W}})^{-\alpha q} \sum_{j=1}^{M} t_j^q (w_{N;j}+1)^{-\alpha q} \ge (5K_{\mathfrak{W}})^{-\alpha q} \frac{|\mathbf{w}_N|}{N} N^{-\alpha q} \left(\sum_{j=1}^{M} t_j^{\frac{q}{\alpha q+1}} \right)^{\alpha q+1}. \tag{A.7}$$

Proof. We start with (i) and (ii). The case $t_j = 0$ for all $j \in \mathbb{N}$ is trivial, and we may assume $t_1 > 0$. Furthermore, for ease of notation we shall drop the index N, i.e. we write $\mathbf{w} = \mathbf{w}_N$ and $w_j = w_{N;j}$. We begin with estimating $|\mathbf{w}|$. For $j \le M$ we either have $w_j = 0$ or $0 \le \lfloor w_j - 1 \rfloor_{\mathfrak{W}} \in \mathfrak{W}$ and $(\lfloor w_j - 1 \rfloor_{\mathfrak{W}} + 1)^{-(\alpha q+1)} t_j^q > \sigma_{N;M}$ since $w_j \in \mathfrak{W}$ was chosen minimal in (A.4). Using Assumption 2.4 we conclude for j with $w_j \ne 0$

$$(K_{\mathfrak{W}}^{-1} w_j)^{-(\alpha q+1)} t_j^q \ge (K_{\mathfrak{W}}^{-1}(w_j+1))^{-(\alpha q+1)} t_j^q \ge (\lfloor w_j - 1 \rfloor_{\mathfrak{W}} + 1)^{-(\alpha q+1)} t_j^q > \sigma_{N,M}. \tag{A.8}$$

From this we deduce $w_j \le K_{\mathfrak{W}} (t_j^q/(\sigma_{N,M}))^{1/(\alpha q+1)}$ for all $j \le M$. Employing the definition of $\sigma_{N,M}$ we get

$$|\mathbf{w}| = \sum_{j=1}^M w_j \le \sigma_{N,M}^{-\frac{1}{\alpha q+1}} K_{\mathfrak{W}} \sum_{j=1}^M t_j^{\frac{q}{\alpha q+1}} \le N. \tag{A.9}$$

Moreover, (A.4) yields $w_j + 1 \ge (t_j^q/\sigma_{N,M})^{1/(\alpha q+1)}$ for all $j \in \{1, \dots, M\}$ and thus

$$|\mathbf{w}| + M = \sum_{j=1}^M (w_j + 1) \ge \sigma_{N,M}^{-\frac{1}{\alpha q+1}} \sum_{j=1}^M t_j^{\frac{q}{\alpha q+1}} = \frac{N}{K_{\mathfrak{W}}}. \tag{A.10}$$

We now prove (ii). Using (A.4) and (A.9), we obtain

$$\sum_{j\in\mathbb{N}} (w_j+1)^{-\alpha q} t_j^q = \sum_{j=1}^M (w_j+1)(w_j+1)^{-(\alpha q+1)} t_j^q \le \sigma_{N,M}(N + |\mathbf{w}|) \le 2N^{-\alpha q} \left(K_{\mathfrak{W}} \sum_{j=1}^M t_j^{\frac{q}{\alpha q+1}} \right)^{\alpha q+1}. \tag{A.11}$$

Finally, let us prove the optimality result (iii). Fix $N \in \mathbb{N}$. With $\tilde{t}_j := t_j$ for $j \le M$ and $t_j = 0$ otherwise, Lemma A.2 states that $l_j \in \mathbb{N}_0$ minimal s.t. $((l_j+1)^{-\alpha q} - (l_j+2)^{-\alpha q}) \tilde{t}_j^q \le \alpha q (4K_{\mathfrak{W}})^{-(\alpha q+1)} \sigma_{N,M}$ for all $j \in \mathbb{N}$ is an optimal choice, under the additional constraint $l_j = 0$ for all $j > M$. By the mean value theorem there exists $\zeta \in (0,1)$ depending on l_j, α and q with $(l_j+1)^{-\alpha q} - (l_j+2)^{-\alpha q} = \alpha q (l_j + 1 + \zeta)^{-(\alpha q+1)}$. Using $K_{\mathfrak{W}} \ge 1$ we conclude for $j \le M$

$$\begin{aligned}\left(\frac{l_j}{K_{\mathfrak{W}}} + 1\right)^{-(\alpha q+1)} t_j^q &\le K_{\mathfrak{W}}^{\alpha q+1} (l_j+1)^{-(\alpha q+1)} t_j^q = K_{\mathfrak{W}}^{\alpha q+1} \frac{(l_j+1+\zeta)^{\alpha q+1}}{(l_j+1)^{\alpha q+1}} t_j^q (l_j+1+\zeta)^{-(\alpha q+1)} \\ &\le \frac{(2K_{\mathfrak{W}})^{\alpha q+1}}{\alpha q} ((l_j+1)^{-\alpha q} - (l_j+2)^{-\alpha q}) t_j^q \le 2^{-(\alpha q+1)} \sigma_{N,M}. \end{aligned} \tag{A.12}$$

The definition of w_j, l_j then gives $w_j \le \lceil K_{\mathfrak{W}}^{-1} l_j \rceil_{\mathfrak{W}}$. Now, either $l_j < K_{\mathfrak{W}}$, in which case (A.12) implies $t_j^q \le \sigma_{N,M}$ and thus $0 = w_j \le l_j$, or there exists $i \ge 1$ with $\mathfrak{w}_i \le l_j \le \mathfrak{w}_{i+1}$. Then $K_{\mathfrak{W}}^{-1} l_j \le K_{\mathfrak{W}}^{-1} \mathfrak{w}_{i+1} \le \mathfrak{w}_i$, and therefore again $w_j \le \lceil K_{\mathfrak{W}}^{-1} l_j \rceil_{\mathfrak{W}} \le l_j$. Similarly,

$$((4K_{\mathfrak{W}} w_j + 3 + 1)^{-\alpha q} - (4K_{\mathfrak{W}} w_j + 3 + 2)^{-\alpha q}) t_j^q \le \alpha q (4K_{\mathfrak{W}}(w_j+1))^{-(\alpha q+1)} \le \alpha q (4K_{\mathfrak{W}})^{-(\alpha q+1)} \sigma_{N,M}, \tag{A.13}$$

implying $l_j \le \lceil 4K_{\mathfrak{W}} w_j + 3 \rceil \le 4K_{\mathfrak{W}} w_j + 4$. Altogether $w_j + 1 \le l_j + 1 \le 5K_{\mathfrak{W}}(w_j+1)$ for all w_j. In particular $|\mathbf{l}| \ge |\mathbf{w}|$. Now let $\tilde{\mathbf{l}} \in \mathbb{N}_0^{\mathbb{N}}$ arbitrary with $|\tilde{\mathbf{l}}| \le |\mathbf{w}|$ and $\operatorname{supp} \tilde{\mathbf{l}} \subseteq \{1, \dots, M\}$. Then, using optimality of $\mathbf{l}$ as well as $|\mathbf{l}| \ge |\mathbf{w}| \ge |\tilde{\mathbf{l}}|$

$$\sum_{j=1}^M t_j^q (\tilde{l}_j + 1)^{-\alpha q} \ge \sum_{j=1}^M t_j^q (l_j+1)^{-\alpha q} \ge (5K_{\mathfrak{W}})^{-\alpha q} \sum_{j=1}^M t_j^q (w_j+1)^{-\alpha q}.$$

Ultimately, whenever $w_j \neq 0$, by (A.8) we have $(K_{\mathfrak{W}}^{-1}(w_j+1))^{-(\alpha q+1)} t_j^q \geq \sigma_{N,M}$ and hence

$$\sum_{j=1}^{M}(w_j+1)^{-\alpha q}t_j^q = \sum_{j=1}^{M}(w_j+1)(w_j+1)^{-(\alpha q+1)}t_j^q \geq K_{\mathfrak{W}}^{-(\alpha q+1)}\sigma_{N,M} \sum_{\{1\leq j\leq M\,:\,w_j\neq 0\}}(w_j+1)$$
$$\geq \frac{|\mathbf{w}|}{N}N^{-\alpha q}\left(\sum_{j=1}^{M} t_j^{\frac{q}{\alpha q+1}}\right)^{\alpha q+1}, \tag{A.14}$$

giving (A.7).

□

Remark A.4. *One verifies that* $\mathbf{w}_N = (w_{N;j})_{j\in\mathbb{N}}$ *in* (A.4) *for* $j \leq M$ *is explicitly given by*

$$w_{N;j} := \left\lceil N t_j^{\frac{q}{\alpha q+1}}\left(K_{\mathfrak{W}}\sum_{i=1}^{M} t_i^{\frac{q}{\alpha q+1}}\right)^{-1} - 1\right\rceil_{\mathfrak{W}}. \tag{A.15}$$

Lemma A.5. *Let* $p \in (0,\infty)$ *and let* $(t_j)_{j\in\mathbb{N}}$ *be as in Assumption A.1. Then for all* $N \in \mathbb{N}$

$$t_N \leq \left(\sum_{j=1}^{N} t_j^p\right)^{\frac{1}{p}} N^{-\frac{1}{p}}. \tag{A.16}$$

For the above lemma, see for instance [26, Section 7.4] and the references there. It is a consequence of the Hölder inequality, and immediately gives Stechkin's lemma, which states that for $\mathbf{t} = (t_j)_{j\in\mathbb{N}} \in \ell^p(\mathbb{N})$, $p \in (0,1)$, as in Assumption A.1

$$\sum_{j>N} t_j \leq \|\mathbf{t}\|_{\ell^p}\int_N^\infty x^{-\frac{1}{p}}\mathrm{d}x \leq \|\mathbf{t}\|_{\ell^p}\frac{p}{1-p}N^{1-\frac{1}{p}}. \tag{A.17}$$

We are now in position to prove Theorem 2.6.

Proof of Theorem 2.6. 1st Step: We start with items (i) - (iii). Due to Remark A.4 and Proposition A.3 it holds $|\mathbf{w}_N| \leq N$. The fact that (ii) is satisfied, follows immediately by the precise formula for $\mathbf{w}_N$ in Remark A.4, and because $\mathbf{t}_1 = (t_{1;j})_{j\in\mathbb{N}}$ is monotonically decreasing, i.e. $i \geq j$ implies $t_{1;i} \geq t_{1;j}$.

Let us next verify (iii). We distinguish the cases $q/(\alpha q+1) \geq p_1$ and $q/(\alpha q+1) < p_1$. In the first case we get $(\sum_{j=1}^M t_{1;j}^{q/(\alpha q+1)})^{\alpha q+1} \leq \|\mathbf{t}_1\|_{\ell^{p_1}(\mathbb{N})}^q$, so that by Proposition A.3 and (A.17) with $M = \lceil N^\beta \rceil$

$$S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{w}_N,q,\alpha) \leq \sum_{j=1}^{M} t_{1;j}^q (w_{N;j}+1)^{-\alpha q} + \sum_{j>M} t_{0;j}^q \leq 2K_{\mathfrak{W}}^{\alpha q+1}\|\mathbf{t}_1\|_{\ell^{p_1}}^q N^{-\alpha q} + \|\mathbf{t}_0\|_{\ell^{p_0}}^q \frac{p_0/q}{1-p_0/q} C N^{\beta(1-\frac{q}{p_0})}. \tag{A.18}$$

The assumption $q/(\alpha q+1) \geq p_1$ is equivalent to $\beta(q/p_0 - 1) \geq \alpha q$ as a straightforward computation shows and thus (2.9) is satisfied.

The second case corresponds to $1 > q/(p_1(\alpha q+1))$, and with Hölder's inequality

$$\left(\sum_{j=1}^{M} t_{1;j}^{\frac{q}{\alpha q+1}}\right)^{\alpha q+1} \le \|\mathbf{t}_1\|_{\ell^{p_1}}^q M^{\alpha q+1-\frac{q}{p_1}}. \tag{A.19}$$

We note that this is equation also holds in case $p_1=\infty$. Proposition A.3 with $M=\lceil N^\beta\rceil$ and (A.17) then give similarly as before

$$\begin{aligned}S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{w}_N,q,\alpha) &\le \sum_{j=1}^{M} t_{1;j}^q (w_{N;j}+1)^{-\alpha q} + \sum_{j>M} t_{0;j}^q \\ &\le 2K_{\mathfrak{W}}^{\alpha q+1}\|\mathbf{t}\|_{\ell^{p_1}}^q M^{\alpha q+1-\frac{q}{p_1}} N^{-\alpha q} + \|\mathbf{t}_0\|_{\ell^{p_0}}^q \frac{p_0/q}{1-p_0/q} M^{1-\frac{q}{p_0}} \\ &\le C(\|\mathbf{t}_1\|_{\ell^{p_1}}^q N^{-\alpha q+\beta(\alpha q+1-\frac{q}{p_1})} + \|\mathbf{t}_0\|_{\ell^{p_0}}^q N^{\beta(1-\frac{q}{p_0})}) = C(\|\mathbf{t}_0\|_{\ell^{p_0}}^q + \|\mathbf{t}_1\|_{\ell^{p_1}}^q) N^{\beta(1-\frac{q}{p_0})},\end{aligned} \tag{A.20}$$

where we have used $-\alpha q+\beta(\alpha q+1-q/p_1)=\beta(1-q/p_0)$ (which again is also true in case $p_1=\infty$). This proves (2.9) in the second case.

2nd Step: We show (iv) and begin with $p_1 \le q/(\alpha q+1)$. Set $\mathbf{t}_0 := \mathbf{t}_1 := (\delta_{1j})_{j\in\mathbb{N}}$. Then for any $\mathbf{v}$ with $|\mathbf{v}|=N$ it holds $S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{v},q,\alpha) \ge (t_{1;1}N^{-\alpha})^q = N^{-\alpha q}$.

Next let $p_1 > q/(\alpha q+1)$. The sequences $\mathbf{t}_0$, $\mathbf{t}_1$ defined by

$$t_{0;j} := j^{-\frac{1}{p_0}}, \qquad t_{1;j} := j^{-\frac{1}{p_1}}, \tag{A.21}$$

are in $\ell^{p_0+\varepsilon}$, $\ell^{p_1+\varepsilon}$ respectively for any $\varepsilon>0$ (in case $p_1=\infty$, then $t_{1;j}=1$ for all j and $\mathbf{t}_1\in\ell^{\infty+\varepsilon}=\ell^\infty$). Let $\tilde r>0$ and $(\mathbf{v}_N)_{N\in\mathbb{N}}$ such that $|\mathbf{v}_N|\le N$ and $S(q,\mathbf{t}_0,\mathbf{t}_1,\mathbf{v}_N)\le CN^{-\tilde r}$. Without loss of generality we assume that $\mathbf{v}_N\in\mathbb{N}_0^{\mathbb{N}}$ minimizes $S(q,\mathbf{t}_0,\mathbf{t}_1,\mathbf{v}_N)$ under the constraint $|\mathbf{v}_N|\le N$. In the case that this multiindex is not unique, let $\mathbf{v}_N$ be one of the multiindices for which additionally $|\mathbf{v}_N|$ becomes minimal. We claim that

$$v_{N;i}\ne 0 \Rightarrow \min\{t_{0;i}, t_{1;i}(v_{N;i}+1)^{-\alpha}\} = t_{1;i}(v_{N;i}+1)^{-\alpha} \qquad \text{and} \qquad i\le j \Rightarrow v_{N;i}\ge v_{N;j}. \tag{A.22}$$

The first implication follows since $\mathbf{v}_N$ additionally minimizes $|\mathbf{v}_N|$. For the second implication assume first $i<j$ and $v_{N;i}=0$, $v_{N;j}\ne 0$. Define $\tilde v_{N;l} := v_{N;l}$ for $l\notin\{i,j\}$ and $\tilde v_{N;i} := v_{N;j}$, $\tilde v_{N;j} := v_{N;i}$. By the first implication we have $0\le t_{0;j}-t_{1;j}(v_{N;j}+1)^{-\alpha}$ and thus $(t_{1;j}/t_{0;j})^{1/\alpha}\le v_{N;j}+1$. Therefore

$$\begin{aligned}S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{v}_N,q,\alpha) - S(\mathbf{t}_0,\mathbf{t}_1,\tilde{\mathbf{v}}_N,q,\alpha) &= (t_{0;i}+t_{1;j}(v_{N;j}+1)^{-\alpha}) - (t_{0;j}+t_{1;i}(v_{N;j}+1)^{-\alpha}) \\ &= (t_{0;i}-t_{0;j}) + (v_{N;j}+1)^{-\alpha}(t_{1;j}-t_{1;i}) \le (t_{0;i}-t_{0;j}) + (v_{N;j}+1)^{-\alpha}(t_{1;j}-t_{1;i}) \\ &\le (t_{0;i}-t_{0;j}) + \frac{t_{0;j}}{t_{1;j}}(t_{1;j}-t_{1;i}) \le t_{0;i} - t_{0;j}\frac{t_{1;i}}{t_{1;j}} \le 0,\end{aligned} \tag{A.23}$$

where the last step follows by $t_{1;i}/t_{1;j}\ge t_{0;i}/t_{0;j}$ due to our assumption that $p_0\le p_1$. Hence, we may assume without loss of generality that $v_{N;i}=0$ implies $v_{N;j}=0$ for all $j\ge i$. Next, assume that for some $i\ne j$ it holds $0\ne v_{N;i}<v_{N;j}$. Then, with $\tilde{\mathbf{v}}_N$ as above,

$$\begin{aligned}&S(\mathbf{t}_0,\mathbf{t}_1,\mathbf{v}_N,q,\alpha) - S(\mathbf{t}_0,\mathbf{t}_1,\tilde{\mathbf{v}}_N,q,\alpha) \\ &\qquad = (t_{1;i}(v_{N;i}+1)^{-\alpha} + t_{1;j}(v_{N;j}+1)^{-\alpha}) - (t_{1;i}(v_{N;j}+1)^{-\alpha} + t_{1;j}(v_{N;i}+1)^{-\alpha}) \\ &\qquad = ((v_{N;i}+1)^{-\alpha} - (v_{N;j}+1)^{-\alpha})(t_{1;i}-t_{1;j}) < 0,\end{aligned} \tag{A.24}$$

which is a contradiction. Hence (A.22) is satisfied.

Define now $M(N) := \max_j v_{N;j} \neq 0$. By (A.22) $M(N) \leq N$ since $|\mathbf{v}_N| \leq N$. Also due to (A.22) we can write

$$S(\mathbf{t}_0, \mathbf{t}_1, \mathbf{v}_N, q, \alpha) = \sum_{j=1}^{M(N)} t_{1;j}^q v_{N;j}^{-\alpha q} + \sum_{j>M(N)} t_{0;j}^q. \tag{A.25}$$

Then there exists a constant $\tilde{C} > 0$ such that

$$\sum_{j>M(N)} t_j^q \geq \int_{M(N)+1}^{\infty} x^{-\frac{q}{p_0}} \mathrm{d}x \geq \tilde{C} M(N)^{1-\frac{q}{p_0}}. \tag{A.26}$$

By assumption $S(\mathbf{t}_0, \mathbf{t}_1, \mathbf{v}_N, q, \alpha) \leq CN^{-\tilde{r}}$, and thus $M(N)^{1-\frac{q}{p_0}} \leq CN^{-\tilde{r}}$, giving

$$M(N) \geq CN^{\frac{-\tilde{r}}{1-\frac{q}{p_0}}}. \tag{A.27}$$

Next define $\tilde{N} = \tilde{N}(N) := 4N$ and let $\mathbf{w}_{\tilde{N}}$ as in Proposition A.3 for the sequence $\mathbf{t} = \mathbf{t}_1$ and with $\mathfrak{W} = \mathbb{N}_0$ (i.e. $K_{\mathfrak{W}} = 2$ in Assumption 2.4) as well as $M = M(N)$. By Proposition A.3 (i) $|\mathbf{w}_{\tilde{N}}| \geq K_{\mathfrak{W}}^{-1} \tilde{N} - M \geq 2^{-1} \cdot 4N - M \geq 2N - N \geq N \geq |\mathbf{v}_N|$. Employing Proposition A.3 (iii) we find

$$\sum_{j=1}^{M(N)} t_{1;j}^q v_{N;j}^{-\alpha q} \geq C \sum_{j=1}^{M(N)} t_{1;j}^q w_{\tilde{N};j}^{-\alpha q} \geq C \frac{|\mathbf{w}_{\tilde{N}}|}{\tilde{N}} \tilde{N}^{-\alpha q} \left(\sum_{j=1}^{M(N)} j^{-\frac{q}{p_1(\alpha q+1)}} \right) \geq CN^{-\alpha q} \left(\sum_{j=1}^{M(N)} j^{-\frac{q}{p_1(\alpha q+1)}} \right). \tag{A.28}$$

Similar as in (A.26), and exploiting $p_1 > q/(\alpha q + 1)$, i.e. $q/(p_1(\alpha q + 1)) < 1$

$$N^{-\alpha q} \left(\sum_{j=1}^{M(N)} j^{-\frac{q}{p_1(\alpha q+1)}} \right)^{\alpha q+1} \geq CN^{-\alpha q} M(N)^{\alpha q+1-\frac{q}{p_1}} \geq CN^{-\alpha q} N^{\frac{\tilde{r}}{\frac{q}{p_0}-1}(\alpha q+1-\frac{q}{p_1})}, \tag{A.29}$$

where the last inequality follows by (A.27) and $\alpha q + 1 - q/p_1 > 0$. Finally, in order for $S(\mathbf{t}_0, \mathbf{t}_1, \mathbf{v}_N, q, \alpha) \leq CN^{-\tilde{r}}$ to be satisfied, it must hold

$$-\tilde{r} \geq -\alpha q + \frac{\tilde{r}}{\frac{q}{p_0} - 1} \left(\alpha q + 1 - \frac{q}{p_1} \right). \tag{A.30}$$

Using $q(\alpha + 1/p_0 - 1/p_1) > 0$, this is equivalent to

$$\frac{\tilde{r}}{\frac{q}{p_0} - 1} \left(-\frac{q}{p_0} + 1 - \alpha q - 1 + \frac{q}{p_1} \right) \geq -\alpha q \quad \iff \quad \tilde{r} \leq \left(\frac{q}{p_0} - 1 \right) \frac{\alpha}{\alpha + \frac{1}{p_0} - \frac{1}{p_1}} = \beta \left(\frac{q}{p_0} - 1 \right). \tag{A.31}$$

Since $0 < p_0 < 1$, $p_1 \in [p_0, \infty]$ was arbitrary, and since $\mathbf{t}_0 \in \ell^{p_0+\varepsilon}$, $\mathbf{t}_1 \in \ell^{p_1+\varepsilon}$ for any $\varepsilon > 0$, it is easy to conclude that (iii) is satisfied in the current case, where $p_1 > q/(\alpha q + 1)$. $\square$

B Proof of Proposition 2.12

Proof of Proposition 2.12. If $p_1 < \infty$, we introduce the new sequences $\tilde{\mathbf{t}}_i$, $i \in \{0,1\}$ by

$$\tilde{t}_{0;\boldsymbol{\nu}} := \max\{t_{0;\boldsymbol{\nu}}, t_{1;\boldsymbol{\nu}}^{p_1/p_0}\}, \qquad \tilde{t}_{1;\boldsymbol{\nu}} := \max\{t_{0;\boldsymbol{\nu}}^{p_0/p_1}, t_{1;\boldsymbol{\nu}}\}. \tag{B.1}$$

They have the properties of being in $\ell_{\mathrm{m}}^{p_i}$ with

$$\|\tilde{\mathbf{t}}_i\|_{\ell^{p_i}} \le \|\mathbf{t}_0\|_{\ell^{p_0}} + \|\mathbf{t}_1\|_{\ell^{p_1}} + \|\mathbf{t}_0\|_{\ell^{p_0}}^{p_0/p_1} + \|\mathbf{t}_1\|_{\ell^{p_1}}^{p_1/p_0}, \tag{B.2}$$

and since

$$\tilde{t}_{0;\boldsymbol{\nu}} = \tilde{t}_{1;\boldsymbol{\nu}}^{p_1/p_0}, \tag{B.3}$$

it holds for all $\boldsymbol{\nu}$, $\boldsymbol{\mu} \in \mathcal{F}$

$$\tilde{t}_{0;\boldsymbol{\nu}} \le \tilde{t}_{0;\boldsymbol{\mu}} \Longrightarrow \tilde{t}_{1;\boldsymbol{\nu}} \le \tilde{t}_{1;\boldsymbol{\mu}}. \tag{B.4}$$

In case $p_1 = \infty$, we set $\tilde{t}_{0;\boldsymbol{\nu}} := t_{0;\boldsymbol{\nu}}$ and $\tilde{t}_{1;\boldsymbol{\nu}} := \|\mathbf{t}_1\|_{\ell^\infty}$, with (B.4) being satisfied again.

Since $(\tilde{t}_{i;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ is an $\ell_{\mathrm{m}}^{p_i}$ sequence, we may assume wlog $\tilde{t}_{i;\boldsymbol{\nu}} \ge \tilde{t}_{i;\boldsymbol{\mu}}$ whenever $\boldsymbol{\mu} \le \boldsymbol{\nu}$, $i \in \{0,1\}$. Consider an enumeration $\pi : \mathbb{N} \to \mathcal{F}$ such that $\tilde{t}_{i;\pi(j)}$ is monotonically decreasing with respect to j (for both $i \in \{0,1\}$, which is possible due to $\tilde{\mathbf{t}}_0 = \tilde{\mathbf{t}}_1^{p_1/p_0}$). By Theorem 2.6, there exists a constant $C = C(\|\tilde{\mathbf{t}}_0\|_{\ell^{p_0}}, \|\tilde{\mathbf{t}}_1\|_{\ell^{p_1}}, \alpha) > 0$ and, for each $\tilde{N} \in \mathbb{N}$ and for every $q \in [1,2]$ there exists a sequence $\tilde{\mathbf{w}}_{q;\tilde{N}} \in \mathfrak{W}^{\mathbb{N}}$ such that by (B.2) (with r depending on q)

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{\tilde{t}_{0;\boldsymbol{\nu}}, (\tilde{w}_{q;\tilde{N};\boldsymbol{\nu}} + 1)^{-\alpha}\tilde{t}_{1;\boldsymbol{\nu}}\}^q \le C\tilde{N}^{-r} \ . \tag{B.5}$$

If $p_1 < \infty$ we set $\overline{q} := p_1/(1-p_1\alpha) \in \mathbb{R}\cup\{\pm\infty\}$. In case that $\overline{q} \in (0,\infty)$, this is the value satisfying $p_1 = \overline{q}/(\alpha\overline{q}+1)$. Introduce $\tilde{\mathbf{w}}_N$ by

$$\tilde{w}_{N;\boldsymbol{\nu}} := \begin{cases} \max\{\tilde{w}_{1;\lfloor N/3\rfloor;\pi^{-1}(\boldsymbol{\nu})}, \tilde{w}_{\overline{q};\lfloor N/3\rfloor;\pi^{-1}(\boldsymbol{\nu})}, \tilde{w}_{2;\lfloor N/3\rfloor;\pi^{-1}(\boldsymbol{\nu})}\} & \text{if } p_1 < \infty \text{ and } \overline{q} \in (1,2) \\ \max\{\tilde{w}_{1;\lfloor N/2\rfloor;\pi^{-1}(\boldsymbol{\nu})}, \tilde{w}_{2;\lfloor N/2\rfloor;\pi^{-1}(\boldsymbol{\nu})}\} & \text{otherwise,} \end{cases} \tag{B.6}$$

and finally define $\mathbf{w}_N = (w_{N;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ via

$$w_{N;\boldsymbol{\nu}} := \begin{cases} \tilde{w}_{N;\boldsymbol{\nu}} & \text{if } \tilde{t}_{0;\boldsymbol{\nu}} \ge (\tilde{w}_{N;\boldsymbol{\nu}}+1)^{-\alpha}\tilde{t}_{1;\boldsymbol{\nu}} \\ 0 & \text{otherwise.} \end{cases} \tag{B.7}$$

We now verify the claims of the corollary and begin with the case where $p_1 < \infty$ and $\overline{q} \in (1,2)$. First note that $|\mathbf{w}_N| \le |\tilde{\mathbf{w}}_N| \le |\tilde{\mathbf{w}}_{1;\lfloor N/3\rfloor}| + |\tilde{\mathbf{w}}_{\overline{q};\lfloor N/3\rfloor}| + |\tilde{\mathbf{w}}_{2;\lfloor N/3\rfloor}| \le N$ by the definition of $\tilde{\mathbf{w}}_{q;N}$, $q \in \{1,\overline{q},2\}$ and Theorem 2.6 (i). Next, we show (2.17). Using (B.5) as well as (B.7) we get for $q \in \{1,\overline{q},2\}$

$$\begin{aligned} &\sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}\neq 0\}} (w_{N;\boldsymbol{\nu}}+1)^{-\alpha q} t_{1;\boldsymbol{\nu}}^q + \sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}=0\}} t_{0;\boldsymbol{\nu}}^q \\ &\quad\le \sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}\neq 0\}} (w_{N;\boldsymbol{\nu}}+1)^{-\alpha q} \tilde{t}_{1;\boldsymbol{\nu}}^q + \sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}=0\}} \tilde{t}_{0;\boldsymbol{\nu}}^q \\ &\quad= \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{\tilde{t}_{0;\boldsymbol{\nu}}, (\tilde{w}_{N;\boldsymbol{\nu}}+1)^{-\alpha}\tilde{t}_{1;\boldsymbol{\nu}}\}^q \\ &\quad\le \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{\tilde{t}_{0;\boldsymbol{\nu}}, (\tilde{w}_{q;\lfloor N/2\rfloor;\boldsymbol{\nu}}+1)^{-\alpha}\tilde{t}_{1;\boldsymbol{\nu}}\}^q \le CN^{-r}, \end{aligned} \tag{B.8}$$

showing (2.17) for $q \in \{1, \overline{q}, 2\}$.

It remains to consider $q \in (1, \overline{q})$ and $q \in (\overline{q}, 2)$. In the first case let $\theta \in (0,1)$ with $\theta + (1-\theta)\overline{q} = q$. Let now $r = r(q)$ be as in (2.7) (for the given p_0, p_1, α). As stated earlier, $\overline{q}$ is the value satisfying $p_1 = \overline{q}/(\alpha\overline{q}+1)$. Hence $r(1) = \beta(1/p_0 - 1)$, $r(\overline{q}) = \beta(\overline{q}/p_0 - 1) = \overline{q}\alpha$ (the last equality holds by the choice of $\overline{q}$) and $r(2) = 2\alpha$ where $\beta = \alpha/(\alpha + p_0^{-1} - p_1^{-1})$ (cp. (2.7)). Then, denoting the left-hand side of (2.17) by $T(q)$ we have with Hölder's inequality

$$T(q) = T(\theta + (1-\theta)\overline{q}) = \sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}\neq 0\}} \left((w_{N;\boldsymbol{\nu}}+1)^{-\alpha} t_{1;\boldsymbol{\nu}}\right)^{\theta+(1-\theta)\overline{q}} + \sum_{\{\boldsymbol{\nu}\in\mathcal{F}\,:\,w_{N;\boldsymbol{\nu}}=0\}} t_{0;\boldsymbol{\nu}}^{\theta+(1-\theta)\overline{q}}$$
$$\leq CT(1)^{\theta}T(\overline{q})^{(1-\theta)} \leq CN^{-\theta r(1)}N^{-(1-\theta)r(\overline{q})} = CN^{-(\theta\beta(1/p_0-1)+(1-\theta)\alpha\overline{q})} = N^{-\beta(q/p_0-1)} = CN^{-r(q)}. \tag{B.9}$$

For $q \in (\overline{q}, 2)$ we proceed using a similar argument by interpolating between $\overline{q}$ and 2.

Finally, if $p_1 = \infty$ or $p_1 < \infty$ and $\overline{q} \notin (1,2)$, the argument is again similar, by interpolating between 1 and 2, due to the choice of $\tilde{\mathbf{w}}$ in (B.6) in this case. □

C Proof of Proposition 2.20

Proof of Proposition 2.20. We proceed as in the proof of Proposition 2.12 and set for $\boldsymbol{\nu} \in \mathcal{F}$

$$\tilde{m}_{0;\boldsymbol{\nu}} := \max\{m_{0;\boldsymbol{\nu}}, m_{1;\boldsymbol{\nu}}^{p_1/p_0}\}, \qquad \tilde{m}_{1;\boldsymbol{\nu}} := \max\{m_{0;\boldsymbol{\nu}}^{p_0/p_1}, m_{1;\boldsymbol{\nu}}\} \tag{C.1}$$

in case $p_1 < \infty$. If $p_1 = \infty$ we define $\tilde{m}_{0;\boldsymbol{\nu}} := m_{0;\boldsymbol{\nu}}$ as well as $\tilde{m}_{1;\boldsymbol{\nu}} := \|\mathbf{m}_1\|_{\ell^\infty(\mathcal{F})} + m_{0;\boldsymbol{\nu}}$. The reason for this definition is, that these sequences are majorants of $\mathbf{m}_0$, $\mathbf{m}_1$ still satisfying (2.48) such that $\tilde{\mathbf{m}}_i \in \ell^{p_i}(\mathcal{F})$ for $i \in \{0,1\}$, and additionally

$$\tilde{m}_{0;\boldsymbol{\nu}} \leq \tilde{m}_{0;\boldsymbol{\mu}} \qquad \text{if and only if} \qquad \tilde{m}_{1;\boldsymbol{\nu}} \leq \tilde{m}_{1;\boldsymbol{\mu}}. \tag{C.2}$$

As in the proof of Proposition 2.12, we then find an enumeration $\pi : \mathbb{N} \to \mathcal{F}$ with the property that $\tilde{m}_{i;\pi(j)}$ is monotonically decaying in j (for both $i \in \{0,1\}$) and such that $\{\pi(1), \ldots, \pi(N)\}$ is downward closed for any $N \in \mathbb{N}$. Employing Theorem 2.6 we obtain $\tilde{\mathbf{w}}_N$ with $|\tilde{\mathbf{w}}_N| \leq N$ satisfying (2.49) for the majorants $\tilde{\mathbf{m}}_0$, $\tilde{\mathbf{m}}_1$, of $\mathbf{m}_0$, $\mathbf{m}_1$ and hence in particular for the latter sequences. By this theorem, for $\boldsymbol{\nu} = \pi(j)$, there holds

$$\tilde{w}_{N;\boldsymbol{\nu}} = \left\lceil N\tilde{m}_{1;\boldsymbol{\nu}}^{\frac{1}{\alpha+1}} \left(K_{\mathfrak{W}} \sum_{i=1}^{\lceil N^\beta \rceil} \tilde{m}_{1;\pi(i)}^{\frac{1}{\alpha+1}}\right)^{-1} - 1\right\rceil_{\mathfrak{W}} \tag{C.3}$$

in case $j \leq \lceil N^\beta \rceil$ (with $\beta \in (0,1]$ as in (2.7)), and $\tilde{w}_{N;\boldsymbol{\nu}} = 0$ otherwise.

For the second part, we have to slightly modify $\tilde{\mathbf{w}}_N$, which is why for every fixed $N \in \mathbb{N}$ we construct $\overline{\mathbf{w}}_N = (\overline{w}_{N;\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \mathbb{N}_0^{\mathbb{N}}$ as follows: Define for $\boldsymbol{\nu} \in \mathcal{F}$

$$\overline{w}_{N;\boldsymbol{\nu}} := \begin{cases} \tilde{w}_{N;\lfloor\boldsymbol{\nu}\rfloor_\jmath} & \text{if } \tilde{w}_{N;\lfloor\boldsymbol{\nu}\rfloor_\jmath} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{C.4}$$

Since $\tilde{m}_{1;\boldsymbol{\nu}}$ has the property (2.48) (cp. (C.1)) and because of its definition (C.3), for all $\boldsymbol{\nu} \in \mathcal{F}$ we either have $\tilde{w}_{N;\boldsymbol{\nu}} = \tilde{w}_{N;\lfloor\boldsymbol{\nu}\rfloor_\jmath}$ or $\tilde{w}_{N;\lfloor\boldsymbol{\nu}\rfloor_\jmath} \neq 0$ and $\tilde{w}_{N;\boldsymbol{\nu}} = 0$ (which can happen if the expression in (C.3) is positive but $\pi^{-1}(\boldsymbol{\nu}) > \lceil N^\beta \rceil$). Therefore $\overline{\mathbf{w}}_N \ge \tilde{\mathbf{w}}_N$ componentwise, so that $\overline{\mathbf{w}}_N$ also satisfies (2.49). Furthermore, we claim that with $K_{\mathfrak{J}}$ as in Assumption 2.18, it holds $|\overline{\mathbf{w}}_N| \le K_{\mathfrak{J}}|\tilde{\mathbf{w}}_N|$: To verify this, first we show that there is at most one

$$\boldsymbol{\nu} \in \mathcal{F} \cap \mathfrak{J}^{\mathbb{N}} \quad \text{s.t.} \quad \exists\, \boldsymbol{\mu} \in \mathcal{F} \quad \text{with} \quad \lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\nu} \quad \text{and} \quad \overline{w}_{N;\boldsymbol{\mu}} > 0 \quad \text{but} \quad \tilde{w}_{N;\boldsymbol{\mu}} = 0. \tag{C.5}$$

Suppose there is $\boldsymbol{\nu}_1 \neq \boldsymbol{\nu}_2$, both satisfying (C.5). Then due to $\boldsymbol{\nu}_i \in \mathfrak{J}^{\mathbb{N}}$, we have $\lfloor \boldsymbol{\nu}_i \rfloor_{\mathfrak{J}} = \boldsymbol{\nu}_i$ for $i \in \{1,2\}$ and wlog $\tilde{m}_{1;\boldsymbol{\nu}_1} > \tilde{m}_{1;\boldsymbol{\nu}_2}$ by (2.48). Thus for any $\boldsymbol{\mu}$ with $\lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\nu}_1$ we have $\tilde{m}_{1;\boldsymbol{\mu}} = \tilde{m}_{1;\boldsymbol{\nu}_1} > \tilde{m}_{1;\boldsymbol{\nu}_2}$ again by (2.48). Hence for the quantity in (C.4) we get

$$\left\lceil N \tilde{m}_{1;\boldsymbol{\mu}}^{\frac{1}{\alpha+1}} \left(K_{\mathfrak{W}} \sum_{i=1}^{\lceil N^\beta \rceil} \tilde{m}_{1;\pi(i)}^{\frac{1}{\alpha+1}} \right)^{-1} - 1 \right\rceil_{\mathfrak{W}} > \left\lceil N \tilde{m}_{1;\boldsymbol{\nu}_2}^{\frac{1}{\alpha+1}} \left(K_{\mathfrak{W}} \sum_{i=1}^{\lceil N^\beta \rceil} \tilde{m}_{1;\pi(i)}^{\frac{1}{\alpha+1}} \right)^{-1} - 1 \right\rceil_{\mathfrak{W}} \tag{C.6}$$

and furthermore the definition of π implies $\pi^{-1}(\boldsymbol{\mu}) < \pi^{-1}(\boldsymbol{\nu}_2)$. Consequently, the definition of $\tilde{\mathbf{w}}_N$ gives that $\tilde{w}_{N;\boldsymbol{\nu}_2} \le \tilde{w}_{N;\boldsymbol{\mu}} = 0$. This is a contradiction, and shows that $\boldsymbol{\nu}_2$ as assumed does not exist. For this reason we either have that $\boldsymbol{\nu}$ as in (C.5) does not exist or such $\boldsymbol{\nu}$ exists and (cp. (C.4))

$$|\overline{\mathbf{w}}_N| - |\tilde{\mathbf{w}}_N| \le \tilde{w}_{N;\boldsymbol{\nu}} |\{\boldsymbol{\mu} \in \mathcal{F} \,:\, \lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\nu}\}|. \tag{C.7}$$

If such $\boldsymbol{\nu}$ does not exist, then $\tilde{\mathbf{w}}_N = \overline{\mathbf{w}}_N$ and $|\overline{\mathbf{w}}_N| \le |\tilde{\mathbf{w}}_N| \le N$. Suppose on the other hand that such $\boldsymbol{\nu}$ exists. Then $\boldsymbol{\nu} \neq \mathbf{0}$ because $\{\boldsymbol{\mu} \in \mathcal{F} \,:\, \lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \mathbf{0}\} = \{\mathbf{0}\}$ by definition of $\mathfrak{J}$, so that (C.5) cannot hold for $\boldsymbol{\nu} = \mathbf{0}$. Wlog assume $\nu_1 \neq 0$ and define $\boldsymbol{\eta} = (\boldsymbol{\eta}_j)_{j\in\mathbb{N}}$ via $\eta_1 := \lfloor \nu_1 - 1 \rfloor_{\mathfrak{J}}$ and $\eta_j := \nu_j$ for $j > 1$. Then

$$\begin{aligned} |\{\boldsymbol{\mu} \in \mathcal{F} \,:\, \lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\nu}\}| &= \prod_{\{j\in\mathbb{N}\,:\,\nu_j\neq 0\}} (\nu_j - \lfloor \nu_j - 1 \rfloor_{\mathfrak{J}}) \le K_{\mathfrak{J}} \prod_{\{j\in\mathbb{N}\,:\,\eta_j\neq 0\}} (\eta_j - \lfloor \eta_j - 1 \rfloor_{\mathfrak{J}}) \\ &= K_{\mathfrak{J}} |\{\boldsymbol{\mu} \in \mathcal{F} \,:\, \lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\eta}\}| \end{aligned} \tag{C.8}$$

by Assumption 2.18. Note that since $\boldsymbol{\eta} \le \boldsymbol{\nu}$ and $w_{N;\boldsymbol{\nu}} > 0$, by (2.48) for any $\boldsymbol{\mu}$ with $\lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\eta}$ we have $\tilde{m}_{N;\boldsymbol{\mu}} = \tilde{m}_{N;\boldsymbol{\eta}} > \tilde{m}_{N;\boldsymbol{\nu}}$ which implies $\pi^{-1}(\boldsymbol{\mu}) < \pi^{-1}(\boldsymbol{\nu})$ and furthermore $\tilde{w}_{N;\boldsymbol{\mu}} \ge \tilde{w}_{N;\boldsymbol{\nu}} > 0$. By (C.4) it holds $\tilde{w}_{N;\boldsymbol{\mu}} = \overline{w}_{N;\boldsymbol{\mu}}$ for all $\boldsymbol{\mu}$ with $\lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\eta}$. With (C.7) and (C.8) we arrive at

$$|\overline{\mathbf{w}}_N| - |\tilde{\mathbf{w}}_N| \le K_{\mathfrak{J}} \sum_{\{\boldsymbol{\mu}\,:\,\lfloor \boldsymbol{\mu} \rfloor_{\mathfrak{J}} = \boldsymbol{\eta}\}} \tilde{w}_{N;\boldsymbol{\mu}} \le K_{\mathfrak{J}} |\tilde{\mathbf{w}}_N|, \tag{C.9}$$

and so we conclude $|\overline{\mathbf{w}}_N| \le (1 + K_{\mathfrak{J}})|\tilde{\mathbf{w}}_N| \le (1 + K_{\mathfrak{J}})N$ for all $N \in \mathbb{N}$.

Finally set

$$\mathbf{w}_N := \overline{\mathbf{w}}_{\lfloor \frac{N}{1+K_{\mathfrak{J}}} \rfloor}. \tag{C.10}$$

It remains to check the claimed properties. As mentioned above, $\overline{\mathbf{w}}_N$ satisfies (2.49) (with $\mathbf{w}_N$ replaced by $\overline{\mathbf{w}}_N$), and thus

$$\begin{aligned} \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{m_{0;\boldsymbol{\nu}}, m_{1;\boldsymbol{\nu}} (w_{N;\boldsymbol{\nu}} + 1)^{-\alpha}\} &= \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{m_{0;\boldsymbol{\nu}}, m_{1;\boldsymbol{\nu}} (\overline{w}_{\lfloor \frac{N}{1+K_{\mathfrak{J}}} \rfloor;\boldsymbol{\nu}} + 1)^{-\alpha}\} \\ &\le \sum_{\boldsymbol{\nu}\in\mathcal{F}} \min\{m_{0;\boldsymbol{\nu}}, m_{1;\boldsymbol{\nu}} (\tilde{w}_{\lfloor \frac{N}{1+K_{\mathfrak{J}}} \rfloor;\boldsymbol{\nu}} + 1)^{-\alpha}\} \le C \left\lfloor \frac{N}{1+K_{\mathfrak{J}}} \right\rfloor^{-r} \le C N^{-r}, \end{aligned} \tag{C.11}$$

and

$$|\mathbf{w}_N| = |\overline{\mathbf{w}}_{\lfloor \frac{N}{1+K_\mathfrak{J}} \rfloor}| \le (1+K_\mathfrak{J})|\tilde{\mathbf{w}}_{\lfloor \frac{N}{1+K_\mathfrak{J}} \rfloor}| \le N \tag{C.12}$$

by (C.9) and because $|\tilde{\mathbf{w}}_{\lfloor N/(1+K_\mathfrak{J})\rfloor}| \le \lfloor N/(1+K_\mathfrak{J})\rfloor$ as stated in Theorem 2.6. Next, (2.51) follows from $w_{N;\boldsymbol{\nu}} = w_{N;\boldsymbol{\mu}}$ for all $\boldsymbol{\mu}$ with $\lfloor \boldsymbol{\mu} \rceil_\mathfrak{J} = \boldsymbol{\nu}$, which in turn holds by the definition of $\mathbf{w}_N$ and (C.4). In order to show downward closedness of $\Lambda_{l;N}$, it is sufficient to prove (2.38), i.e.

$$\boldsymbol{\nu} \le \boldsymbol{\mu} \Rightarrow w_{N;\boldsymbol{\nu}} \ge w_{N;\boldsymbol{\mu}}. \tag{C.13}$$

Let $\boldsymbol{\mu} \le \boldsymbol{\nu}$ with $\pi(j) = \boldsymbol{\mu}$, $\pi(i) = \boldsymbol{\nu}$. Then $j \le i$ since otherwise $\{\pi(1), \dots, \pi(j)\}$ would not be downward closed, as it would contain $\boldsymbol{\nu}$ but not $\boldsymbol{\mu}$. Hence $\tilde{w}_{M;\boldsymbol{\mu}} \ge \tilde{w}_{M;\boldsymbol{\nu}}$ by Theorem 2.6 (for all $M \in \mathbb{N}$). This implies $w_{N;\boldsymbol{\mu}} \ge w_{N;\boldsymbol{\nu}}$ by (C.4) and (C.10).

Finally we verify (2.50). If $\boldsymbol{\nu}$, $\boldsymbol{\mu}$ are such that $\tilde{m}_{1;\boldsymbol{\nu}} > \tilde{m}_{1;\boldsymbol{\mu}}$, then $\pi^{-1}(\boldsymbol{\nu}) < \pi^{-1}(\boldsymbol{\mu})$ by definition of π. Thus the definition of $\tilde{\mathbf{w}}_N$ in (C.3) gives the implication

$$\tilde{m}_{1;\boldsymbol{\nu}} > \tilde{m}_{1;\boldsymbol{\mu}} \Rightarrow \tilde{w}_{N;\boldsymbol{\nu}} \ge \tilde{w}_{N;\boldsymbol{\mu}}. \tag{C.14}$$

Hence, due to (C.4), (C.10) and (2.48) for $\tilde{\mathbf{m}}_1$, we conclude with $\varepsilon_{i;l;N} := \min_{\boldsymbol{\nu} \in \Lambda_{l;N}} \tilde{m}_{i;\boldsymbol{\nu}}$ and (C.2)

$$\Lambda_{l;N} = \{\boldsymbol{\nu} \in \mathcal{F} \,:\, \tilde{m}_{1;\boldsymbol{\nu}} \ge \varepsilon_{1;l;N}\} = \{\boldsymbol{\nu} \in \mathcal{F} \,:\, \tilde{m}_{0;\boldsymbol{\nu}} \ge \varepsilon_{0;l;N}\}. \tag{C.15}$$

For $p_1 < \infty$, the sequence $\tilde{\mathbf{m}}_0$ was defined in (C.1) as the elementwise maximum of the two sequences $\mathbf{m}_0$ and $\mathbf{m}_1^{p_1/p_0}$, both of which fulfil (2.45). But then $\tilde{\mathbf{m}}_0$ also fulfils (2.45) if $p_1 < \infty$: Since $\tilde{m}_{0;\boldsymbol{\nu}} = \max\{m_{0;\boldsymbol{\nu}}, m_{1;\boldsymbol{\nu}}^{p_0/p_1}\}$, for $x > 0$ it holds

$$\Lambda(\tilde{\mathbf{m}}_0; x) = \{\boldsymbol{\nu} \in \mathcal{F} \,:\, \tilde{m}_{0;\boldsymbol{\nu}} \ge x\} = \Lambda(\mathbf{m}_0; x) \cup \Lambda(\mathbf{m}_1^{p_1/p_0}; x). \tag{C.16}$$

Thus, with d as in (2.44)

$$\begin{aligned} d(\Lambda(\tilde{\mathbf{m}}_0; x)) &\le d(\Lambda(\mathbf{m}_0; x)) + d(\Lambda(\mathbf{m}_1^{p_1/p_0}; x)) \le o(\log(|\Lambda(\mathbf{m}_0; x)|)) + o(\log(|\Lambda(\mathbf{m}_1^{p_1/p_0}; x)|)) \\ &\le o(\log(|\Lambda(\tilde{\mathbf{m}}_{0;\boldsymbol{\nu}}; x)|)) \qquad \text{as } x \to 0. \end{aligned} \tag{C.17}$$

The corresponding statement for m in (2.45) can be shown analogously. Next, if $p_1 = \infty$ we have $\tilde{\mathbf{m}}_0 = \mathbf{m}_0$ and thus $\tilde{\mathbf{m}}_0$ satisfies (2.45) by assumption.

Therefore $\tilde{\mathbf{m}}_0$ satisfies (2.45) for any p_1, and consequently (C.15) implies (2.50). The constants in (2.50) do not depend on l because the sequence $\tilde{\mathbf{m}}_0$ in (C.15) does not depend on l. □

D Proof of Lemma 5.1

Proof of Lemma 5.1. It is easy to check that $\tilde{\mathbf{t}}_i$, $i \in \{0, 1\}$, also satisfy (5.2) (possibly with different constants C_1, C_2), and those sequences are monotonically decreasing. For simplicity we omit the $\delta > 0$ argument in the following, and simply assume that a decreasing rearrangement $(\tilde{t}_{1;j})_{j \in \mathbb{N}}$ of $(\tilde{t}_{1;\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}}$ satisfies $\tilde{t}_{1;j} \sim j^{-r_1}$, where by this notation we mean $C_1 j^{-r} \le \tilde{t}_{1;j} \le C_2 j^{-r}$ for all $j \in \mathbb{N}$ and some fixed positive constants C_1, C_2. In the following let $N := |\Lambda(\varepsilon)|$. We have

$$\sum_{\boldsymbol{\nu} \in \Lambda(\varepsilon)^c} t_{0;\boldsymbol{\nu}} \le C \sum_{j > N} j^{-r_0} \le C N^{1-r_0}. \tag{D.1}$$

To prove (5.4), we start with the first case where $r_1 < 1+\alpha$. It holds

$$\sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} \tilde{t}_{1;\boldsymbol{\nu}}^{\frac{1}{1+\alpha}} \le C\sum_{j=1}^{N} j^{-\frac{r_1}{1+\alpha}} \le CN^{1-\frac{r_1}{1+\alpha}}. \tag{D.2}$$

Moreover

$$C_\varepsilon = \max_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} \tilde{t}_{1;\boldsymbol{\nu}}^{\frac{1}{\alpha}-\frac{r_0}{\alpha r_1}-\frac{1}{\alpha+1}} \sim N^{\frac{r_1}{\alpha+1}+\frac{r_0-r_1}{\alpha}}. \tag{D.3}$$

Hence employing Assumption 2.4

$$|\mathbf{w}_\varepsilon| \le CC_\varepsilon N^{1-\frac{r_1}{\alpha+1}} \le CN^{\frac{r_1}{\alpha+1}+\frac{r_0-r_1}{\alpha}} N^{1-\frac{r_1}{\alpha+1}} = CN^{1+\frac{r_0-r_1}{\alpha}}. \tag{D.4}$$

In a similar fashion one gets $|\mathbf{w}_\varepsilon| \ge CN^{1+(r_0-r_1)/\alpha}$. Moreover, due to $w_{\varepsilon;\boldsymbol{\nu}} \ge C_\varepsilon \tilde{t}_{1;\boldsymbol{\nu}}^{1/(\alpha+1)}$ we have $\tilde{t}_{1;\boldsymbol{\nu}} w_\varepsilon^{-(\alpha+1)} \le C_\varepsilon^{-(\alpha+1)}$ and thus

$$\begin{aligned}\sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} (w_{\varepsilon;\boldsymbol{\nu}}+1)^{-\alpha} t_{1;\boldsymbol{\nu}} &\le \sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} w_{\varepsilon;\boldsymbol{\nu}}^{-\alpha} t_{1;\boldsymbol{\nu}} = \sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} w_{\varepsilon;\boldsymbol{\nu}} w_{\varepsilon;\boldsymbol{\nu}}^{-\alpha-1} t_{1;\boldsymbol{\nu}} \le |\mathbf{w}_\varepsilon| C_\varepsilon^{-(\alpha+1)} \\ &\le CN^{1+\frac{r_0-r_1}{\alpha}} N^{-r_1-(r_0-r_1)\frac{\alpha+1}{\alpha}} = CN^{1-r_0}. \end{aligned} \tag{D.5}$$

Thus, overall

$$\sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)} (w_{\varepsilon;\boldsymbol{\nu}}+1)^{-\alpha} t_{1;\boldsymbol{\nu}} + \sum_{\boldsymbol{\nu}\in\Lambda(\varepsilon)^c} t_{0;\boldsymbol{\nu}} \le CN^{1-r_0} \le C|\mathbf{w}_\varepsilon|^{\frac{1-r_0}{1+\frac{r_0-r_1}{\alpha}}} = C|\mathbf{w}_\varepsilon|^{(1-r_0)\beta}, \tag{D.6}$$

with $\beta = \alpha/(\alpha+r_0-r_1)$.

The proof for the second case is similar. □

References

[1] C. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Studies in Economic Theory. Springer, third edition, 2006.

[2] M. Bachmayr, A. Cohen, D. Dũng, and C. Schwab. Fully discrete approximation of parametric and stochastic elliptic PDEs. *SIAM J. Numer. Anal.*, 55(5):2151–2186, 2017.

[3] J. Beck, R. Tempone, F. Nobile, and L. Tamellini. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Models Methods Appl. Sci.*, 22(9):1250023, 33, 2012.

[4] J.-P. Berrut and L. N. Trefethen. Barycentric Lagrange interpolation. *SIAM Rev.*, 46(3):501–517, 2004.

[5] M. Bieri, R. Andreev, and C. Schwab. Sparse tensor discretization of elliptic SPDEs. *SIAM J. Sci. Comput.*, 31(6):4281–4304, 2009/10.

[6] M. Bieri and C. Schwab. Sparse high order FEM for elliptic sPDEs. *Comput. Methods Appl. Mech. Engrg.*, 198(13-14):1149–1170, 2009.

[7] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[8] C. Băcuţă, H. Li, and V. Nistor. Differential operators on domains with conical points: precise uniform regularity estimates. *Rev. Roumaine Math. Pures Appl.*, 62(3):383–411, 2017.

[9] H.-J. Bungartz and M. Griebel. A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives. *J. Complexity*, 15(2):167–199, 1999.

[10] J.-P. Calvi and P. V. Manh. Lagrange interpolation at real projections of Leja sequences for the unit disk. *Proc. Amer. Math. Soc.*, 140(12):4271–4284, 2012.

[11] J.-P. Calvi and M. Phung Van. On the Lebesgue constant of Leja sequences for the unit disk and its applications to multivariate interpolation. *J. Approx. Theory*, 163(5):608–622, 2011.

[12] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab. Adaptive algorithms for sparse polynomial approximation of parametric and stochastic elliptic PDEs. *M2AN Math. Mod. and Num. Anal.*, 47(1):253–280, 2013.

[13] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Journ. Found. Comp. Math.*, 14(4):601–633, 2013.

[14] M. A. Chkifa. On the Lebesgue constant of Leja sequences for the complex unit disk and of their real projection. *J. Approx. Theory*, 166:176–200, 2013.

[15] A. Cohen, A. Chkifa, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. *Journ. Math. Pures et Appliquees*, 103(2):400–428, 2015.

[16] A. Cohen and R. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numer.*, 24:1–159, 2015.

[17] A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic pdes. *Analysis and Applications*, 9(1):11–47, 2010.

[18] A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10(6):615–646, 2010.

[19] A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. *Anal. Appl. (Singap.)*, 9(1):11–47, 2011.

[20] A. Cohen, G. Migliorati, and F. Nobile. Discrete least-squares approximations over optimized downward closed polynomial spaces in arbitrary dimension. *Constr. Approx.*, 45(3):497–519, 2017.

[21] A. Cohen, C. Schwab, and J. Zech. Shape holomorphy of the stationary Navier-Stokes Equations. Technical Report 2016-45, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2016. (to appear in SIAM Journ. Mathematical Analysis (2018)).

[22] D. Dũng. Linear collective collocation and Galerkin approximations for parametric and stochastic elliptic pdes, 2015, arXiv:1511.03377.

[23] K. Deimling. *Nonlinear functional analysis.* Springer-Verlag, Berlin, 1985.

[24] J. Dick, R. N. Gantner, Q. T. Le Gia, and C. Schwab. Multilevel higher-order quasi-Monte Carlo Bayesian estimation. *Math. Models Methods Appl. Sci.*, 27(5):953–995, 2017.

[25] J. Dick, Q. T. LeGia, and C. Schwab. Higher order quasi monte carlo integration for holomorphic, parametric operator equations. *SIAM Journ. Uncertainty Quantification*, 4(1):48–79, 2016.

[26] D. Dũng, V. N. Temlyakov, and T. Ullrich. Hyperbolic cross approximation, 2016, arXiv:1601.03978.

[27] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.

[28] C. Gittelson. Convergence rates of multilevel and sparse tensor approximations for a random elliptic pde. *SIAM Journal on Numerical Analysis*, 51(4):2426–2447, 2013.

[29] M. Griebel and H. Harbrecht. A note on the construction of L-fold sparse tensor product spaces. *Constr. Approx.*, 38(2):235–251, 2013.

[30] M. Griebel and H. Harbrecht. On the convergence of the combination technique. In *Sparse grids and applications—Munich 2012*, volume 97 of *Lect. Notes Comput. Sci. Eng.*, pages 55–74. Springer, Cham, 2014.

[31] M. D. Gunzburger, C. G. Webster, and G. Zhang. Stochastic finite element methods for partial differential equations with random input data. *Acta Numer.*, 23:521–650, 2014.

[32] A.-L. Haji-Ali, F. Nobile, L. Tamellini, and R. Tempone. Multi-index stochastic collocation convergence rates for random PDEs with parametric regularity. *Found. Comput. Math.*, 16(6):1555–1605, 2016.

[33] A.-L. Haji-Ali, F. Nobile, L. Tamellini, and R. Tempone. Multi-Index Stochastic Collocation for random PDEs. *Comput. Methods Appl. Mech. Engrg.*, 306:95–122, 2016.

[34] M. Hansen and C. Schwab. Sparse adaptive approximation of high dimensional parametric initial value problems. *Vietnam Journal of Mathematics*, 41(2):181–215, 2013.

[35] S. Heinrich. *Multilevel Monte Carlo Methods*, pages 58–67. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[36] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Comput. Methods Appl. Mech. Engrg.*, 194(39-41):4135–4195, 2005.

[37] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces, Volume I: Martingales and Littlewood-Paley Theory.* Springer International Publishing, 2016.

[38] C. Jerez-Hanckes, C. Schwab, and J. Zech. Electromagnetic wave scattering by random surfaces: shape holomorphy. *Math. Models Methods Appl. Sci.*, 27(12):2229–2259, 2017.

[39] H. Li, A. Mazzucato, and V. Nistor. Analysis of the finite element method for transmission/mixed boundary value problems on general polygonal domains. *Electron. Trans. Numer. Anal.*, 37:41–69, 2010.

[40] J. Mujica. *Complex analysis in Banach spaces*, volume 120 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1986. Holomorphic functions and domains of holomorphy in finite and infinite dimensions, Notas de Matemática [Mathematical Notes], 107.

[41] G. A. Muñoz, Y. Sarantopoulos, and A. Tonge. Complexifications of real Banach spaces, polynomials and multilinear maps. *Studia Math.*, 134(1):1–33, 1999.

[42] P.-A. Nitsche. Sparse approximation of singularity functions. *Constr. Approx.*, 21(1):63–81, 2005.

[43] F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs. *Numer. Math.*, 134(2):343–388, 2016.

[44] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[45] J. Pousin and J. Rappaz. Consistency, stability, a priori and a posteriori errors for Petrov-Galerkin methods applied to nonlinear problems. *Numer. Math.*, 69(2):213–231, 1994.

[46] H. Rauhut and C. Schwab. Compressive sensing petrov-galerkin approximation of high-dimensional parametric operator equations. *Mathematics of Computation*, 86(304):661–700, 2017.

[47] C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6):1–28, 2013.

[48] C. Schillings and C. Schwab. Sparsity in Bayesian inversion of parametric operator equations. *Inverse Problems*, 30(6), 2014.

[49] C. Schwab and R. A. Todor. Convergence rates of sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA Journal of Numerical Analysis*, 44:232–261, 2007.

[50] A. G. Werschulz. The complexity of the Poisson problem for spaces of bounded mixed derivatives. In *The mathematics of numerical analysis (Park City, UT, 1995)*, volume 32 of *Lectures in Appl. Math.*, pages 895–914. Amer. Math. Soc., Providence, RI, 1996.

[51] J. Zech, 2018. Dissertation, ETH Zürich, in preparation.

[52] J. Zech and C. Schwab. Convergence rates of high dimensional Smolyak quadrature. Technical Report 2017-27, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2017.

[53] C. Zenger. Sparse grids. In *Parallel algorithms for partial differential equations (Kiel, 1990)*, volume 31 of *Notes Numer. Fluid Mech.*, pages 241–251. Friedr. Vieweg, Braunschweig, 1991.