Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn
Friedrich-Hirzebruch-Allee 7 • 53115 Bonn • Germany
phone +49 228 73-69828 • fax +49 228 73-69847
www.ins.uni-bonn.de

B. Adcock, M. Griebel, G. Maier

The Sample Complexity of Learning Lipschitz Operators with respect to Gaussian Measures

INS Preprint No. 2402

October 2024

The Sample Complexity of Learning Lipschitz Operators with respect to Gaussian Measures

Ben Adcock
Department of Mathematics, Simon Fraser University, 8888 University Drive, V5A 1S6, Burnaby BC, Canada

and

Michael Griebel and Gregor Maier*
Institute for Numerical Simulation, University of Bonn, Friedrich-Hirzebruch-Allee 7, 53115 Bonn, Germany

Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven 1, 53754 Sankt Augustin, Germany
*Corresponding author: maier@ins.uni-bonn.de

Operator learning, the approximation of mappings between infinite-dimensional function spaces using machine learning, has gained increasing research attention in recent years. Approximate operators, learned from data, can serve as efficient surrogate models for problems in computational science and engineering, complementing traditional methods. However, despite their empirical success, our understanding of the underlying mathematical theory is in large part still incomplete. In this paper, we study the approximation of Lipschitz operators with respect to Gaussian measures. We prove higher Gaussian Sobolev regularity of Lipschitz operators and establish lower and upper bounds on the Hermite polynomial approximation error. We then study general reconstruction strategies of Lipschitz operators from m arbitrary (potentially adaptive) linear samples. As a key finding, we tightly characterize the corresponding sample complexity, that is, the smallest achievable worst-case error among all possible choices of (adaptive) sampling and reconstruction strategies in terms of m. As a consequence, we identify an inherent curse of sample complexity: No method to approximate Lipschitz operators based on m linear samples can achieve algebraic convergence rates in m. On the positive side, we prove that a sufficiently fast spectral decay of the covariance operator of the underlying Gaussian measure guarantees convergence rates which are arbitrarily close to any algebraic rate. Overall, by tightly characterizing the sample complexity, our work confirms the intrinsic difficulty of learning Lipschitz operators, regardless of the data or learning technique.

Keywords: Operator learning; Sample complexity; Lipschitz operators; Gaussian measures.

1. Introduction

We study the approximation of generic Lipschitz operators which map between (infinite-) dimensional Hilbert spaces. The approximation error is measured in expectation in L^2 with input samples drawn from a Gaussian probability distribution. We commence with a detailed literature review in Section 1.1, where we put our work in the context of operator learning and motivate the Gaussian setting as a natural framework for analyzing Lipschitz operators.

We subsequently summarize our main contributions in Section 1.2 and give an overview of the organization of the remainder of the paper.

1.1. *Motivation and literature review*

With the rise of machine learning, in particular deep learning, in computational science and engineering (CSE), operator learning has recently emerged as a new paradigm for the data-driven approximation of mappings between infinite-dimensional function spaces. Multiple deep learning architectures, typically referred to as *neural operators*, such as DeepONet (Lu et al., 2021), FNO (Li et al., 2021), non-local neural operators (Kovachki et al., 2023), and PCA-Net (Bhattacharya et al., 2021), have been proposed and their efficiency has been demonstrated in various practical applications. We refer to the recent reviews Kovachki et al. (2024b) and Boullé and Townsend (2024) and references therein. However, the empirical success of these neural operators has to large extent not yet been supported so far by a general mathematical theory. A thorough understanding of theoretical approximation guarantees is important though for a reliable deployment of operator learning methods in CSE applications.

1.1.1. Theory of operator learning

A typical starting point in the theoretical analysis of operator learning are universal approximation results. They guarantee the existence of a neural operator of certain type which approximates a target operator up to some arbitrarily small error (Chen and Chen, 1995; Lanthaler et al., 2022; Kovachki et al., 2021; Lanthaler et al., 2024; Lanthaler, 2023). Albeit being necessary for assessing the basic utility of a neural operator, mere existence results are of limited use in practical applications, where instead questions about quantitative approximation guarantees and explicit convergence rates are of greater importance.

To address the latter, two quantities are of key interest. On the one hand, the *parametric complexity* quantifies the convergence of the approximation error in terms of the number of tunable parameters employed by the approximation method. In the context of (deep) neural network (NN) approximations, this is often referred to as *expression rates*. On the other hand, the *sample complexity* quantifies the convergence of the approximation error in terms of the number of samples used for fitting the parameters to data. Previous research efforts mainly focused on deriving expression rates for NN approximations of specific (classes of) operators, whereas there has been comparably little work on sample complexity estimates.

Holomorphic operators have been widely-studied in operator learning. They arise, for example, as parameter-to-solution mappings for parametric partial differential equations (PDEs), see, e.g., Cohen and DeVore (2015) and Adcock et al. (2022, Chpt. 4) and references therein. Such operators can be learned with algebraic or (on finite-dimensional domains) even exponential parametric complexity with NNs (Opschoor et al., 2022; Herrmann et al., 2024; Schwab and Zech, 2023). Moreover, they can be approximated with near-optimal sample complexity with least-squares and compressed sensing methods (Adcock et al., 2024a; Bartel and Dũng, 2024; Adcock et al., 2024c, 2025b) as well as with NNs (Adcock et al., 2024d,b, 2025a). We mention in passing that algebraic NN expression rates and exponential sample complexity estimates (in probability) have also been derived for classes of (non-holomorphic) solution operators of certain PDEs (De Ryck and Mishra, 2022; Boullé et al., 2023). Moreover, algebraic complexity estimates are also available for infinite-dimensional functionals (with one-dimensional codomain) with mixed regularity (Dũng and Griebel, 2016).

1.1.2. Analysis of Lipschitz operators

However, many operators fail to be holomorphic. Amongst families of non-holomorphic operators, Lipschitz operators—which arise, for example, as parameter-to-solution mappings for parametric elliptic variational inequalities—are an important class. Applications are numerous and include, inter alia, obstacle problems in mechanics and control problems, as well as uncertainty quantification in equilibrium models. The respective operators in these contexts do generically not exhibit any regularity beyond Lipschitz continuity. We refer to Schwab et al. (2023) and Gwinner et al. (2021, Chpt. 6) and references therein for further details.

The study of approximating Lipschitz operators has recently attracted increased research attention as they exhibit fundamentally different properties to holomorphic operators in terms of learnability. It was shown in Lanthaler (2023) that bounded Lipschitz (and C^k-Fréchet differentiable) operators cannot be approximated with algebraic parametric complexity using PCA-Net. More specifically, the number of real-valued PCA-Net parameters scales exponentially with the inverse of the approximation error. This result, termed the *curse of parametric complexity*, can be interpreted as the infinite-dimensional analogue to the classical curse of dimensionality in finite dimensions, see also Lanthaler and Stuart (2024). It can be seemingly overcome by neural operators which use hyper-expressive activation functions or non-standard NN architectures (Schwab et al., 2023). In practical implementations, however, each real-valued parameter can only be represented by a sequence of bits of finite length. In Lanthaler (2024), the cost model of counting real-valued parameters was therefore replaced by instead counting the number of bits that are necessary to digitally encode each parameter to some finite accuracy. The resulting cost-accuracy scaling law reveals a curse of parametric complexity that is independent of the activation functions used in any NN approximation. It states that the number of bits required to encode each NN parameter, in fact, still scales exponentially with the inverse of the approximation error.

Based on the theory of widths, it was shown in Kovachki et al. (2024a) that bounded Lipschitz and C^k-operators also exhibit a *curse of data complexity*. That is, the error in expectation with respect to a Gaussian measure with at most algebraically decaying PCA eigenvalues converges at most logarithmically in the number of samples for any learning algorithm which is based on i.i.d. pointwise samples. However, Lipschitz operators that arise in practical applications as mentioned above are usually not bounded and the just reviewed results are thus not directly applicable.

1.1.3. Analysis of Gaussian measures

In the present paper, we prove a *curse of sample complexity* which generalizes the result of Kovachki et al. (2024a) to unbounded Lipschitz operators and arbitrary (centered, nondegenerate) Gaussian measures. Unlike in previous studies, we tightly characterize the sample complexity of learning such operators in terms of the PCA eigenvalues. We then show that algebraic convergence cannot be achieved, regardless of the decay of these eigenvalues. Nevertheless, we show that with sufficiently fast spectral decay, error decay rates which are arbitrarily close to any algebraic rate can be achieved. We mention the recent work Liu et al. (2024) for further results in the statistical theory of deep nonparametric estimation of Lipschitz operators. Therein, however, the authors work with probability measures with compact support. Consequently, their results are not directly applicable to Gaussian measures.

Our work focuses on Gaussian measures as they are the typical choice of input measures. They also allow us to draw on results from infinite-dimensional analysis (Bogachev, 1998;

Da Prato, 2006, 2014; Lunardi et al., 2015/2016). The theory of Gaussian Sobolev spaces is key in our analysis as it is well-known that Lipschitz functionals are Gaussian Sobolev functionals. This connection yields explicit control over error bounds in terms of the spectral properties of the covariance operator of the Gaussian measure. The Gaussian setting has been considered previously to prove expression rates for NN approximations of operators, see, e.g., Schwab and Zech (2023); Dũng et al. (2023). It can also be studied within the abstract framework developed in Griebel and Oswald (2017), see Example 1 therein, to derive dimension-independent results for the approximation of high- and infinite-dimensional function(al)s. Its connection to Lipschitz regularity, however, has, to the best of our knowledge, not yet been exploited to derive sample complexity estimates for Lipschitz operators.

1.2. *Contributions*

Let $\mathcal{X},\mathcal{Y}$ be separable Hilbert spaces with $\dim(\mathcal{X})=\infty$ and μ be a Gaussian measure on $\mathcal{X}$. Detailed notation and further preliminaries are introduced in Section 2. Additional notions and technical results from operator theory and infinite-dimensional analysis are discussed in the appendix. We now give an (informal) overview of our three main contributions

1. In Section 3, we extend standard results from infinite-dimensional analysis and define the (weighted) Gaussian Sobolev space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ by means of a sequence of positive real-valued weights $\boldsymbol{b}=(b_i)_{i\in\mathbb{N}}$ with $0<b_i\leq 1$. As our first main contribution, we show that this space contains the set $\mathrm{Lip}(\mathcal{X},\mathcal{Y})$ of Lipschitz operators which map from $\mathcal{X}$ to $\mathcal{Y}$:

Result 1 (Lipschitz operators are Gaussian Sobolev operators, cf. Thm. 3.10)**.** *If $\mathcal{Y}$ is finite-dimensional, then* $\mathrm{Lip}(\mathcal{X},\mathcal{Y})\subset W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$*. If $\mathcal{Y}$ is infinite-dimensional and if $\boldsymbol{b}\in\ell^2(\mathbb{N})$, then* $\mathrm{Lip}(\mathcal{X},\mathcal{Y})\subset W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$*. In both cases, the space of bounded Lipschitz operators is continuously embedded in* $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$*.*

The sequence $\boldsymbol{b}$ is essential to treat the case $\dim(\mathcal{Y})=\infty$ and it can be interpreted as a sequence of parameters which control the degree of (weak) differentiability of the Sobolev operators. Result 1 is crucial in our subsequent analysis. Elements in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ are characterized as operators whose polynomial expansion coefficients with respect to the (infinite-dimensional) Hermite polynomials $\{H_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma}\in\Gamma}$, with countable index set Γ, are weighted ℓ^2-summable. The corresponding weights $\boldsymbol{u}=(u_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma}$ are given in terms of the ($\boldsymbol{b}$-weighted) PCA eigenvalues $\lambda_{\boldsymbol{b},i}$ (of the covariance operator) of μ. As a result, we can study the approximation of a given Lipschitz operator by considering its Hermite polynomial expansions.

2. In Section 4, we give upper and lower bounds for the convergence of these expansions in terms of the PCA eigenvalues. In particular, we show the following *curse of parametric complexity*: No s-term Hermite polynomial expansion can converge with an algebraic rate uniformly for all Lipschitz operators as $s\to\infty$. This holds regardless of the decay rate of the PCA eigenvalues. More specifically, let $S\subset\Gamma$ be a finite index set with at most s elements and let F_S denote the polynomial approximation of an operator $F\in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ by Hermite polynomials $H_{\boldsymbol{\gamma}}$ with $\boldsymbol{\gamma}\in S$. Moreover, let $\pi:\mathbb{N}\to\Gamma$ be a bijection such that $(u_{\boldsymbol{\pi(i)}})_{i\in\mathbb{N}}$ is a nonincreasing rearrangement of $\boldsymbol{u}$. Our second main contribution is the following result:

Result 2 (Curse of parametric complexity, cf. Thms. 4.1, 4.6, 4.7)**.** *For every $s \in \mathbb{N}$, we have*

$$\inf_{S \subset \Gamma, |S| \leq s} \sup_F \|F - F_S\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} = \sup_F \left\|F - F_{\{\boldsymbol{\pi}(\mathbf{1}),\ldots,\boldsymbol{\pi}(\boldsymbol{s})\}}\right\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} = u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})},$$

where the suprema are taken either over the class of all Lipschitz operators or all $W^{1,2}_{\mu,\boldsymbol{b}}$-operators with $\|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} \leq 1$. Moreover, $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$ cannot decay with algebraic order as $s \to \infty$, regardless of the spectral properties of μ. On the positive side, the decay of $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$ can become arbitrarily close to any algebraic rate if the PCA eigenvalues $\lambda_{\boldsymbol{b},i}$ decay sufficiently fast (e.g., double exponentially).

As we will discuss in Section 4.4, this curse of parametric complexity is consistent with the one from Lanthaler (2023), which we mentioned in Section 1.1.

3. Up to this point, we study best polynomial approximation of Lipschitz operators. In Section 5, we adopt a more general view and consider learning Lipschitz and $W^{1,2}_{\mu,\boldsymbol{b}}$-operators from finitely many arbitrary linear samples. Using tools from information-based complexity (Novak and Woźniakowski, 2008), we define the adaptive m-width $\Theta_m(\mathcal{K})$ of a set $\mathcal{K}$ of operators. It quantifies the best worst-case error that can be achieved by learning operators in $\mathcal{K}$ from m measurements $\mathcal{L}(F) \in \mathcal{Y}^m$ which are generated by an adaptive sampling operator $\mathcal{L}$ and used as inputs for an arbitrary (generally nonlinear) reconstruction map $\mathcal{T} : \mathcal{Y}^m \to L^2_\mu(\mathcal{X};\mathcal{Y})$. Specifically

$$\Theta_m(\mathcal{K}) := \inf\left\{\sup_{F \in \mathcal{K}} \|F - \mathcal{T}(\mathcal{L}(F))\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} : \mathcal{L}, \mathcal{T} \text{ as above}\right\}.$$

Our key result is the characterization of $\Theta_m(\mathcal{K})$ in terms of the weights $u_{\boldsymbol{\gamma}}$:

Result 3 (Characterization of the adaptive m-width, cf. Thms. 5.4, 4.6, 4.7)**.** *For any number of samples $m \in \mathbb{N}$, we have*

$$\Theta_m(\mathcal{K}) = u_{\boldsymbol{\pi}(\boldsymbol{m}+\mathbf{1})},$$

where $\mathcal{K}$ is either the set of all Lipschitz or all $W^{1,2}_{\mu,\boldsymbol{b}}$-operators of at most unit $W^{1,2}_{\mu,\boldsymbol{b}}$-norm. Again, $u_{\boldsymbol{\pi}(\boldsymbol{m}+\mathbf{1})}$ cannot decay with algebraic order as $m \to \infty$, regardless of the spectral properties of μ. But its decay can become arbitrarily close to any algebraic rate if the PCA eigenvalues $\lambda_{\boldsymbol{b},i}$ of μ decay sufficiently fast (e.g., double exponentially).

This result tightly characterizes the sample complexity of learning Lipschitz operators and gives rise to the following *curse of sample complexity*: No procedure (e.g., NNs, polynomial approximation, random features, kernel methods, etc.) for learning Lipschitz operators can achieve algebraic convergence rates for the worst-case L^2_μ-approximation error. This holds for general (centered, nondegenerate) Gaussian measures μ. In light of Result 3, note that Result 2 shows that Hermite polynomial approximation is optimal among all possible (adaptive) sampling and reconstruction operators for approximating Lipschitz and $W^{1,2}_{\mu,\boldsymbol{b}}$-operators. Finally, we emphasize that Result 3 implies that from an information-based complexity point of view, there is no difference between the space $\text{Lip}(\mathcal{X}, \mathcal{Y})$, equipped with the $W^{1,2}_{\mu,\boldsymbol{b}}$-norm, and the whole space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ as the adaptive m-widths of the unit balls in both spaces coincide.

1.3. *Limitations and future work*

In this work, we do not consider reconstruction strategies based on i.i.d. pointwise samples but based on general linear information. In practice, however, recovery from pointwise samples is most relevant as it is non-intrusive. While the inherent curse of sample complexity cannot be overcome (see also below), it is of key interest to design algorithms with (near-)optimal sample complexity which achieve an optimal approximation error with the minimum amount of operator evaluations. Moreover, in practice, only finite-dimensional data is available, which is usually corrupted by noise. Hence, encoding and decoding errors and errors due to noisy observations arise. We aim to address this problem in future work by studying the approximation of Lipschitz operators from noisy pointwise samples with methods from machine learning, e.g., in terms of practical existence theorems, as studied in Adcock et al. (2024b,d, 2025a); Franco and Brugiapaglia (2025), or with techniques from statistical learning theory (see, e.g., Schmidt-Hieber (2020); Liu et al. (2024)).

On the other hand, by tightly characterizing the curse of sample complexity, our results suggest that learning Lipschitz operators with arbitrary Gaussian measures may be practically challenging. For instance, if the PCA eigenvalues decay algebraically, our characterization implies at best polylogarithmic decay of the error in m of *any* learning algorithm—a rate that may be too slow in practice. Extending and corroborating previous studies (e.g., Kovachki et al. (2024a); Lanthaler and Stuart (2024)), our results suggest that the class of Lipschitz operators may simply be too large. By contrast, the class of holomorphic operators is known to be too small since operators found in various applications are non-holomorphic. Hence, our results highlight another important problem for future work, namely, the search for spaces that better describe classes of operators found in applications.

2. Preliminaries

We recall some standard notions and fix the notation which we use throughout the text. Further notation will be introduced in the text as needed.

2.1. *Basic notation*

As usual, $\mathbb{N}$ denotes the set of all positive integers, $\mathbb{N}_0$ the set of all nonnegative integers, and $\mathbb{R}$ the real numbers. We denote by $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$ and $(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}})$ two separable Hilbert spaces with corresponding inner products. For simplicity, we focus on real Hilbert spaces, but all results can be readily generalized to the complex case as well. We use capital letters $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ (at times also H, K, Z) for elements of the respective Hilbert spaces. Operators which map from $\mathcal{X}$ to $\mathcal{Y}$ are typically denoted by the capital letters F or G. For functionals, i.e., in the case $\mathcal{Y} = \mathbb{R}$, we also use lower case letters at times. The space of all continuous operators from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $C(\mathcal{X},\mathcal{Y})$ and we write $C(\mathcal{X}) := C(\mathcal{X},\mathbb{R})$. The space of all Hilbert-Schmidt operators from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $(HS(\mathcal{X},\mathcal{Y}), \langle\cdot,\cdot\rangle_{HS(\mathcal{X},\mathcal{Y})})$. It is again a separable Hilbert space with induced norm $\|F\|_{HS(\mathcal{X},\mathcal{Y})} := \langle F, F\rangle_{HS(\mathcal{X},\mathcal{Y})}^{1/2}$.

We equip $\mathcal{X}$ with a centered, nondegenerate Gaussian measure μ with covariance operator $Q := \int_{\mathcal{X}} X \otimes X \, d\mu(X)$. We recall that $Q : \mathcal{X} \to \mathcal{X}$ is positive definite, self-adjoint and trace class (Da Prato, 2006, Prop. 1.8). As such, there exists an orthonormal eigenbasis (PCA basis) $\{\phi_i\}_{i\in\mathbb{N}}$ of $\mathcal{X}$ and a sequence of corresponding PCA eigenvalues $\boldsymbol{\lambda} = (\lambda_i)_{i\in\mathbb{N}}$. To be explicit, we have $Q\phi_i = \lambda_i \phi_i$ with $\lambda_i > 0$ for every $i \in \mathbb{N}$ and $\sum_{i=1}^{\infty} \lambda_i < \infty$. We denote the standard Gaussian measure on $\mathbb{R}$ by $\mu_1 := \mathcal{N}(0,1)$, the standard Gaussian measure on $\mathbb{R}^n$, $n \in \mathbb{N}$, by $\mu_n := \bigotimes_{i=1}^n \mu_1$, and the standard Gaussian measure on the space of sequences $\mathbb{R}^{\mathbb{N}}$ by $\mu_\infty := \bigotimes_{i=1}^\infty \mu_1$.

For a positive integer N, we use the shorthand notation $[N]$ for the set $\{1,2,\dots,N\}$ and also write $[\infty] := \mathbb{N}$. Sequences of real numbers with (possibly finite) index set I are denoted by lower case bold letters $\boldsymbol{x} = (x_i)_{i\in I} \in \mathbb{R}^I$. Sequences with elements in a Hilbert space $\mathcal{Z}$ are denoted by bold capital letters $\boldsymbol{Z} = (Z_i)_{i\in I} \in \mathcal{Z}^I$. We write $\mathbf{0}$ and $\mathbf{1}$ for the constant zero and constant one sequence, respectively. Algebraic operations on a sequence $\boldsymbol{x} \in \mathbb{R}^I$ are defined componentwise: We write $\sqrt{\boldsymbol{x}} := (\sqrt{x_i})_{i\in I}$ and $1/\boldsymbol{x} := (1/x_i)_{i\in I}$, whenever these expressions make sense. Given a scalar $c \in \mathbb{R}$, we write $c\boldsymbol{x} := (cx_i)_{i\in I}$ for the scaled sequence. In a similar vein, inequalities of the form $\boldsymbol{x} \le \boldsymbol{y}$ between sequences $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^I$ are understood componentwise, that is, $x_i \le y_i$ for every $i \in I$. The expressions $\boldsymbol{x} \ge \boldsymbol{y}$, $\boldsymbol{x} < \boldsymbol{y}$, and $\boldsymbol{x} > \boldsymbol{y}$ are understood in a similar sense. Given $J \subset I$, we write $\boldsymbol{x}_J$ for the subsequence $(x_i)_{i\in J}$. We write $\delta_{i,j}$ for the Dirac delta function (i.e. $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise), and for two sequences $\boldsymbol{\gamma},\boldsymbol{\gamma'} \in \mathbb{N}_0^{\mathbb{N}}$, we set

$$\delta_{\boldsymbol{\gamma},\boldsymbol{\gamma'}} := \prod_{i=1}^\infty \delta_{\gamma_i,\gamma_i'}.$$

We denote the Euclidean norm and inner product on $\mathbb{R}^n$ by $\|\cdot\|_{\mathbb{R}^n}$ and $\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$, respectively, and the standard basis vectors by $\boldsymbol{e_i} := (\delta_{i,j})_{j\in[n]}$ for $i \in [n]$. The symbol $\mathcal{L}^n$ denotes the Lebesgue measure on $\mathbb{R}^n$. Finally, we use the notation $x \lesssim y$ for $x,y \in \mathbb{R}$ if there exists a global constant $C > 0$, independent of any parameters, such that $x \le Cy$. We write $x \gtrsim y$ if $y \lesssim x$, and $x \sim y$ if both $x \lesssim y$ and $x \gtrsim y$.

2.2. *Sequence spaces*

Let $1 \le p \le \infty$. Given an index set I, positive weights $\boldsymbol{w} = (w_i)_{i\in I} > \mathbf{0}$, and a Hilbert space $\mathcal{Z}$, we define the space $\ell^p_{\boldsymbol{w}}(I;\mathcal{Z})$ as the set of all $\mathcal{Z}$-valued sequences $\boldsymbol{Z} = (Z_i)_{i\in I}$ whose norm $\|\boldsymbol{Z}\|_{\ell^p_{\boldsymbol{w}}(I;\mathcal{Z})}$ is finite, where

$$\|\boldsymbol{Z}\|_{\ell^p_{\boldsymbol{w}}(I;\mathcal{Z})} := \begin{cases} \left(\sum_{i\in I} w_i^{-p}\|Z_i\|_{\mathcal{Z}}^p\right)^{1/p} & \text{if } 1 \le p < \infty, \\ \sup_{i\in I}\left\{w_i^{-1}\|Z_i\|_{\mathcal{Z}}\right\} & \text{if } p = \infty. \end{cases}$$

We denote the closed unit ball in $\ell^p_{\boldsymbol{w}}(I;\mathcal{Z})$ by

$$B^p_{\boldsymbol{w}}(I;\mathcal{Z}) := \left\{\boldsymbol{x} \in \ell^p_{\boldsymbol{w}}(I;\mathcal{Z}) : \|\boldsymbol{x}\|_{\ell^p_{\boldsymbol{w}}(I;\mathcal{Z})} \le 1\right\}.$$

If $\mathcal{Z} = \mathbb{R}$, we write $(\ell^p_{\boldsymbol{w}}(I), \|\cdot\|_{\ell^p_{\boldsymbol{w}}(I)})$ and $B^p_{\boldsymbol{w}}(I)$, and if $\boldsymbol{w} = \mathbf{1}$, we write $(\ell^p(I;\mathcal{Z}), \|\cdot\|_{\ell^p(I;\mathcal{Z})})$ and $B^p(I;\mathcal{Z})$.

2.3. *The weighted space $\mathcal{X}_{\boldsymbol{b}}$*

Let $\boldsymbol{b} = (b_i)_{i\in\mathbb{N}}$ be a sequence of positive weights with $\mathbf{0} < \boldsymbol{b} \le \mathbf{1}$. By means of the PCA basis $\{\phi_i\}_{i\in\mathbb{N}}$ of $\mathcal{X}$ we define the space

$$\mathcal{X}_{\boldsymbol{b}} := \left\{X \in \mathcal{X} : \sum_{i=1}^\infty b_i^{-2}\left|\langle X,\phi_i\rangle_{\mathcal{X}}\right|^2 < \infty\right\}.$$

Note that $\mathcal{X}_{\boldsymbol{b}}$ is a Hilbert subspace of $\mathcal{X}$ with inner product

$$\langle X, Z\rangle_{\mathcal{X}_{\boldsymbol{b}}} := \sum_{i=1}^{\infty} b_i^{-2} \langle X, \phi_i\rangle_{\mathcal{X}} \langle Z, \phi_i\rangle_{\mathcal{X}}, \quad X, Z \in \mathcal{X}_{\boldsymbol{b}},$$

which induces the norm

$$\|X\|_{\mathcal{X}_{\boldsymbol{b}}} := \sqrt{\langle X, X\rangle_{\mathcal{X}_{\boldsymbol{b}}}}, \quad X \in \mathcal{X}_{\boldsymbol{b}}.$$

Moreover, it is easy to see that $\mathcal{X}_{\boldsymbol{b}}$ has an orthonormal basis $\{\eta_i\}_{i\in\mathbb{N}}$, given by

$$\eta_i := b_i \phi_i, \quad i \in \mathbb{N}. \tag{2.1}$$

Remark 2.1. *We highlight two important cases. If $\boldsymbol{b} = \mathbf{1}$, we recover the full space $\mathcal{X}_{\mathbf{1}} = \mathcal{X}$. If $\boldsymbol{b} = \sqrt{\boldsymbol{\lambda}}$, we obtain the Cameron-Martin space $\mathcal{X}_{\sqrt{\boldsymbol{\lambda}}} = \mathcal{H}$ of μ, see (C.1).*

2.4. *Lebesgue-Bochner spaces and Hermite polynomials*

We write $L^2_\mu(\mathcal{X};\mathcal{Y})$ for the Lebesgue-Bochner space of (equivalence classes of) strongly measurable operators $F : \mathcal{X} \to \mathcal{Y}$ with finite Bochner norm

$$\|F\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} := \left(\int_{\mathcal{X}} \|F(X)\|^2_{\mathcal{Y}} d\mu(X) \right)^{1/2}.$$

If $\mathcal{Y} = \mathbb{R}$, the Lebesgue-Bochner space $L^2_\mu(\mathcal{X};\mathbb{R})$ coincides with the usual Lebesgue space and we write $L^2_\mu(\mathcal{X}) := L^2_\mu(\mathcal{X};\mathbb{R})$. See, e.g., Hytönen et al. (2016, Chpt. 1), for more information.

Next, we introduce the (infinite-dimensional) Hermite polynomials. For $n \in \mathbb{N}_0$, we first define the nth normalized (probabilist's) Hermite polynomial on $\mathbb{R}$ by

$$H_n : \mathbb{R} \to \mathbb{R}, \quad H_n(x) := \frac{(-1)^n}{\sqrt{n!}} \exp\left(\frac{x^2}{2}\right) \frac{d^n}{dx^n} \exp\left(-\frac{x^2}{2}\right).$$

The family $\{H_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $L^2_{\mu_1}(\mathbb{R})$ (Da Prato, 2006, Prop. 9.4). We now define the higher-dimensional Hermite polynomials as products of the one-dimensional ones. To this end, we introduce the set of all sequences of nonnegative integers with finite support,

$$\Gamma := \left\{ \boldsymbol{\gamma} \in \mathbb{N}_0^{\mathbb{N}} : \mathrm{supp}(\boldsymbol{\gamma}) < \infty \right\},$$

with the support of $\boldsymbol{\gamma}$ defined as $\mathrm{supp}(\boldsymbol{\gamma}) := \{i \in \mathbb{N} : \gamma_i \neq 0\}$. It is easy to see that Γ is countable. For $\boldsymbol{\gamma} \in \Gamma$ and $d \in \mathbb{N}$, we set

$$H_{\boldsymbol{\gamma},d} : \mathbb{R}^d \to \mathbb{R}, \quad H_{\boldsymbol{\gamma},d}(\boldsymbol{x}) := \prod_{i=1}^{d} H_{\gamma_i}(x_i). \tag{2.2}$$

Remark 2.2. *Since $\boldsymbol{\gamma} \in \Gamma$ has finite support and $H_0 = 1$, each Hermite polynomial $H_{\boldsymbol{\gamma},d}$ can also be seen as a function $H_{\boldsymbol{\gamma},\infty}$ with infinite-dimensional input $\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}}$ (simply by ignoring all x_i with $i \notin \mathrm{supp}(\boldsymbol{\gamma})$).*

Finally, we define Hermite polynomials on the infinite-dimensional space $\mathcal{X}$ by means of the PCA basis $\{\phi_i\}_{i\in\mathbb{N}}$ and the PCA eigenvalues $\boldsymbol{\lambda} = (\lambda_i)_{i\in\mathbb{N}}$:

$$H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} : \mathcal{X} \to \mathbb{R}, \quad H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}(X) := \prod_{i=1}^{\infty} H_{\gamma_i}\left(\frac{\langle X, \phi_i\rangle_{\mathcal{X}}}{\sqrt{\lambda_i}}\right). \tag{2.3}$$

As only finitely many factors in (2.3) are different from 1, each $H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}$ is a smooth function on $\mathcal{X}$ with polynomial growth at infinity. Fernique's theorem (Theorem C.1) then implies that $H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} \in L^2_\mu(\mathcal{X})$ for every $\boldsymbol{\gamma} \in \Gamma$. Similarly as in finite dimensions, the family $\{H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}\}_{\boldsymbol{\gamma}\in\Gamma}$ of infinite-dimensional Hermite polynomials is an orthonormal basis of $L^2_\mu(\mathcal{X})$ (Da Prato, 2006, Thm. 9.7). Recalling that $L^2_\mu(\mathcal{X};\mathcal{Y}) = L^2_\mu(\mathcal{X}) \otimes \mathcal{Y}$ with Hilbertian tensor product, any $F \in L^2_\mu(\mathcal{X};\mathcal{Y})$ can thus be written as an unconditionally $L^2_\mu(\mathcal{X};\mathcal{Y})$-convergent expansion in Hermite polynomials, also called the *Wiener-Hermite Polynomial Chaos (PC) expansion*,

$$F = \sum_{\boldsymbol{\gamma}\in\Gamma} Y_{\boldsymbol{\gamma}} H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} \quad \text{with} \quad Y_{\boldsymbol{\gamma}} := \int_{\mathcal{X}} F(X) H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}(X)\, d\mu(X) \in \mathcal{Y}. \tag{2.4}$$

The $Y_{\boldsymbol{\gamma}}$ are called the *Wiener-Hermite PC coefficients* of F. Moreover, *Parseval's identity* holds, that is,

$$\|F\|^2_{L^2_\mu(\mathcal{X};\mathcal{Y})} = \sum_{\boldsymbol{\gamma}\in\Gamma} \|Y_{\boldsymbol{\gamma}}\|^2_{\mathcal{Y}}. \tag{2.5}$$

3. Gaussian Sobolev and Lipschitz operators

Let $\boldsymbol{b}$ be a sequence of weights and $\mathcal{X}_{\boldsymbol{b}}$ the corresponding weighted space as introduced in Section 2.3. We first sketch the definition of the weighted Gaussian Sobolev space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and state its characterization as a weighted ℓ^2-sequence space. Details are provided in Appendix C.2. We then prove that all Lipschitz operators from $\mathcal{X}$ to $\mathcal{Y}$ lie in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ under some sufficient conditions on $\boldsymbol{b}$. The results in this section are the basis for the approximation theoretical analysis carried out in the remainder of this paper.

3.1. *The space* $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$

The definition of the space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ is based on the operator $D_{\mathcal{X}_{\boldsymbol{b}}}$ which denotes the Fréchet differential operator along the space $\mathcal{X}_{\boldsymbol{b}}$ (see Appendix A). We first define $D_{\mathcal{X}_{\boldsymbol{b}}}$ as an operator which maps from a set $\mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$ of cylindrical boundedly differentiable operators to $L^2_\mu(\mathcal{X}; HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))$. The theory of the Cameron-Martin space $\mathcal{H}$ of μ allows us to show that $D_{\mathcal{X}_{\boldsymbol{b}}}$ is closable in $L^2_\mu(\mathcal{X};\mathcal{Y})$, see Proposition C.5. Details about the closability and closure of operators are recalled in Appendix B.

Definition 3.1 (The Sobolev space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$)**.** We define the space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ as the domain of the closure of the operator $D_{\mathcal{X}_{\boldsymbol{b}}} : \mathcal{F}C^1_b(\mathcal{X},\mathcal{Y}) \to L^2_\mu(\mathcal{X}; HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))$ in $L^2_\mu(\mathcal{X};\mathcal{Y})$ (still denoted by $D_{\mathcal{X}_{\boldsymbol{b}}}$).

The space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ is a Hilbert space with the graph norm

$$\|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} := \left(\int_{\mathcal{X}} \|F(X)\|^2_{\mathcal{Y}} d\mu(X) + \int_{\mathcal{X}} \left\|D_{\mathcal{X}_{\boldsymbol{b}}} F(X)\right\|^2_{HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y})} d\mu(X)\right)^{1/2},$$

which is induced by the inner product

$$\langle F, G\rangle_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} := \int_{\mathcal{X}} \langle F(X), G(X)\rangle_{\mathcal{Y}} d\mu(X) + \int_{\mathcal{X}} \left\langle D_{\mathcal{X}_{\boldsymbol{b}}} F(X), D_{\mathcal{X}_{\boldsymbol{b}}} G(X)\right\rangle_{HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y})} d\mu(X).$$

As usual, we write $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X}) := W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathbb{R})$ for the space of Sobolev functionals.

Remark 3.2. *Defining Gaussian Sobolev spaces as the domain of the closure of a suitable differential operator is standard. See, e.g., in Chojnowska-Michalik and Goldys (2001); Da Prato (2006, 2014); Lunardi et al. (2015/2016). For an equivalent definition via the completion of $\mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$ under an appropriate Sobolev norm, see Bogachev (1998, Chpt. 5).*

Remark 3.3. *Weighted Gaussian Sobolev spaces have been considered in the literature in the study of continuous and compact Sobolev embeddings (Shigekawa, 1992; Da Prato and Zabczyk, 1995; Chojnowska-Michalik and Goldys, 2001) by composing the differential operator with an additional self-adjoint nonnegative operator. Recently, in Luo et al. (2023), the authors defined weighted Gaussian Sobolev spaces of functionals on $\ell^r(\mathbb{N})$, $r \geq 1$, by weighting the partial derivatives with elements of a weight sequence $\boldsymbol{b} \in \ell^\infty(\mathbb{N})$ and remarked that their construction unifies various definitions of Gaussian Sobolev spaces via different choices of $\boldsymbol{b}$. Our definition of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X})$ is equivalent to this construction in the Hilbert space case $r = 2$. However, our approach via the differential operator along the space $\mathcal{X}_{\boldsymbol{b}}$ highlights the role of the latter as the underlying differential structure of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X})$. For $\boldsymbol{b} = \sqrt{\boldsymbol{\lambda}}$, we obtain the same space as defined in Bogachev (1998); Lunardi et al. (2015/2016); Da Prato (2014). For $\boldsymbol{b} = \boldsymbol{1}$, we obtain a space as defined in Da Prato (2006), which is smaller.*

Next, we use the Wiener-Hermite PC expansion (2.4) of $L^2_\mu(\mathcal{X};\mathcal{Y})$-operators to characterize the space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ by a weighted ℓ^2-space.

Definition 3.4 (Weighted PCA eigenvalues)**.** The $\boldsymbol{b}$*-weighted PCA eigenvalues* are given by

$$\boldsymbol{\lambda}_{\boldsymbol{b}} = (\lambda_{\boldsymbol{b},i})_{i\in\mathbb{N}}, \quad \lambda_{\boldsymbol{b},i} := \lambda_i / b_i^2. \tag{3.1}$$

The following theorem is an immediate consequence of Proposition C.7.

Theorem 3.5 (ℓ^2-characterization of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$)**.** *The map*

$$\ell^2_{\boldsymbol{u}}(\Gamma;\mathcal{Y}) \to W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}), \quad \boldsymbol{Y} = (Y_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma} \mapsto \sum_{\boldsymbol{\gamma}\in\Gamma} Y_{\boldsymbol{\gamma}} H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}$$

with the family of weights

$$\boldsymbol{u} = (u_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma}, \quad u_{\boldsymbol{\gamma}} = u_{\boldsymbol{\gamma}}(\boldsymbol{\lambda}_{\boldsymbol{b}}) := \left(1 + \sum_{i=1}^{\infty} \frac{\gamma_i}{\lambda_{\boldsymbol{b},i}}\right)^{-1/2}, \tag{3.2}$$

is an isometric isomorphism. In particular, using the representation (2.4)*, we have*

$$\|F\|^2_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} = \sum_{\boldsymbol{\gamma}\in\Gamma} u_{\boldsymbol{\gamma}}^{-2} \left\| \int_{\mathcal{X}} F H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu \right\|^2_{\mathcal{Y}}, \quad \forall F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}).$$

Assumption 3.6 (Properties of $\boldsymbol{b}$)**.** *The sequence $\boldsymbol{b} = (b_i)_{i\in\mathbb{N}}$ is a sequence of positive real numbers with $\mathbf{0} < \boldsymbol{b} \leq \mathbf{1}$ such that the sequence of weighted PCA eigenvalues $\boldsymbol{\lambda_b} = (\lambda_{\boldsymbol{b},i})_{i\in\mathbb{N}}$, defined in* (3.1)*, is nonincreasing. If* $\dim(\mathcal{Y}) = \infty$*, we assume in addition that $\boldsymbol{b} \in \ell^2(\mathbb{N})$.*

The weights $u_{\boldsymbol{\gamma}}$ in (3.2) are key to our analysis. By Assumption 3.6, we can them in a nonincreasing way, that is, there exists a *nonincreasing rearrangement* $\pi : \mathbb{N} \to \Gamma$ of $\boldsymbol{u}$ such that

$$u_{\boldsymbol{\pi(1)}} \geq u_{\boldsymbol{\pi(2)}} \geq \cdots > 0. \tag{3.3}$$

The map π is unique up to permutations of weights of the same value. The additional requirement of ℓ^2-summability of $\boldsymbol{b}$ in Assumption 3.6 in the case $\dim(\mathcal{Y}) = \infty$ will become clear by Theorem 3.10. It states that the set of Lipschitz operators is a subset of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$.

Remark 3.7. *If $\mathcal{Y}$ is finite-dimensional, we can choose $\boldsymbol{b} = \mathbf{1}$, which gives $\boldsymbol{\lambda_b} = \boldsymbol{\lambda}$. This is not possible if $\mathcal{Y}$ is infinite-dimensional. However, since $\boldsymbol{\lambda} \in \ell^1(\mathbb{N})$, a valid choice for $\boldsymbol{b}$ in any case is $\boldsymbol{b} = \sqrt{\boldsymbol{\lambda}}$, which leads to $\boldsymbol{\lambda_b} = \mathbf{1}$ and thus to no decay of the $\lambda_{\boldsymbol{b},i}$ at all.*

Remark 3.8. *Assumption 3.6 implies that $\limsup_{i\to\infty} \lambda_{\boldsymbol{b},i} < \infty$. Interestingly, this condition is equivalent to the continuous embedding of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X})$ in the Orlicz space $L^p \log^{\frac{p}{2}} L(\mathcal{X},\mu)$ for $p \in [1,\infty)$ (Luo et al., 2023, Thm. 4.2). The stronger condition $\lim_{i\to\infty} \lambda_{\boldsymbol{b},i} = 0$ is equivalent to the compact embedding of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X})$ in $L^2_\mu(\mathcal{X})$ (Da Prato and Zabczyk, 1995, Prop. 2.2), and, more generally, in the Orlicz space $L^2 \log^q L(\mathcal{X},\mu)$ for $q \in [0,1)$ (Luo et al., 2023, Thm. 5.2).*

3.2. *Lipschitz operators*

We now turn to Lipschitz continuous operators and recall their definition.

Definition 3.9 (Lipschitz operators)**.** An operator $F : \mathcal{X} \to \mathcal{Y}$ is called *(L-)Lipschitz (continuous)* if there exists a constant $L > 0$ such that

$$\|F(X) - F(Z)\|_{\mathcal{Y}} \leq L\|X - Z\|_{\mathcal{X}}, \quad \forall X, Z \in \mathcal{X}.$$

The number L is a *Lipschitz constant* of F. The smallest Lipschitz constant of F is given by

$$[F]_{\mathrm{Lip}(\mathcal{X},\mathcal{Y})} := \sup_{\substack{X,Z\in\mathcal{X}\\ X\neq Z}} \frac{\|F(X) - F(Z)\|_{\mathcal{Y}}}{\|X - Z\|_{\mathcal{X}}}.$$

We denote the space of all Lipschitz operators from $\mathcal{X}$ to $\mathcal{Y}$ by $\mathrm{Lip}(\mathcal{X},\mathcal{Y})$ and write $\mathrm{Lip}(\mathcal{X}) := \mathrm{Lip}(\mathcal{X};\mathbb{R})$. We further define the space of all *bounded* Lipschitz operators $C^{0,1}(\mathcal{X};\mathcal{Y})$ as the set of all Lipschitz operators $F \in \mathrm{Lip}(\mathcal{X},\mathcal{Y})$ with finite norm

$$\|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})} := \sup_{X\in\mathcal{X}} \|F(X)\|_{\mathcal{Y}} + [F]_{\mathrm{Lip}(\mathcal{X},\mathcal{Y})}.$$

Note that $C^{0,1}(\mathcal{X},\mathcal{Y})$ is a strict subset of $\mathrm{Lip}(\mathcal{X},\mathcal{Y})$ as operators in $\mathrm{Lip}(\mathcal{X},\mathcal{Y})$ are not necessarily bounded. The next result, which is the main result of this section, motivates Gaussian Sobolev spaces as a natural setting for the study of Lipschitz operators. We present a sketch of the proof, highlighting the main ideas. A detailed proof is given in Appendix D.

Theorem 3.10 (Lipschitz operators are Gaussian Sobolev operators)**.** *Let $\boldsymbol{b}=(b_i)_{i\in\mathbb{N}}$ be a sequence of positive numbers with $\boldsymbol{0}<\boldsymbol{b}\le\boldsymbol{1}$.*

(i) If $\mathcal{Y}$ is finite-dimensional, then $\mathrm{Lip}(\mathcal{X},\mathcal{Y})\subset W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and

$$\left\|D_{\mathcal{X}_{\boldsymbol{b}}}F\right\|_{L^2_\mu(\mathcal{X}_{\boldsymbol{b}};HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))}\le\sqrt{\dim(\mathcal{Y})}\cdot[F]_{\mathrm{Lip}},\quad\forall F\in\mathrm{Lip}(\mathcal{X},\mathcal{Y}).$$

In particular, the embedding $C^{0,1}(\mathcal{X},\mathcal{Y})\hookrightarrow W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ is continuous with

$$\|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})}\le\sqrt{\dim(\mathcal{Y})}\cdot\|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})},\quad\forall F\in C^{0,1}(\mathcal{X},\mathcal{Y}).$$

(ii) If $\mathcal{Y}$ is infinite-dimensional and if $\boldsymbol{b}\in\ell^2(\mathbb{N})$, then $\mathrm{Lip}(\mathcal{X},\mathcal{Y})\subset W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and

$$\left\|D_{\mathcal{X}_{\boldsymbol{b}}}F\right\|_{L^2_\mu(\mathcal{X}_{\boldsymbol{b}};HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))}\le\|\boldsymbol{b}\|_{\ell^2(\mathbb{N})}\cdot[F]_{\mathrm{Lip}},\quad\forall F\in\mathrm{Lip}(\mathcal{X},\mathcal{Y}).$$

In particular, the embedding $C^{0,1}(\mathcal{X},\mathcal{Y})\hookrightarrow W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ is continuous with

$$\|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})}\le\max\left\{1,\|\boldsymbol{b}\|_{\ell^2(\mathbb{N})}\right\}\cdot\|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})},\quad\forall F\in C^{0,1}(\mathcal{X},\mathcal{Y}).$$

Proof (Sketch) Let $F\in\mathrm{Lip}(\mathcal{X},\mathcal{Y})$. In order to show that F lies in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, it suffices to find a sequence $(F_n)_{n\in\mathbb{N}}$ of operators which converge to F in $L^2_\mu(\mathcal{X};\mathcal{Y})$ and which are uniformly bounded in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ (see Lemma C.6). To this end, we construct a specific sequence of operators of the form $F_n=V_n\circ T_n$ with $V_n:\mathbb{R}^n\to\mathcal{Y}$ and $T_n:\mathcal{X}\to\mathbb{R}^n$ which converge to F in $L^2_\mu(\mathcal{X};\mathcal{Y})$ and, in fact, in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. It can be shown that the operator V_n inherits Lipschitz continuity of F. Hence, by Rademacher's theorem, V_n is differentiable $\mathcal{L}^n$-almost everywhere. We are now left with showing that F_n is differentiable μ-almost everywhere and that there exists a constant $C>0$ such that

$$\left\|D_{\mathcal{X}_{\boldsymbol{b}}}F_n(X)\right\|_{HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y})}\le C \tag{3.4}$$

for all $n\in\mathbb{N}$ and μ-almost every $X\in\mathcal{X}$.

In case (i), let $m:=\dim(\mathcal{Y})$. We fix an orthonormal basis $\{\psi_j\}_{j\in[m]}$ of $\mathcal{Y}$ and consider the (Lipschitz continuous) coordinate functions $V_n^{(j)}:=\langle V_n,\psi_j\rangle_{\mathcal{Y}}:\mathbb{R}^n\to\mathbb{R}$. Unwinding definitions, a straight-forward calculation then leads to (3.4) for any weight sequence $\boldsymbol{0}<\boldsymbol{b}\le\boldsymbol{1}$. The resulting constant C is given by $\sqrt{\dim(\mathcal{Y})}\cdot[F]_{\mathrm{Lip}}$. The same argument does not work in case (ii) where $\mathcal{Y}$ has infinite dimension. At this point, we additionally require that $\boldsymbol{b}\in\ell^2(\mathbb{N})$. A slight modification of the argument in (i) then yields (3.4) with C given by $\|\boldsymbol{b}\|_{\ell^2(\mathbb{N})}\cdot[F]_{\mathrm{Lip}}$.

The continuous embedding of $C^{0,1}(\mathcal{X},\mathcal{Y})$ in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ follows in both cases from (3.4) and the fact that $F_n \to F$ in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. $\square$

In light of Theorem 3.10, note that Assumption 3.6 implies that $\mathrm{Lip}(\mathcal{X},\mathcal{Y}) \subset W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ regardless of whether $\mathcal{Y}$ is finite- or infinite-dimensional.

Remark 3.11. *For functionals, i.e., in the case $\mathcal{Y}=\mathbb{R}$, it is well-known that Lipschitz continuity implies Gaussian Sobolev regularity. We refer to Da Prato (2006, Prop. 10.11) and Da Prato (2014, Prop. 3.18), where it is shown that* $\mathrm{Lip}(\mathcal{X}) \subset W^{1,2}_{\mu,\boldsymbol{1}}(\mathcal{X})$ *and* $\mathrm{Lip}(\mathcal{X}) \subset W^{1,2}_{\mu,\sqrt{\boldsymbol{\lambda}}}(\mathcal{X})$*, respectively.*

Remark 3.12. *One can define Gaussian Sobolev spaces $W^{1,p}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ for any order $1 \le p < \infty$ and a proof analogous to the one of Theorem 3.10 shows that they contain* $\mathrm{Lip}(\mathcal{X},\mathcal{Y})$ *as a subset. For the case $\mathcal{Y}=\mathbb{R}$, we mention Bogachev (1998, Ex. 5.4.10(i)) and Lunardi et al. (2015/2016, Prop. 10.1.4). However, only in the Hilbert space case $p=2$ there is a simple characterization of $W^{1,p}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ in terms of a weighted ℓ^2-space as given by Theorem 3.5.*

Finally, we introduce the following notation:

Definition 3.13 (Sobolev unit (Lipschitz) ball)**.** We define the *Sobolev unit ball* and the *Sobolev unit Lipschitz ball* as, respectively,

$$B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}) := \left\{ F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}) : \|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} \le 1 \right\},$$

$$B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}) := \left\{ F \in \mathrm{Lip}(\mathcal{X},\mathcal{Y}) : \|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} \le 1 \right\}.$$

4. Polynomial s-term approximation

The ℓ^2-characterization of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ via Wiener-Hermite PC expansions (see Theorem 3.5) motivates studying polynomial s-term approximations of $W^{1,2}_{\mu,\boldsymbol{b}}$-operators and quantifying the smallest achievable worst-case s-term error. To this end, for any index set $S \subset \Gamma$, we define the space of $\mathcal{Y}$-valued polynomials

$$\mathcal{P}_{S;\mathcal{Y}} := \left\{ \sum_{\boldsymbol{\gamma}\in S} Y_{\boldsymbol{\gamma}} H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} : Y_{\boldsymbol{\gamma}} \in \mathcal{Y} \right\}$$

and the corresponding orthogonal L^2_μ-projection

$$(\cdot)_S : L^2_\mu(\mathcal{X};\mathcal{Y}) \to \mathcal{P}_{S;\mathcal{Y}}, \quad F \mapsto F_S := \sum_{\boldsymbol{\gamma}\in S} \left(\int_{\mathcal{X}} F H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu \right) H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}.$$

Next, let $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and let $S \subset \Gamma$ be finite with $|S| \leq s$. Setting $Y_{\boldsymbol{\gamma}} := \int_{\mathcal{X}} F H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu$, it follows from Parseval's identity (2.5) and Theorem 3.5 that

$$\|F - F_S\|^2_{L^2_\mu(\mathcal{X};\mathcal{Y})} = \sum_{\boldsymbol{\gamma}\in\Gamma\setminus S} \|Y_{\boldsymbol{\gamma}}\|^2_{\mathcal{Y}} \leq \big(\max_{\boldsymbol{\gamma}\in\Gamma\setminus S} u^2_{\boldsymbol{\gamma}}\big) \sum_{\boldsymbol{\gamma}\in\Gamma} u^{-2}_{\boldsymbol{\gamma}} \|Y_{\boldsymbol{\gamma}}\|^2_{\mathcal{Y}} = \big(\max_{\boldsymbol{\gamma}\in\Gamma\setminus S} u^2_{\boldsymbol{\gamma}}\big) \|F\|^2_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})}.$$

Let us recall from (3.3) the nonincreasing rearrangement $\pi : \mathbb{N} \to \Gamma$ of $\boldsymbol{u} = (u_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma}$. For $s \in \mathbb{N}$, we set $S = \pi([s]) = \{\boldsymbol{\pi(1)}, \ldots, \boldsymbol{\pi(s)}\}$ and conclude for $\mathcal{K} \in \{B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}), B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})\}$ that

$$\inf_{S\subset\Gamma,|S|\leq s} \sup_{F\in\mathcal{K}} \|F - F_S\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq \sup_{F\in\mathcal{K}} \|F - F_{\pi([s])}\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq \max_{\boldsymbol{\gamma}\in\Gamma\setminus\pi([s])} u_{\boldsymbol{\gamma}} = u_{\boldsymbol{\pi(s+1)}}. \tag{4.1}$$

Our first main result in this section shows that this chain of inequalities can, in fact, be improved to equality and hence gives a tight characterization of the best polynomial s-term error. The proof is an immediate consequence of Theorem 5.4 in the special case $\mathcal{V} = L^2_\mu(\mathcal{X};\mathcal{Y})$.

Theorem 4.1 (Best polynomial s-term error). *For $\mathcal{K} \in \{B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}), B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})\}$ and every $s \in \mathbb{N}$, we have*

$$\inf_{S\subset\Gamma,|S|\leq s} \sup_{F\in\mathcal{K}} \|F - F_S\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} = \sup_{F\in\mathcal{K}} \|F - F_{\pi([s])}\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} = u_{\boldsymbol{\pi(s+1)}}.$$

Motivated by Theorem 4.1, we study in the rest of this section the decay of $u_{\boldsymbol{\pi(s+1)}}$ as $s \to \infty$. The proofs are based on the relation of the set $\pi([s])$ to an anisotropic total degree index set, which we discuss in Section 4.1. Subsequently, we prove lower and upper bounds for $u_{\boldsymbol{\pi(s+1)}}$ in Sections 4.2 and 4.3.

4.1. *Relation to anisotropic total degree index sets*

We first recall the notion of anisotropic total degree (TD) index sets and provide lower and upper size bounds. We then identify a specific such set to which we can relate $\pi([s])$.

Definition 4.2 (Anisotropic TD index set). For $d \in \mathbb{N}$ and $\boldsymbol{a} = (a_1, \ldots, a_d) \in \mathbb{R}^d$, $\boldsymbol{a} > \boldsymbol{0}$, we define the anisotropic TD index set in d dimensions with weight $\boldsymbol{a}$ by

$$\Lambda^{\mathrm{TD}}_{d,\boldsymbol{a}} := \left\{ \nu \in \mathbb{N}^d_0 : \sum_{i=1}^d a_i \nu_i \leq 1 \right\}.$$

Lemma 4.3 (Lower and upper size bounds for anisotropic TD index sets). *Let $d \in \mathbb{N}$ and $\boldsymbol{a} = (a_1, \ldots, a_d) \in \mathbb{R}^d$ with $0 < a_1 \leq \cdots \leq a_d$. We have*

$$\prod_{i=1}^d \frac{1}{a_i i} \leq |\Lambda^{\mathrm{TD}}_{d,\boldsymbol{a}}| \leq \prod_{i=1}^d \left(\frac{1}{a_i i} + 1 \right). \tag{4.2}$$

Proof The upper bound is Lemma 5.3 in Haji-Ali et al. (2018). The lower bound is proved in Beged-dov (1972). See Griebel and Oettershagen (2016) for further discussion. □

For any $\varepsilon > 0$, we now define the set

$$S(\varepsilon) := \left\{ \boldsymbol{\gamma} \in \Gamma : u_{\boldsymbol{\gamma}}^{-2} \leq 1 + \frac{1}{\varepsilon^2} \right\} = \left\{ \boldsymbol{\gamma} \in \Gamma : \sum_{i=1}^{\infty} \frac{\gamma_i}{\lambda_{\boldsymbol{b},i}} \leq \frac{1}{\varepsilon^2} \right\} \tag{4.3}$$

as well as the quantity

$$d(\varepsilon) := \min \left\{ l \in \mathbb{N} : \lambda_{\boldsymbol{b},l+1} < \varepsilon^2 \right\} \in \mathbb{N} \cup \{\infty\} \tag{4.4}$$

with the convention $\min(\emptyset) = \infty$. By definition, we have $S(\varepsilon) = \pi([|S(\varepsilon)|])$. If $d(\varepsilon)$ is finite, then $S(\varepsilon)$ is isomorphic to an anisotropic TD index set,

$$S(\varepsilon) \cong \Lambda^{\mathrm{TD}}_{d(\varepsilon),\boldsymbol{a}}, \tag{4.5}$$

with weights $a_i := \varepsilon^2/\lambda_{\boldsymbol{b},i}$, $i \in [d(\varepsilon)]$, under the isomorphism

$$\{\boldsymbol{\gamma} \in \Gamma : \mathrm{supp}(\boldsymbol{\gamma}) \subset [d(\varepsilon)]\} \to \mathbb{N}_0^{d(\varepsilon)}, \quad \boldsymbol{\gamma} \mapsto (\gamma_1, \ldots, \gamma_{d(\varepsilon)}).$$

Observe that by Assumption 3.6 we have $0 < a_1 \leq a_2 \leq \cdots \leq a_{d(\varepsilon)}$.

Remark 4.4 (Effective dimension)**.** *The number $d(\varepsilon)$, defined in (4.4), can be interpreted as the* effective dimension *of the approximation problem in the following sense: For $i > d(\varepsilon)$, we have $\lambda_{\boldsymbol{b},i} < \varepsilon^2$, that is, the variance of μ in the ith coordinate direction (w.r.t. the PCA basis $\{\phi_i\}_{i\in\mathbb{N}}$) essentially vanishes for very small values of ε. Consequently, $\mathcal{N}(0, \lambda_{\boldsymbol{b},i}) \approx \delta_0$ for $i > d(\varepsilon)$, where $\mathcal{N}(0, \lambda_{\boldsymbol{b},i})$ denotes the one-dimensional Gaussian measure on $\mathbb{R}$ with mean 0 and variance $\lambda_{\boldsymbol{b},i}$ and δ_0 is the Dirac delta measure centered at 0. Measuring an operator F on $\mathcal{X}$ with respect to μ thus essentially reduces to measuring F in its first $d(\varepsilon)$ coordinates with respect to the Gaussian product measure $\bigotimes_{i=1}^{d(\varepsilon)} \mathcal{N}(0, \lambda_{\boldsymbol{b},i})$.*

Remark 4.5 (Finiteness of $d(\varepsilon)$)**.** *Note that requiring $d(\varepsilon)$ to be finite for every $\varepsilon > 0$ together with Assumption 3.6 implies $\lim_{i\to\infty} \lambda_{\boldsymbol{b},i} = 0$. On the other hand, the limit condition $\lim_{i\to\infty} \lambda_{\boldsymbol{b},i} = 0$ implies Assumption 3.6 after a suitable reordering of the $\lambda_{\boldsymbol{b},i}$ as well as finiteness of $d(\varepsilon)$ for every $\varepsilon > 0$. In this context, we also recall Remark 3.8.*

4.2. *Lower bound*

We now prove the second main result of this section which is a lower bound for $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$. It states that, regardless of the (unweighted) PCA eigenvalues $\boldsymbol{\lambda}$ and choice of $\boldsymbol{b}$, one cannot achieve an algebraic decay of $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$ as $s \to \infty$. In light of Remark 4.5, we emphasize that we do not assume that $\lim_{i\to\infty} \lambda_{\boldsymbol{b},i} = 0$, but only that the $\lambda_{\boldsymbol{b},i}$ are nonincreasing (Assumption 3.6). In particular, the effective dimension $d(\varepsilon)$ in (4.4) may be infinite for a given ε.

Theorem 4.6 (Impossibility of algebraic decay of $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$)**.** *For any $p \in \mathbb{N}$, there exists $\bar{s} \in \mathbb{N}$, depending on $\lambda_{\boldsymbol{b},1}, \ldots, \lambda_{\boldsymbol{b},p}$, and p, such that*

$$u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})} \geq C s^{-\frac{1}{2p}}, \quad \forall s \geq \bar{s},$$

with constant $C = C(\lambda_{\boldsymbol{b},1}, \ldots, \lambda_{\boldsymbol{b},p}, p) := \frac{1}{2} \left(\prod_{i=1}^{p} \frac{\lambda_{\boldsymbol{b},i}}{i} \right)^{\frac{1}{2p}}$.

Proof Fix $\varepsilon > 0$, whose exact value will be chosen later, and define for $n \in \mathbb{N}$ the set

$$S(\varepsilon, n) := \left\{ \boldsymbol{\gamma} \in \mathbb{N}_0^{\mathbb{N}} : \sum_{i=1}^{\infty} \frac{\gamma_i}{\lambda_{\boldsymbol{b},i}} \leq \frac{1}{\varepsilon^2},\ \operatorname{supp}(\boldsymbol{\gamma}) \subset [n] \right\}.$$

We make a couple of simple but important observations. First note that $S(\varepsilon) = S(\varepsilon, d(\varepsilon))$, where $S(\varepsilon)$ and $d(\varepsilon)$ are defined in (4.3) and (4.4), respectively. Second, we have $S(\varepsilon, n') \subset S(\varepsilon, n)$ for every $1 \leq n' \leq n$. Third, $S(\varepsilon, n)$ is isomorphic to the anisotropic TD index set $\Lambda^{\mathrm{TD}}_{\boldsymbol{a},n}$ with weight $\boldsymbol{a} = (a_1, \ldots, a_n)$, $a_i := \varepsilon^2/\lambda_{\boldsymbol{b},i}$, under the isomorphism

$$\{\boldsymbol{\gamma} \in \mathbb{N}_0^{\mathbb{N}} : \operatorname{supp}(\boldsymbol{\gamma}) \subset [n]\} \to \mathbb{N}_0^n, \quad \boldsymbol{\gamma} \mapsto (\gamma_1, \ldots, \gamma_n).$$

We combine the preceding observations with the lower size bound (4.2) to conclude

$$|S(\varepsilon)| = |S(\varepsilon, d(\varepsilon))| \geq \left|S(\varepsilon, d')\right| \geq \prod_{i=1}^{d'} \frac{\lambda_{\boldsymbol{b},i}}{\varepsilon^2 i}, \quad \forall 1 \leq d' \leq d(\varepsilon). \tag{4.6}$$

Analogous to the definition of $d(\varepsilon)$ in (4.4), we set

$$\widetilde{d}(\varepsilon) := \min\left\{ l \in \mathbb{N} : \frac{\lambda_{\boldsymbol{b},l+1}}{l+1} < \varepsilon^2 \right\} \in \mathbb{N}.$$

Note that $\widetilde{d}(\varepsilon)$ is well-defined because $\lambda_{\boldsymbol{b},i}$ is bounded from above by $\lambda_{\boldsymbol{b},1}$ for every $i \in \mathbb{N}$. Moreover, we have $\widetilde{d}(\varepsilon) \leq d(\varepsilon)$ as well as $\widetilde{d}(\varepsilon) \to \infty$ as $\varepsilon \to 0$.

Next, let $p \in \mathbb{N}$ be arbitrary. We fix some $0 < \bar{\varepsilon} = \bar{\varepsilon}(\lambda_{\boldsymbol{b},p}, p) \leq \min\{\sqrt{\lambda_{\boldsymbol{b},p}/p}, 1\}$. By definition of $\widetilde{d}(\varepsilon)$, we have $\widetilde{d}(\varepsilon) \geq p$ for every $0 < \varepsilon \leq \bar{\varepsilon}$. Since $\lambda_{\boldsymbol{b},i}/(\varepsilon^2 i) \geq 1$ for every $1 \leq i \leq \widetilde{d}(\varepsilon)$, it follows from (4.6) that

$$|S(\varepsilon)| \geq \prod_{i=1}^{\widetilde{d}(\varepsilon)} \frac{\lambda_{\boldsymbol{b},i}}{\varepsilon^2 i} \geq \prod_{i=1}^{p} \frac{\lambda_{\boldsymbol{b},i}}{\varepsilon^2 i} = \widetilde{C} \varepsilon^{-2p}, \quad \forall 0 < \varepsilon \leq \bar{\varepsilon},$$

with constant $\widetilde{C} = \widetilde{C}(\lambda_{\boldsymbol{b},1}, \ldots, \lambda_{\boldsymbol{b},p}, p) := \prod_{i=1}^{p} \frac{\lambda_{\boldsymbol{b},i}}{i}$. We choose $\bar{s} = \bar{s}(\lambda_{\boldsymbol{b},1}, \ldots, \lambda_{\boldsymbol{b},p}, p) \in \mathbb{N}$ sufficiently large such that $s + 1 \geq \lceil \widetilde{C} \bar{\varepsilon}^{-2p} \rceil$ for every $s \geq \bar{s}$. We then fix some arbitrary $s \geq \bar{s}$ and pick $0 < \widetilde{\varepsilon} \leq \bar{\varepsilon}$ such that $\widetilde{C} \widetilde{\varepsilon}^{-2p} = s + 1$. Solving for $\widetilde{\varepsilon}^2$ yields

$$\widetilde{\varepsilon}^2 = \widetilde{\varepsilon}^2(\lambda_{\boldsymbol{b},1}, \ldots, \lambda_{\boldsymbol{b},p}, p, s) = \widetilde{C}^{1/p}(s+1)^{-1/p}.$$

By our choice of $\widetilde{\varepsilon}$, we have $|S(\widetilde{\varepsilon})| \geq s + 1$. Since $S(\widetilde{\varepsilon}) = \pi([|S(\widetilde{\varepsilon})|])$, we conclude $\boldsymbol{\pi(s+1)} \in S(\widetilde{\varepsilon})$ and therefore

$$u^2_{\boldsymbol{\pi(s+1)}} \geq \left(\frac{1}{\widetilde{\varepsilon}^2} + 1 \right)^{-1} \geq \frac{1}{2} \widetilde{\varepsilon}^2 = \frac{1}{2} \widetilde{C}^{1/p} (s+1)^{-1/p} \geq \frac{1}{4} \widetilde{C}^{1/p} s^{-1/p}, \quad \forall s \geq \bar{s},$$

where the second inequality holds because $\widetilde{\varepsilon} \leq \bar{\varepsilon} \leq 1$. This completes the proof. ☐

4.3. *Upper bounds*

We have now seen that $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$ cannot decay algebraically as $s\to\infty$, regardless of the decay of the $\boldsymbol{b}$-weighted PCA eigenvalues $\lambda_{\boldsymbol{b},i}$. We now study the decay of $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$ for three different decays of these eigenvalues: Algebraic and exponential decays, which are typically considered in the context of (functional) PCA (see, e.g., Reiß and Wahl (2020); Milbradt and Wahl (2020) and references therein), as well as double exponential decays. In all three cases we have $\lim_{i\to\infty}\lambda_{\boldsymbol{b},i}=0$ so that the effective dimension $d(\varepsilon)$ is finite for every $\varepsilon>0$, see Remark 4.5. In principle, the proof of the following result can be adapted to any other spectral decay rate, as long as this vanishing limit condition is satisfied so that a suitable choice of ε is possible.

Theorem 4.7 (Typical decays of $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$). *Let $\alpha,\beta>0$.*

(a) Algebraic spectral decay: *Let $\lambda_{\boldsymbol{b},i}=i^{-\alpha}$ for every $i\in\mathbb{N}$. Then, for every $\delta,\eta>0$, there exists $\bar{s}=\bar{s}(\alpha,\delta,\eta)\in\mathbb{N}$ such that for every $s\ge\bar{s}$,*

$$u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}\le\eta\log(s)^{-\frac{1}{2(1/\alpha+\delta)}}. \tag{4.7}$$

(b) Exponential spectral decay: *Let $\lambda_{\boldsymbol{b},i}=e^{-\alpha i^{\beta}}$ for every $i\in\mathbb{N}$. Then, for every $\delta>0$, there exists $\bar{s}=\bar{s}(\alpha,\beta,\delta)\in\mathbb{N}$ such that for every $s\ge\bar{s}$,*

$$u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}\le e^{-\frac{1}{2}\alpha^{\frac{1}{\beta+1}}\left(\frac{\beta+1}{\beta+\delta}\right)^{\frac{1}{1+1/\beta}}\log(s)^{\frac{1}{1+1/\beta}}}. \tag{4.8}$$

(c) Double exponential spectral decay: *Let $\lambda_{\boldsymbol{b},i}=e^{-e^{\alpha i}}$ for every $i\in\mathbb{N}$. Then, for every $\delta,\eta>0$, there exists $\bar{s}=\bar{s}(\alpha,\delta,\eta)\in\mathbb{N}$ such that for every $s\ge\bar{s}$,*

$$u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}\le e^{-\eta\log(s)^{\frac{1}{1+\delta}}}. \tag{4.9}$$

Proof All rates can be derived by suitably choosing the parameter ε in the set $S(\varepsilon)$, defined in (4.3). By the relation (4.5), we may use the the upper size bound (4.2) to compute

$$|S(\varepsilon)|\le\prod_{i=1}^{d(\varepsilon)}\left(\frac{\lambda_{\boldsymbol{b},i}}{i\varepsilon^2}+1\right)\le\prod_{i=1}^{d(\varepsilon)}\frac{\lambda_{\boldsymbol{b},i}}{\varepsilon^2}\left(\frac{1}{i}+1\right)\le(2\varepsilon^{-2})^{d(\varepsilon)}\prod_{i=1}^{d(\varepsilon)}\lambda_{\boldsymbol{b},i}. \tag{4.10}$$

In the second step we used the fact that $\lambda_{\boldsymbol{b},i}\ge\varepsilon^2$ for $i\in[d(\varepsilon)]$ by definition of $d(\varepsilon)$ (see (4.4)). For brevity, we write in the following $d=d(\varepsilon)$.

Case (a). Using (4.10) as well as the Stirling type estimate $d^d\le e^d d!$, we obtain

$$|S(\varepsilon)|\le(2\varepsilon^{-2})^d\prod_{i=1}^{d}i^{-\alpha}=(2\varepsilon^{-2})^d(d!)^{-\alpha}\le(2\varepsilon^{-2})^d e^{\alpha d}d^{-\alpha d}. \tag{4.11}$$

Let $0<\varepsilon\leq 2^{-\alpha/2}$. By definition, we have $d\leq\varepsilon^{-2/\alpha}<d+1\leq 2d$. We take the logarithm on both sides in (4.11) and compute

$$\begin{aligned}\log(|S(\varepsilon)|)&\leq d\log(2\varepsilon^{-2})+\alpha d-\alpha d\log(d)\\&\leq\varepsilon^{-2/\alpha}\log(2\varepsilon^{-2})+\alpha\varepsilon^{-2/\alpha}-\frac{1}{2}\alpha\varepsilon^{-2/\alpha}\log\left(\frac{1}{2}\varepsilon^{-2/\alpha}\right)\\&=\frac{1}{2}\varepsilon^{-2/\alpha}\log(2\varepsilon^{-2})+\left(\alpha+\frac{1}{2}\log(2^{\alpha+1})\right)\varepsilon^{-2/\alpha}.\end{aligned}$$

Next, let $\delta,\eta>0$ be arbitrary. There exists $0<\bar{\varepsilon}=\bar{\varepsilon}(\alpha,\delta,\eta)\leq 2^{-\alpha/2}$ such that for every $0<\varepsilon\leq\bar{\varepsilon}$, we have $\log(2\varepsilon^{-2})\leq\eta^{2(1/\alpha+\delta)}\varepsilon^{-2\delta}$ as well as $\alpha+\frac{1}{2}\log(2^{\alpha+1})\leq\frac{\eta^{2(1/\alpha+\delta)}}{2}\varepsilon^{-2\delta}$. It follows that

$$\log(|S(\varepsilon)|)\leq\eta^{2(1/\alpha+\delta)}\varepsilon^{-2/\alpha-2\delta},\quad\forall\varepsilon\leq\bar{\varepsilon}. \tag{4.12}$$

We set the right hand-side equal to $\log(s)$ and solve for ε^2, which gives

$$\varepsilon^2=\varepsilon(s)^2=\eta^2\log(s)^{-\frac{1}{1/\alpha+\delta}}. \tag{4.13}$$

We can now choose $\bar{s}=\bar{s}(\alpha,\delta,\eta)\in\mathbb{N}$ sufficiently large such that the right-hand side in (4.13) is smaller than $\bar{\varepsilon}$ for every $s\geq\bar{s}$. Then, (4.12) holds with $\varepsilon=\varepsilon(s)$ and we conclude $|S(\varepsilon(s))|\leq s$ for every $s\geq\bar{s}$. Since $S(\varepsilon)=\pi([|S(\varepsilon)|])$, it follows that $u_{\boldsymbol{\pi}(\boldsymbol{s}+\boldsymbol{1})}\notin S(\varepsilon)$ and consequently,

$$u^2_{\boldsymbol{\pi}(\boldsymbol{s}+\boldsymbol{1})}\leq(\varepsilon(s)^{-2}+1)^{-1}\leq\varepsilon(s)^2\leq\eta^2\log(s)^{-\frac{1}{1/\alpha+\delta}},\quad\forall s\geq\bar{s}.$$

Case (b). By (4.10), we find

$$|S(\varepsilon)|\leq(2\varepsilon^{-2})^d\prod_{i=1}^{d}e^{-\alpha i^\beta}=(2\varepsilon^{-2})^d e^{-\alpha\sum_{i=1}^d i^\beta}. \tag{4.14}$$

Let $0<\varepsilon\leq 1$. By definition, we have $d\leq\alpha^{-1/\beta}\log(\varepsilon^{-2})^{1/\beta}<d+1$. We take the logarithm on both sides of (4.14) and compute

$$\begin{aligned}\log(|S(\varepsilon)|)&\leq d\log(2\varepsilon^{-2})-\alpha\sum_{i=1}^{d}i^\beta\leq d\log(2\varepsilon^{-2})-\alpha\int_0^d t^\beta dt\\&\leq\alpha^{-1/\beta}\log(\varepsilon^{-2})^{1+1/\beta}+\alpha^{-1/\beta}\log(2)\log(\varepsilon^{-2})^{1/\beta}\\&\qquad-\frac{\alpha}{\beta+1}(\alpha^{-1/\beta}\log(\varepsilon^{-2})^{1/\beta}-1)^{\beta+1}.\end{aligned}$$

Next, let $\widetilde{\delta}>0$ and $0<\eta<1$ be arbitrary. There exists $\bar{\varepsilon}=\bar{\varepsilon}(\alpha,\beta,\widetilde{\delta},\eta)>0$ such that $\log(2)\leq\widetilde{\delta}\log(\varepsilon^{-2})$ and $\alpha^{-1/\beta}\log(\varepsilon^{-2})^{1/\beta}-1\geq\eta^{\frac{1}{\beta+1}}\alpha^{-1/\beta}\log(\varepsilon^{-2})^{1/\beta}$ for every $0<\varepsilon\leq\bar{\varepsilon}$. Hence

$$\log(|S(\varepsilon)|)\leq\alpha^{-1/\beta}\left(1+\widetilde{\delta}-\frac{\eta}{\beta+1}\right)\log(\varepsilon^{-2})^{1+1/\beta},\quad\forall\varepsilon\leq\bar{\varepsilon}.$$

We now set $\delta=1+\widetilde{\delta}(\beta+1)-\eta$ and conclude similarly as in case (a).

Case (c). By (4.10), we find that

$$|S(\varepsilon)| \leq (2\varepsilon^{-2})^d \prod_{i=1}^{d} e^{-e^{\alpha i}} = (2\varepsilon^{-2})^d e^{-\sum_{i=1}^{d} e^{\alpha i}}. \tag{4.15}$$

Let $0 < \varepsilon \leq \sqrt{1/e}$. Then, by definition, we have $d \leq \log(\log(\varepsilon^{-2}))/\alpha < d+1$. We take the logarithm on both sides in (4.15) and compute

$$\log(|S(\varepsilon)|) \leq d\log(2\varepsilon^{-2}) - \sum_{i=1}^{d} e^{\alpha i} \leq \frac{1}{\alpha}\log(\log(\varepsilon^{-2}))\log(2\varepsilon^{-2}) - \frac{\log(\varepsilon^{-2})-1}{e^{\alpha}-1} + 1. \tag{4.16}$$

Next, let $\delta, \eta > 0$ be arbitrary. There exists $0 < \bar{\varepsilon} = \bar{\varepsilon}(\alpha,\delta,\eta) \leq \sqrt{1/e}$ such that for every $0 < \varepsilon \leq \bar{\varepsilon}$, the rightmost hand-side in (4.16) is dominated by $(2\eta)^{-(1+\delta)}\log(\varepsilon^{-2})^{1+\delta}$. We can now conclude similarly as in case (a). $\square$

Similar computations (which we omit for brevity) based on the lower size bound (4.2) show that the bounds in Theorem 4.7 are asymptotically sharp in the limit $s \to \infty$—at least in the exponential and double exponential case. Indeed, in case (b), for every $0 < \delta < \beta$, there exists $\bar{s} = \bar{s}(\alpha,\beta,\delta) \in \mathbb{N}$ such that

$$u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})} \gtrsim e^{-\frac{1}{2}\alpha^{\frac{1}{\beta+1}}\left(\frac{\beta+1}{\beta-\delta}\right)^{\frac{1}{1+1/\beta}}\log(s+1)^{\frac{1}{1+1/\beta}}}, \quad \forall s \geq \bar{s}. \tag{4.17}$$

In case (c), for every $\eta > 0$, there exists $\bar{s} = \bar{s}(\alpha,\eta) \in \mathbb{N}$ such that

$$u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})} \gtrsim (s+1)^{-\eta}, \quad \forall s \geq \bar{s}.$$

In case (a) with algebraic spectral decay, the lower size bound (4.2) is too weak and leads to inconclusive results which do not asymptotically match the upper bound (4.7).

4.4. *Discussion*

Theorem 4.6 shows that $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$ decays subalgebraically as $s \to \infty$. In particular, as $p \in \mathbb{N}$ can be chosen arbitrarily large, we conclude that the decay is slower than *any* algebraic decay rate for all sufficiently large s. In combination with Theorem 4.1, we deduce the following **curse of parametric complexity**: *No s-term Wiener-Hermite PC expansion can converge with an algebraic rate uniformly for all operators in the Sobolev unit (Lipschitz) ball as $s \to \infty$. This holds regardless of the decay rate of the PCA eigenvalues.*

We have now seen that the approximation of Lipschitz operators by Wiener-Hermite PC expansions cannot be done efficiently with algebraic convergence rates. On the other hand, Theorem 4.7 shows the connection between the decay of the eigenvalues $\lambda_{\boldsymbol{b},i}$ in i and the decay of $u_{\boldsymbol{\pi}(\boldsymbol{s}+\mathbf{1})}$ in s. It illustrates the obvious qualitative fact that a faster eigenvalue decay implies a faster decay of the polynomial s-term error. But it also provides quantitative decay rates and shows that the curse of parametric complexity can be overcome at least asymptotically in the sense that decay rates arbitrarily close to any algebraic rate can be attained, provided the $\lambda_{\boldsymbol{b},i}$ decay sufficiently fast. Since the lower bound (4.17) asymptotically matches the upper bound (4.8), we see, however, that to this end, the eigenvalue decay has to be faster than exponential. Equation (4.9) shows that a double exponential decay is sufficient.

In the context of the above stated parametric curse of complexity, the parameters, that is, the polynomial coefficients in the truncated Wiener-Hermite PC expansion, are elements of $\mathcal{Y}$. In Lanthaler (2023), a related curse of parametric complexity for learning bounded Lipschitz operators with PCA-Net was proved, which for distinction we call *curse of parametric PCA-Net complexity*. It relates the learnability of Lipschitz operators by PCA-Nets to the size, i.e., the number of neural network parameters, of the latter. For succinctness, we refer to Lanthaler (2023) for more details and definitions. The following result follows from Theorem 9 therein.

Theorem 4.8 (Curse of parametric PCA-Net complexity)**.** *For any $\alpha > 0$, there exists a bounded Lipschitz operator $F \in C^{0,1}(\mathcal{X}, \mathcal{Y})$ and a constant $c_\alpha > 0$ such that*

$$\|F - \Psi\|^2_{L^2_\mu(\mathcal{X};\mathcal{Y})} \geq c_\alpha \, (\mathrm{size}(\psi))^{-\alpha}$$

for every PCA-Net $\Psi = \mathcal{D}_\mathcal{Y} \circ \psi \circ \mathcal{E}_\mathcal{X}$. Here, $\psi : \mathbb{R}^{d_\mathcal{X}} \to \mathbb{R}^{d_\mathcal{Y}}$ is a (ReLU) neural network and $\mathcal{E}_\mathcal{X} : \mathcal{X} \to \mathbb{R}^{d_\mathcal{X}}$ and $\mathcal{D}_\mathcal{Y} : \mathbb{R}^{d_\mathcal{Y}} \to \mathcal{Y}$ are encoder and decoder mappings, respectively, which are defined in terms of the empirical PCA eigenvalues based on finitely many fixed sample points.

Remark 4.9. *Inspection of the proof of Lanthaler (2023, Thm. 9) shows that the empirical PCA eigenvalues used in the definition of $\mathcal{E}_\mathcal{X}, \mathcal{D}_\mathcal{Y}$ can be replaced by the exact PCA eigenvalues, which puts Theorem 4.8 in the setting of the present paper.*

We now argue that the curse of parametric complexity described by Theorems 4.1 and 4.6 is consistent with the curse of parametric PCA-Net complexity described by Theorem 4.8. To this end, suppose that Hermite polynomial s-term approximations of Lipschitz operators in the Sobolev unit Lipschitz ball were possible with some algebraic rate of order $\beta > 0$.

Assumption 4.10 (Algebraic s-term convergence for Lipschitz operators)**.** *There exists $\beta > 0$ such that for every $F \in B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, there are constants $c(F) > 0$ and $\bar{s}(F) \in \mathbb{N}$ such that*

$$\left\|F - F_{\pi([s])}\right\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq c(F) s^{-\beta}, \quad \forall s \geq \bar{s}(F). \tag{4.18}$$

Proposition 4.11. *Grant Assumption 4.10. Then there exists $\alpha > 0$ with the following property: For all $F \in C^{0,1}(\mathcal{X}, \mathcal{Y})$, there exist constants $C = C(F) > 0$ and $\bar{\epsilon} = \bar{\epsilon}(F) > 0$ such that for all $0 < \epsilon \leq \bar{\epsilon}$, there is a PCA-Net $\Psi = \widetilde{\mathcal{D}}_\mathcal{Y} \circ \psi \circ \widetilde{\mathcal{E}}_\mathcal{X}$ such that*

$$\|F - \Psi\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq \epsilon \quad \text{and} \quad \mathrm{size}(\psi) \leq C \epsilon^{-1/\alpha}.$$

With the notion introduced in Lanthaler (2023), Proposition 4.11 asserts that α is an algebraic convergence rate for the class of $C^{0,1}(\mathcal{X}, \mathcal{Y})$-operators. It is easy to see that this leads to a contradiction to Theorem 4.8, hence implying Assumption 4.10 to be false. The proof of Proposition 4.11 is based on Schwab and Zech (2023, Thm. 3.9), which provides expression rate bounds for the approximation of multivariate Hermite polynomials by ReLU neural networks. As a preliminary step, we make the following straightforward observation: Let S be a downward closed subset of $\mathbb{N}_0^{\mathbb{N}}$, that is, $\boldsymbol{\gamma} \in S$ implies $\boldsymbol{\gamma'} \in S$ for every $\boldsymbol{\gamma'} \leq \boldsymbol{\gamma}$. Then

$$\max_{\boldsymbol{\gamma} \in S} \|\boldsymbol{\gamma}\|_{\ell^1(\mathbb{N})} \leq |S| - 1 \quad \text{and} \quad \max_{\boldsymbol{\gamma} \in S} |\mathrm{supp}(\boldsymbol{\gamma})| \leq |S| - 1. \tag{4.19}$$

Proof of Proposition 4.11 Suppose that Assumption 4.10 is true. Let $F \in C^{0,1}(\mathcal{X}, \mathcal{Y})$ with $F \neq 0$ (otherwise there is nothing to show). We define the rescaled operator $\widetilde{F} := r^{-1} \|F\|^{-1}_{C^{0,1}(\mathcal{X},\mathcal{Y})} F$, where $r := \sqrt{\dim(\mathcal{Y})}$ if $\mathcal{Y}$ is finite-dimensional and $r := \max\{1, \|\boldsymbol{b}\|_{\ell^2(\mathbb{N})}\}$ if $\mathcal{Y}$ is infinite-dimensional. This implies $\widetilde{F} \in B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X}; \mathcal{Y})$ by Theorem 3.10. Next, let $\bar{s} = \bar{s}(F)$ be the constant in Assumption 4.10 and set $\bar{\epsilon} := \min\{\bar{s}^{-\beta/2}, e^{-\beta/(\beta+1)}\}$ so that $\bar{\epsilon}^{(\beta+1)/\beta} = \min\{\bar{s}^{-(\beta+1)/2}, e^{-1}\}$. Let $0 < \epsilon \leq \bar{\epsilon}$ be arbitrary and choose $s \geq \bar{s}$ with $s^{-\beta/2} \sim \epsilon$.

We proceed with several observations: First, recall from (2.2) that the truncated Wiener-Hermite PC expansion $\widetilde{F}_{\pi([s])}$ is defined via the Hermite polynomials $H_{\boldsymbol{\pi(1)},d_1}, \ldots, H_{\boldsymbol{\pi(s)},d_s}$ with $d_i := \max\{j : j \in \mathrm{supp}(\boldsymbol{\pi(i)})\}$. By Remark 2.2, we can interpret each $H_{\boldsymbol{\pi(i)},d_i}$ as a function $H_{\boldsymbol{\pi(i)},\infty}$ on $\mathbb{R}^{\mathbb{N}}$. Second, observe that the set $\pi([s])$ is a downward closed subset of $\mathbb{N}_0^{\mathbb{N}}$ of size s. Hence, by (4.19), we have

$$\max_{\boldsymbol{\gamma} \in \pi([s])} \|\boldsymbol{\gamma}\|_{\ell^1(\mathbb{N})} \leq s - 1 \quad \text{and} \quad \max_{\boldsymbol{\gamma} \in \pi([s])} |\mathrm{supp}(\boldsymbol{\gamma})| \leq s - 1, \quad \forall s \in \mathbb{N}.$$

With these facts in hand, we can now directly apply Schwab and Zech (2023, Thm. 3.9) to conclude that there exists a ReLU neural network $\psi = (\psi_1, \ldots, \psi_s) : \mathbb{R}^d \to \mathbb{R}^s$ with $d := \max\{d_1, \ldots, d_s\}$ and with each ψ_i depending solely on the variables $(x_i)_{i \in [d]}$ such that

$$\max_{i \in [s]} \left\| H_{\boldsymbol{\pi(i)},\infty} - \psi_i \right\|_{L^2_{\mu_\infty}(\mathbb{R}^{\mathbb{N}})} \leq \epsilon^{(\beta+1)/\beta}. \tag{4.20}$$

Here, we interpret the ψ_i as functionals on $\mathbb{R}^{\mathbb{N}}$ by ignoring all variables x_i with $i > d$. Moreover,

$$\mathrm{size}(\psi) \lesssim s^6 \log(s) \log(\epsilon^{-(\beta+1)/\beta}) \leq \left(1 + \frac{1}{\beta}\right) \epsilon^{-14/\beta - 1}. \tag{4.21}$$

We now define the encoder and decoder mappings

$$\widetilde{\mathcal{E}}_{\mathcal{X}} : \mathcal{X} \to \mathbb{R}^d, \quad \widetilde{\mathcal{E}}_{\mathcal{X}}(X) := \left(\frac{\langle X, \phi_1 \rangle_{\mathcal{X}}}{\sqrt{\lambda_1}}, \ldots, \frac{\langle X, \phi_d \rangle_{\mathcal{X}}}{\sqrt{\lambda_d}} \right),$$

$$\widetilde{\mathcal{D}}_{\mathcal{Y}} : \mathbb{R}^s \to \mathcal{Y}, \quad \widetilde{\mathcal{D}}_{\mathcal{Y}}(\boldsymbol{x}) := \sum_{i=1}^{s} \left(\int_{\mathcal{X}} \widetilde{F} H_{\boldsymbol{\pi(i)},\boldsymbol{\lambda}} d\mu \right) x_i.$$

The corresponding (rescaled) PCA-Net $\Psi := r\|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})} \left(\widetilde{\mathcal{D}}_{\mathcal{Y}} \circ \psi \circ \widetilde{\mathcal{E}}_{\mathcal{X}} \right)$ satisfies

$$\|F - \Psi\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq r\|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})} \Bigg(\underbrace{\left\| \widetilde{F} - \widetilde{F}_{\pi([s])} \right\|_{L^2_\mu(\mathcal{X};\mathcal{Y})}}_{=:T_1(F)} + \underbrace{\left\| \widetilde{F}_{\pi([s])} - \widetilde{\mathcal{D}}_{\mathcal{Y}} \circ \psi \circ \widetilde{\mathcal{E}}_{\mathcal{X}} \right\|_{L^2_\mu(\mathcal{X};\mathcal{Y})}}_{=:T_2(F)} \Bigg).$$

By Assumption 4.10, we can bound $T_1(F)$ by $T_1(F) \leq c(F)s^{-\beta} \sim c(F)\epsilon^2 \leq c(F)\epsilon$, where $c(F)$ is the constant in (4.18). For $T_2(F)$, we find by an application of (4.20), the Cauchy-Schwarz inequality, and Parseval's identity that

$$T_2(F) \le \sum_{i=1}^{s} \left\| \int_{\mathcal{X}} \widetilde{F} H_{\boldsymbol{\pi(i)},\boldsymbol{\lambda}} d\mu \right\|_{\mathcal{Y}} \|H_{\boldsymbol{\pi(i)},\infty} - \psi_i\|_{L^2_{\mu_\infty}(\mathbb{R}^{\mathbb{N}})}$$

$$\le \epsilon^{(\beta+1)/\beta} s^{1/2} \left(\sum_{i=1}^{s} \left\| \int_{\mathcal{X}} \widetilde{F} H_{\boldsymbol{\pi(i)},\boldsymbol{\lambda}} d\mu \right\|_{\mathcal{Y}}^2 \right)^{1/2} \lesssim \epsilon \|\widetilde{F}\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \le \epsilon.$$

Altogether, we conclude that $\|F - \Psi\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \le C' r \|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})}(C(F)+1)\epsilon$ for some global constant $C' > 0$. Upon rescaling ϵ and $\bar{\epsilon}$ by the factor $C' r \|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})}(c(F)+1)$, the claim follows in light of (4.21) with $\alpha = \beta/(14+\beta)$. □

5. Optimal sampling and (adaptive) m-widths

Up to this point, we have studied the best approximation of $W^{1,2}_{\mu,\boldsymbol{b}}$- and Lipschitz operators by Hermite polynomials. This is a certain type of what in the field of information-based complexity is referred to as *linear information* (Novak and Woźniakowski, 2008, Chpt. 4.1.1). In this section, we consider more general sampling and reconstruction schemes based on *adaptive information*, that is, (nonlinear) reconstruction from m adaptively chosen samples. We define adaptive sampling operators and the adaptive m-width and characterize the latter in terms of the weights $u_{\boldsymbol{\gamma}}$. We follow in parts ideas from Adcock et al. (2024c), which studied holomorphic operators. It is known that adaptive methods can be better than nonadaptive methods only by a factor of at most 2 and there are examples where adaptive methods perform slightly better than nonadaptive ones. We refer to Theorem 2 in Novak (1996) as well as to Krieg et al. (2024) and references therein. For this reason, we consider adaptive sampling operators instead of nonadaptive ones. However, we will prove that for $W^{1,2}_{\mu,\boldsymbol{b}}$- and Lipschitz operators, linear approximation based on nonadaptive information is, in fact, optimal, see Theorem 5.4.

5.1. *Adaptive sampling operators*

We first introduce in scalar- and Hilbert-valued adaptive sampling operators.

Definition 5.1 (Adaptive sampling operator; scalar-valued case)**.** Let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be a normed vector space and $m \in \mathbb{N}$. A (scalar-valued) *adaptive sampling operator* is a mapping of the form

$$\mathcal{L} : \mathcal{V} \to \mathbb{R}^m, \quad \mathcal{L}(F) = \begin{pmatrix} \mathcal{L}_1(F) \\ \mathcal{L}_2(F; \mathcal{L}_1(F)) \\ \vdots \\ \mathcal{L}_m(F; \mathcal{L}_1(F), \dots, \mathcal{L}_{m-1}(F)) \end{pmatrix},$$

where $\mathcal{L}_1 : \mathcal{V} \to \mathbb{R}$ is a bounded linear functional and $\mathcal{L}_i : \mathcal{V} \times \mathbb{R}^{i-1} \to \mathbb{R}$ is bounded and linear in its first component for $i = 2, \dots, m$.

Trivially, any bounded linear mapping $\mathcal{L} : \mathcal{V} \to \mathbb{R}^m$ is an adaptive sampling operator (which, in fact, generates nonadaptive information). Different choices for $\mathcal{V}$ lead to important special cases. If $\mathcal{V} = L^2_\mu(\mathcal{X})$ and $\boldsymbol{\gamma}^{(1)},\dots,\boldsymbol{\gamma}^{(m)} \in \Gamma$, we can define a sampling operator, generating (nonadaptive) *linear information*, by

$$\mathcal{L}(F) := \left(\int_{\mathcal{X}} F H_{\boldsymbol{\gamma}^{(i)},\boldsymbol{\lambda}} d\mu\right)_{i\in[m]} \in \mathbb{R}^m, \quad \forall F \in L^2_\mu(\mathcal{X}). \tag{5.1}$$

If $\mathcal{V} = C(\mathcal{X})$ and $X_1,\dots,X_m \in \mathcal{V}$, we can define a pointwise sampling operator, generating (nonadaptive) *standard information* (Novak and Woźniakowski, 2008, Chpt. 4.1.1), by

$$\mathcal{L}(F) := (F(X_i))_{i\in[m]} \in \mathbb{R}^m, \quad \forall F \in C(\mathcal{X}). \tag{5.2}$$

In both cases, the $\boldsymbol{\gamma}^{(i)}$ and the X_i can, in principle, also be chosen adaptively based on previous measurements $\int_{\mathcal{X}} F H_{\boldsymbol{\gamma}^{(j)}} d\mu$ and $F(X_j)$, respectively, for $j \in [i-1]$.

Next, we consider the Hilbert-valued case. For any $Y \in \mathcal{Y}$, $\boldsymbol{v} = (v_i)_{i\in[m]} \in \mathbb{R}^m$, and $F \in L^2_\mu(\mathcal{X})$, we write $Y\boldsymbol{v}$ for the vector $(Yv_i)_{i\in[m]} \in \mathcal{Y}^m$ and YF for the mapping $X \mapsto YF(X)$.

Definition 5.2 (Adaptive sampling operator; Hilbert-valued case)**.** Let $\mathcal{V} \subset L^2_\mu(\mathcal{X};\mathcal{Y})$ be a vector subspace with norm $\|\cdot\|_{\mathcal{V}}$ and consider a mapping

$$\mathcal{L}: \mathcal{V} \to \mathcal{Y}^m, \;\; \mathcal{L}(F) = \begin{pmatrix} \mathcal{L}_1(F) \\ \mathcal{L}_2(F;\mathcal{L}_1(F)) \\ \vdots \\ \mathcal{L}_m(F;\mathcal{L}_1(F),\dots,\mathcal{L}_{m-1}(F)) \end{pmatrix},$$

where $\mathcal{L}_1 : \mathcal{V} \to \mathcal{Y}$ is a bounded linear operator and $\mathcal{L}_i : \mathcal{V} \times \mathcal{Y}^{i-1} \to \mathcal{Y}$ is bounded and linear in its first component for $i = 2,\dots,m$. Then $\mathcal{L}$ is a *Hilbert-valued adaptive sampling operator* if the following condition holds: There exist $Y, \widetilde{Y} \in \mathcal{Y}\setminus\{0\}$, a normed vector space $\widetilde{\mathcal{V}} \subset L^2_\mu(\mathcal{X})$, and a scalar-valued adaptive sampling operator $\widetilde{\mathcal{L}} : \widetilde{\mathcal{V}} \to \mathbb{R}^m$ such that, if $YF \in \mathcal{V}$ for some $F \in L^2_\mu(\mathcal{X})$, then $F \in \widetilde{\mathcal{V}}$ and $\mathcal{L}(YF) = \widetilde{Y}\widetilde{\mathcal{L}}(F)$.

This definition involves a technical assumption which links the Hilbert-valued case to the scalar-valued case and which we will use to establish a lower bound for the adaptive m-width. However, this condition is not too strong. It holds, for example, in the case of adaptive pointwise sampling. Here, we choose $\mathcal{V} = C(\mathcal{X},\mathcal{Y})$ (seen as a subspace of $L^2_\mu(\mathcal{X};\mathcal{Y})$ and equipped with the L^∞-norm) and define

$$\mathcal{L}(F) := (F(X_i))_{i\in[m]} \in \mathcal{Y}^m, \quad \forall F \in \mathcal{V},$$

where the ith sample point X_i is potentially chosen based on the previous measurements $F(X_1),\dots,F(X_{i-1})$. We then have

$$\mathcal{L}(YF) = Y\widetilde{\mathcal{L}}(F), \quad \forall F \in \widetilde{\mathcal{V}} := C(\mathcal{X}), \forall Y \in \mathcal{Y},$$

where $\widetilde{\mathcal{L}} : \widetilde{\mathcal{V}} \to \mathbb{R}^m$ is the adaptive pointwise sampling operator in (5.2). As another example, we will see in the proof of the upper bound of the adaptive m-width that the Hilbert-valued version of linear sampling, as defined in (5.1), is a Hilbert-valued sampling operator in the sense of Definition 5.2 (under a mild condition on $\mathcal{V}$), see Section 5.3.3.

5.2. *Adaptive m-widths and main result*

We now formally define the adaptive m-width and state our main result.

Definition 5.3 (Adaptive m-width)**.** Let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be a normed vector subspace of $L^2_\mu(\mathcal{X};\mathcal{Y})$ and let $\mathcal{K} \subset \mathcal{V}$ be a subset. The adaptive m-width of $\mathcal{K}$ in $\mathcal{V}$ is given by

$$\begin{aligned} &\Theta_m(\mathcal{K};\mathcal{V}, L^2_\mu(\mathcal{X};\mathcal{Y})) \\ &:= \inf\left\{ \sup_{F\in\mathcal{K}} \|F - \mathcal{T}(\mathcal{L}(F))\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} : \mathcal{L}: \mathcal{V} \to \mathcal{Y}^m \text{ adaptive}, \mathcal{T}: \mathcal{Y}^m \to L^2_\mu(\mathcal{X};\mathcal{Y}) \right\}. \end{aligned} \tag{5.3}$$

The adaptive m-width describes the smallest worst-case error that can be achieved when we reconstruct all operators in a set $\mathcal{K}$ by a reconstruction mapping $\mathcal{T}$ from m samples that have been generated by an adaptive Hilbert-valued sampling operator $\mathcal{L}$. It thus quantifies the largest error that can occur with optimally chosen sampling and reconstruction mappings. Note that in (5.3) we allow for any (possibly nonlinear) reconstruction mappings. The choice of $\mathcal{V}$, however, determines which sampling operators are allowed. If $\mathcal{V} = C(\mathcal{X},\mathcal{Y})$, we can use pointwise sampling, whereas, if $\mathcal{V} = L^2_\mu(\mathcal{X};\mathcal{Y})$, we can not.

We consider two choices for $\mathcal{K}$, namely $\mathcal{K} = B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and $\mathcal{K} = B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, see Definition 3.13. The lower bound for the adaptive m-width pertains to arbitrary $\mathcal{V}$. For the upper bound, we require the additional assumption that $\mathcal{V}$ is continuously embedded in $L^2_\mu(\mathcal{X};\mathcal{Y})$. Our main result in this section is the tight characterization of the adaptive m-width of the Sobolev unit (Lipschitz) ball in terms of the weights $u_{\boldsymbol{\gamma}}$.

Theorem 5.4 (Tight characterization of adapt. m-width)**.** *Let $\mathcal{K} \in \{B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}), B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})\}$. For every $m \in \mathbb{N}$, we have the lower bound*

$$\Theta_m(\mathcal{K};\mathcal{V}, L^2_\mu(\mathcal{X};\mathcal{Y})) \geq u_{\boldsymbol{\pi}(\boldsymbol{m}+\boldsymbol{1})},$$

where $\pi : \mathbb{N} \to \Gamma$ is a nonincreasing rearrangement of $\boldsymbol{u} = (u_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma}$, see (3.3). If, in addition, $\mathcal{V}$ is continuously embedded in $L^2_\mu(\mathcal{X};\mathcal{Y})$, we have for every $m \in \mathbb{N}$ the matching upper bound

$$\begin{aligned} \Theta_m(\mathcal{K};\mathcal{V}, L^2_\mu(\mathcal{X};\mathcal{Y})) &\leq \inf_{S\subset\Gamma, |S|\leq m} \sup_{F\in\mathcal{K}} \|F - F_S\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \\ &\leq \sup_{F\in\mathcal{K}} \left\|F - F_{\{\boldsymbol{\pi}(\boldsymbol{1}),\ldots,\boldsymbol{\pi}(\boldsymbol{m})\}}\right\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq u_{\boldsymbol{\pi}(\boldsymbol{m}+\boldsymbol{1})}. \end{aligned}$$

5.3. *Proof of Theorem 5.4*

The proof of the lower bound is based on the theory of Gelfand and Kolmogorov m-widths. We recall relevant results in Section 5.3.1. Further information can be found in Foucart and Rauhut (2013, Chpt. 10). Proofs of the lower and upper bound are then given in Sections 5.3.2 and 5.3.3, respectively.

5.3.1. Results about widths

Let $\mathcal{K}$ be a subset of a normed vector space $(\mathcal{Z}, \|\cdot\|_{\mathcal{Z}})$ and let $m \in \mathbb{N}$. The *Gelfand m-width* of $\mathcal{K}$ is defined by

$$d^m(\mathcal{K}, \mathcal{Z}) := \inf\left\{ \sup_{Z \in \mathcal{K} \cap L^m} \|Z\|_{\mathcal{Z}} : L^m \text{ subspace of } \mathcal{X} \text{ with } \operatorname{codim}(L^m) \le m \right\}.$$

An equivalent characterization is given by

$$d^m(\mathcal{K}, \mathcal{Z}) = \inf\left\{ \sup_{Z \in \mathcal{K} \cap \ker(A)} \|Z\|_{\mathcal{Z}} : A : \mathcal{Z} \to \mathbb{R}^m \text{ linear} \right\}.$$

We also recall the *adaptive compressive m-width* of $\mathcal{K}$,

$$E^m_{\text{ada}}(\mathcal{K}, \mathcal{Z}) := \inf\left\{ \sup_{Z \in \mathcal{K}} \|Z - \Delta(\Gamma(Z))\|_{\mathcal{Z}} : \Gamma : \mathcal{Z} \to \mathbb{R}^m \text{ adaptive}, \ \Delta : \mathbb{R}^m \to \mathcal{Z} \right\},$$

and the *Kolmogorov m-width* of $\mathcal{K}$,

$$d_m(\mathcal{K}, \mathcal{Z}) := \inf\left\{ \sup_{K \in \mathcal{K}} \inf_{Z \in \mathcal{Z}_m} \|Z - K\|_{\mathcal{Z}} : \mathcal{Z}_m \text{ subspace of } \mathcal{Z} \text{ with } \dim(\mathcal{Z}_m) \le m \right\}.$$

Next, we state some standard results which relate these various notions of m-width.

Theorem 5.5 (Foucart and Rauhut (2013, Thm. 10.4))**.** *If $\mathcal{K}$ is symmetric with respect to the origin, i.e., $-\mathcal{K} = \mathcal{K}$, then $d^m(\mathcal{K}, \mathcal{Z}) \le E^m_{\text{ada}}(\mathcal{K}, \mathcal{Z})$.*

Pietsch (Pietsch, 1974, Thm. 7.2) and Stesin (Stesin, 1975, Thm. 3) found explicit characterizations of the Kolmogorov m-width in finite sequence spaces. For this, let us recall from Section 2.2 the notation $B^p_{\boldsymbol{w}}(I)$ to denote the unit ball in $\ell^p_{\boldsymbol{w}}(I)$.

Theorem 5.6 (Characterization of Kolmogorov width)**.** *Let $N \in \mathbb{N}$ with $N > m$, $1 \le q < p \le \infty$, and $\boldsymbol{w} \in \mathbb{R}^N$ be a vector of positive weights. Then,*

$$d_m(B^p_{\boldsymbol{w}}([N]), \ell^q([N])) = \left(\max_{\substack{i_1, \ldots, i_{N-m} \in [N] \\ i_k \ne i_j}} \left(\sum_{j=1}^{N-m} w_{i_j}^{\frac{pq}{p-q}} \right)^{\frac{1}{p} - \frac{1}{q}} \right)^{-1}.$$

Pietsch showed that the Kolmogorov width and the Gelfand width are dual to each other (Pietsch, 1974, Thm. 6.2). For a direct proof of this fact in our setting we also refer to Adcock et al. (2024c, Thm. B.3). Recall that for a (possibly finite) sequence $\boldsymbol{w} = (w_i)_{i \in I}$ of nonvanishing real numbers, we write $1/\boldsymbol{w} := (1/w_i)_{i \in I}$ for the sequence of reciprocals.

Theorem 5.7 (Duality of Kolmogorov width and Gelfand width)**.** *For $1 \le p, q \le \infty$, let $\boldsymbol{w} \in \mathbb{R}^N$ be a vector of positive weights and let $1 \le p^*, q^* \le \infty$ be such that $1/p + 1/p^* = 1$ and $1/q + 1/q^* = 1$. Then $d_m(B^p([N]), \ell^q_{\boldsymbol{w}}([N])) = d^m(B^{q^*}_{1/\boldsymbol{w}}([N]), \ell^{p^*}([N]))$.*

Finally, we need one more technical lemma which states that suitably changing weights in sequence spaces does not change the Kolmogorov width.

Lemma 5.8 (Adcock et al. (2024c, Lem. B.4))**.** *Let $\boldsymbol{w} \in \mathbb{R}^N$ be a vector of positive weights and $1 \leq p, q \leq \infty$. Then $d_m(B^p([N]), \ell^q_{\boldsymbol{w}}([N])) = d_m(B^p_{1/\boldsymbol{w}}([N]), \ell^q([N]))$.*

5.3.2. Lower bound

Since $B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ is a subset of $B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, it suffices to prove the lower bound for the adaptive m-width in the case $\mathcal{K} = B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, that is,

$$\Theta_m(B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y});\mathcal{V}, L^2_\mu(\mathcal{X};\mathcal{Y})) \geq u_{\boldsymbol{\pi(m+1)}}, \quad \forall m \in \mathbb{N}. \tag{5.4}$$

This implies the same lower bound in the case $\mathcal{K} = B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. The proof consists of two main steps. We first reduce the problem to a discrete one which involves the adaptive compressive m-width of the unit ball in a space of suitably weighted finite sequences, see Lemma 5.9. In the discrete setting, we can then use Theorems 5.5 and 5.7, and Lemma 5.8 to relate the adaptive compressive m-width to the Kolmogorov m-width. We then apply Theorem 5.6 in combination with a limiting argument to conclude the claim.

For the next result, let us recall the notation $(c\boldsymbol{u})_I$ with $\boldsymbol{u} = (u_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma}$, $c \in \mathbb{R}$, and $I \subset \Gamma$ to denote the scaled subsequence $(cu_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in I}$.

Lemma 5.9 (Reduction to discrete problem)**.** *Let $I \subset \Gamma$ be a finite index set. Then, for every constant $c \in (0,1)$, we have*

$$\Theta_m(B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y});\mathcal{V}, L^2_\mu(\mathcal{X};\mathcal{Y})) \geq d^m(B^2_{(c\boldsymbol{u})_I}(I), \ell^2(I)).$$

The proof of this lemma is based on the construction of a suitable Lipschitz operator. To this end, let $R > 0$ and $n \in \mathbb{N}$, and consider the capped one-dimensional Hermite polynomials

$$\widetilde{H}_{n,R}(x) := \begin{cases} H_n(x) & \text{if } -R \leq x \leq R, \\ H_n(R) & \text{if } x > R, \\ H_n(-R) & \text{if } x < -R. \end{cases} \tag{5.5}$$

For $\boldsymbol{\gamma} \in \Gamma$ and $d \in \mathbb{N}$, we define

$$\widetilde{H}_{\boldsymbol{\gamma},R,d} : \mathbb{R}^d \to \mathbb{R}, \quad \widetilde{H}_{\boldsymbol{\gamma},R,d}(\boldsymbol{x}) := \prod_{i=1}^d \widetilde{H}_{\gamma_i,R}(x_i), \tag{5.6}$$

as well as

$$\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} : \mathcal{X} \to \mathbb{R}, \quad \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}(X) := \prod_{i=1}^\infty \widetilde{H}_{\gamma_i,R}\left(\frac{\langle X, \phi_i\rangle_{\mathcal{X}}}{\sqrt{\lambda_i}}\right). \tag{5.7}$$

Before proving Lemma 5.9, we need a couple of preliminary results.

Lemma 5.10 (Lipschitz continuity)**.** *For every $R > 0$ and every $\boldsymbol{\gamma} \in \Gamma$, the functional $\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} : \mathcal{X} \to \mathbb{R}$, defined in (5.7), is Lipschitz continuous.*

Proof Fix $R>0$ and $\boldsymbol{\gamma}\in\Gamma$ with $\operatorname{supp}(\boldsymbol{\gamma})\subset[d]$ for some $d\in\mathbb{N}$. We define the scaling functional

$$S_{\boldsymbol{\lambda},d}:\mathcal{X}\to\mathbb{R}^d,\quad S_{\boldsymbol{\lambda},d}(X):=\left(\frac{\langle X,\phi_i\rangle_{\mathcal{X}}}{\sqrt{\lambda_i}}\right)_{i\in[d]},$$

and note that $\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}=\widetilde{H}_{\boldsymbol{\gamma},R,d}\circ S_{\boldsymbol{\lambda},d}$, with $\widetilde{H}_{\boldsymbol{\gamma},R,d}$ given by (5.6). As $S_{\boldsymbol{\lambda},d}$ is Lipschitz continuous, it suffices to show that $\widetilde{H}_{\boldsymbol{\gamma},R,d}$ is Lipschitz continuous. This, in turn, follows by a simple induction argument on d. □

By Lemma 5.10 and Theorem 3.10, we have $\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}\in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ for every $R>0$. The next result establishes the connection between $\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}$ and $H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}$ in the limit $R\to\infty$. For its proof, we introduce the following notation: For $\boldsymbol{x}\in\mathbb{R}^d$, $d\in\mathbb{N}$, and $1\le k\le d$, we write $\boldsymbol{x}_{[k]}:=(x_1,\ldots,x_k)\in\mathbb{R}^k$. We also recall the complementary error function $\operatorname{erfc}:\mathbb{R}\to\mathbb{R}, x\mapsto \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$, which satisfies

$$\lim_{t\to\infty} t^m\operatorname{erfc}(t)=0,\quad \forall m\in\mathbb{N}. \tag{5.8}$$

Lemma 5.11 (Convergence in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X})$)**.** *For every $\boldsymbol{\gamma}\in\Gamma$, we have*

$$\lim_{R\to\infty}\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}=H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}\quad\text{in } W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X}).$$

Proof Let $\boldsymbol{\gamma}\in\Gamma$ with $\operatorname{supp}(\boldsymbol{\gamma})\subset[d]$ for some $d\in\mathbb{N}$. We first consider convergence in $L^2_\mu(\mathcal{X})$. By Fubini's theorem and a change of variables, one has

$$\left\|\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}-H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}\right\|^2_{L^2_\mu(\mathcal{X})}=\int_{\mathbb{R}^d}\left|\widetilde{H}_{\boldsymbol{\gamma},R,d}(\boldsymbol{x})-H_{\boldsymbol{\gamma},d}(\boldsymbol{x})\right|^2 d\mu_d(\boldsymbol{x}).$$

It thus suffices to show that

$$\lim_{R\to\infty}\widetilde{H}_{\boldsymbol{\gamma},R,d}=H_{\boldsymbol{\gamma},d}\quad\text{in } L^2_{\mu_d}(\mathbb{R}^d),\ \forall d\in\mathbb{N}. \tag{5.9}$$

For this, we use induction on d and start with $d=1$. For $n\in\mathbb{N}_0$, we compute

$$\begin{aligned}
&\int_{\mathbb{R}}\left|\widetilde{H}_{n,R}(x)-H_n(x)\right|^2 d\mu_1(x)\\
&=\int_{-\infty}^{-R}|H_n(-R)-H_n(x)|^2 d\mu_1(x)+\int_R^\infty |H_n(R)-H_n(x)|^2 d\mu_1(x)\\
&\le \left(H_n(-R)^2+H_n(R)^2\right)\operatorname{erfc}\left(\frac{R}{\sqrt{2}}\right)+2\int_{[-R,R]^c}H_n(x)^2 d\mu_1(x)\\
&=: T_1(R)+T_2(R),
\end{aligned}$$

where we used the notation $[-R,R]^c:=\mathbb{R}\setminus[-R,R]$. By (5.8), we have $\lim_{R\to\infty}T_1(R)=0$. Since $H_n\in L^2_{\mu_1}(\mathbb{R})$, the second term $T_2(R)$ converges to zero as $R\to\infty$ by the dominated convergence theorem. Next let $d>1$ and suppose that (5.9) holds for any $1 \le d' < d$. Without loss of generality we may assume $R \ge 1$. Then, by Fubini's theorem, we have

$$\begin{aligned}
&\int_{\mathbb{R}^d} \left|\widetilde{H}_{\boldsymbol{\gamma},R,d}(\boldsymbol{x}) - H_{\boldsymbol{\gamma},d}(\boldsymbol{x})\right|^2 d\mu_d(\boldsymbol{x}) \\
&\le 2\int_{\mathbb{R}}\int_{\mathbb{R}^{d-1}} \left|\widetilde{H}_{\boldsymbol{\gamma},d-1,R}(\boldsymbol{x}_{[d-1]})\widetilde{H}_{\gamma_d,R}(x_d) - \widetilde{H}_{\boldsymbol{\gamma},d-1,R}(\boldsymbol{x}_{[d-1]})H_{\gamma_d}(x_d)\right|^2 d\mu_{d-1}(\boldsymbol{x}_{[d-1]})d\mu_1(x_d) \\
&\quad + 2\int_{\mathbb{R}}\int_{\mathbb{R}^{d-1}} \left|\widetilde{H}_{\boldsymbol{\gamma},d-1,R}(\boldsymbol{x}_{[d-1]})H_{\gamma_d}(x_d) - H_{\boldsymbol{\gamma}_{[d-1]}}(\boldsymbol{x}_{[d-1]})H_{\gamma_d}(x_d)\right|^2 d\mu_{d-1}(\boldsymbol{x}_{[d-1]})d\mu_1(x_d) \\
&=: t_1(R) + t_2(R).
\end{aligned}$$

The term $t_1(R)$ can be bounded from above as

$$t_1(R) \le 2\sup_{R\ge 1}\int_{\mathbb{R}^{d-1}} \left|\widetilde{H}_{\boldsymbol{\gamma},d-1,R}(\boldsymbol{x}_{[d-1]})\right|^2 d\mu_{d-1}(\boldsymbol{x}_{[d-1]}) \int_{\mathbb{R}} \left|\widetilde{H}_{\gamma_d,R}(x_d) - H_{\gamma_d}(x_d)\right|^2 d\mu_1(x_d).$$

By induction hypothesis for $d' = d-1$, the term $\widetilde{H}_{\boldsymbol{\gamma},d-1,R}$ converges in $L^2_{\mu_{d-1}}(\mathbb{R}^{d-1})$ as $R \to \infty$. Hence, the supremum over $R \ge 1$ of the first integral is finite. Applying the induction hypotheses for $d' = 1$, we conclude that the second integral over $\mathbb{R}$ converges to 0 as $R \to \infty$. A similar argument shows that $\lim_{R\to\infty} t_2(R) = 0$. This completes the proof of (5.9).

Next, we consider convergence of $\nabla_{\mathcal{X}_{\boldsymbol{b}}}\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}$ in $L^2_\mu(\mathcal{X};\mathcal{X}_{\boldsymbol{b}})$. For this, we recall the basis $\{\eta_i\}_{i\in\mathbb{N}}$ of $\mathcal{X}_{\boldsymbol{b}}$, defined in (2.1), and the d-dimensional Hermite polynomials $H_{\boldsymbol{\gamma},d}$, defined in (2.2). Note that the capped Hermite polynomials $\widetilde{H}_{n,R}$, as defined in (5.5), lie in the Sobolev space $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$ of weakly differentiable functions which are up to their first derivatives locally integrable. We write $x_i := \langle X, \phi_i\rangle_{\mathcal{X}}$ for $X \in \mathcal{X}$ and $i \in [d]$. From Lemma C.11, it follows that $\widetilde{H}_{n,R} \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and for μ-a.e. $X \in \mathcal{X}$ the partial derivatives are given by

$$\begin{aligned}
\frac{\partial}{\partial\eta_i}\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}(X) &= b_i\lambda_i^{-1/2}\partial_i\widetilde{H}_{\boldsymbol{\gamma},R,d}(\lambda_1^{-1/2}x_1,\dots,\lambda_d^{-1/2}x_d) \\
&= \begin{cases} b_i\lambda_i^{-1/2}\partial_i H_{\boldsymbol{\gamma},d}(\lambda_1^{-1/2}x_1,\dots,\lambda_d^{-1/2}x_d) & \text{if } x_i \in [-R,R], \\ 0 & \text{if } x_i \in [-R,R]^c. \end{cases}
\end{aligned}$$

Moreover, $\frac{\partial}{\partial\eta_i}\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} = \frac{\partial}{\partial\eta_i}H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} = 0$ for $i > d$. Consequently,

$$\begin{aligned}
\left\|\nabla_{\mathcal{X}_{\boldsymbol{b}}}\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} - \nabla_{\mathcal{X}_{\boldsymbol{b}}}H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}\right\|^2_{L^2_\mu(\mathcal{X};\mathcal{X}_{\boldsymbol{b}})} &= \int_{\mathcal{X}}\sum_{i=1}^d \left|\frac{\partial}{\partial\eta_i}\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}(X) - \frac{\partial}{\partial\eta_i}H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}(X)\right|^2 d\mu(X) \\
&= \sum_{i=1}^d \int_{\mathbb{R}^{i-1}\times[-R,R]^c\times\mathbb{R}^{d-i}} \left|b_i\lambda_i^{-1/2}\partial_i H_{\boldsymbol{\gamma},d}(x_1,\dots,x_d)\right|^2 d\mu_d(\boldsymbol{x}).
\end{aligned}$$

Since $\partial_i H_{\boldsymbol{\gamma},d} \in L^2_{\mu_d}(\mathbb{R}^d)$, the right-hand side converges to zero as $R \to \infty$ by the dominated convergence theorem. The proof is now complete. □

The next lemma shows that an approximate version of Parseval's identity also holds for systems of finitely many of the capped infinite-dimensional Hermite polynomials $\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}$:

Lemma 5.12 (Riesz basis)**.** *Let $I \subset \Gamma$ be finite. Then, for every $\varepsilon > 0$, there exists $\bar{R} > 0$ such that for every $R \geq \bar{R}$, we have*

$$(1-\varepsilon)\|\boldsymbol{x}\|_{\ell^2(I)}^2 \leq \left\| \sum_{\boldsymbol{\gamma}\in I} x_{\boldsymbol{\gamma}} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} \right\|_{L^2_\mu(\mathcal{X})}^2 \leq (1+\varepsilon)\|\boldsymbol{x}\|_{\ell^2(I)}^2, \quad \forall \boldsymbol{x} = (x_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in I} \in \mathbb{R}^I. \tag{5.10}$$

In particular, $\{\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}\}_{\boldsymbol{\gamma}\in I}$ is a Riesz basis of $\operatorname{span}\{\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} : \boldsymbol{\gamma} \in I\}$ *for every $R \geq \bar{R}$ with Riesz constants at worst $1 \pm \varepsilon$.*

Proof We first work in $d = 1$ dimension and compute for $n, m \in \mathbb{N}_0$,

$$\begin{aligned}
&\left\langle \widetilde{H}_{n,R}, \widetilde{H}_{m,R} \right\rangle_{L^2_{\mu_1}(\mathbb{R})} = \int_{\mathbb{R}} \widetilde{H}_{n,R}(x)\widetilde{H}_{m,R}(x)d\mu_1(x) \\
&= \int_{-R}^{R} H_n(x)H_m(x)d\mu_1(x) + \int_{-\infty}^{-R} H_n(-R)H_m(-R)d\mu_1(x) + \int_{R}^{\infty} H_n(R)H_m(R)d\mu_1(x) \\
&= \int_{-R}^{R} H_n(x)H_m(x)d\mu_1(x) + \frac{1}{2}H_n(-R)H_m(-R)\operatorname{erfc}\left(\frac{R}{\sqrt{2}}\right) + \frac{1}{2}H_n(R)H_m(R)\operatorname{erfc}\left(\frac{R}{\sqrt{2}}\right) \\
&=: T_1(R) + T_2(R) + T_3(R).
\end{aligned}$$

Note that $H_n H_m$ is a polynomial of order $n+m$. Hence, by (5.8), we deduce $\lim_{R\to\infty} T_2(R) = 0$ and $\lim_{R\to\infty} T_3(R) = 0$. Moreover, the dominated convergence theorem and orthonormality of the Hermite polynomials imply that

$$\lim_{R\to\infty} T_1(R) = \int_{-\infty}^{\infty} H_n(x)H_m(x)d\mu_1(x) = \delta_{n,m}.$$

Altogether,

$$\lim_{R\to\infty} \left\langle \widetilde{H}_{n,R}, \widetilde{H}_{m,R} \right\rangle_{L^2_{\mu_1}(\mathbb{R})} = \delta_{n,m}. \tag{5.11}$$

Now, let $\boldsymbol{\gamma}, \boldsymbol{\gamma'} \in \Gamma$ with $\operatorname{supp}(\boldsymbol{\gamma}), \operatorname{supp}(\boldsymbol{\gamma'}) \subset [d]$ for some $d \in \mathbb{N}$. Since

$$\begin{aligned}
\left\langle \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}, \widetilde{H}_{\boldsymbol{\gamma'},R,\boldsymbol{\lambda}} \right\rangle_{L^2_\mu(\mathcal{X})} &= \int_{\mathbb{R}^d} \widetilde{H}_{\boldsymbol{\gamma},R,d}(\boldsymbol{x})\widetilde{H}_{\boldsymbol{\gamma'},d,R}(\boldsymbol{x})d\mu_d(\boldsymbol{x}) \\
&= \prod_{i=1}^{d} \int_{\mathbb{R}} \widetilde{H}_{\gamma_i,R}(x_i)\widetilde{H}_{\gamma_i',R}(x_i)d\mu_1(x_i),
\end{aligned}$$

we conclude by (5.11) that

$$\lim_{R\to\infty}\left\langle \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}, \widetilde{H}_{\boldsymbol{\gamma}',R,\boldsymbol{\lambda}}\right\rangle_{L^2_\mu(\mathcal{X})} = \delta_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}.$$

Next, fix some arbitrary $\varepsilon > 0$. Then there exists $\bar{R} > 0$ such that for every $R \geq \bar{R}$, we have

$$\left|\left\langle \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}, \widetilde{H}_{\boldsymbol{\gamma}',R,\boldsymbol{\lambda}}\right\rangle_{L^2_\mu(\mathcal{X})} - \delta_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}\right| \leq \varepsilon.$$

Since $I \subset \Gamma$ is finite, we obtain for any $\boldsymbol{x} = (x_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in I} \in \mathbb{R}^I$,

$$\begin{aligned}
&\left\|\sum_{\boldsymbol{\gamma}\in I} x_{\boldsymbol{\gamma}} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}\right\|^2_{L^2_\mu(\mathcal{X})} \\
&= \sum_{\boldsymbol{\gamma}\in I}\sum_{\boldsymbol{\gamma}'\in I \boldsymbol{\gamma}'\neq\boldsymbol{\gamma}} x_{\boldsymbol{\gamma}} x_{\boldsymbol{\gamma}'} \underbrace{\left\langle \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}, \widetilde{H}_{\boldsymbol{\gamma}',R,\boldsymbol{\lambda}}\right\rangle_{L^2_\mu(\mathcal{X})}}_{\leq\varepsilon} + \sum_{\boldsymbol{\gamma}\in\Gamma} x^2_{\boldsymbol{\gamma}} \underbrace{\left\langle \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}, \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}\right\rangle_{L^2_\mu(\mathcal{X})}}_{\leq 1+\varepsilon} \\
&\leq \varepsilon\|\boldsymbol{x}\|^2_{\ell^1(I)} + (1+\varepsilon)\|\boldsymbol{x}\|^2_{\ell^2(I)} \leq (\varepsilon|I| + 1 + \varepsilon)\|\boldsymbol{x}\|^2_{\ell^2(I)}.
\end{aligned}$$

Similarly, we have

$$\left\|\sum_{\boldsymbol{\gamma}\in I} x_{\boldsymbol{\gamma}} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}\right\|^2_{L^2_\mu(\mathcal{X})} \geq -\varepsilon\|\boldsymbol{x}\|^2_{\ell^1(I)} + (1-\varepsilon)\|\boldsymbol{x}\|^2_{\ell^2(I)} \geq (-\varepsilon|I| + 1 - \varepsilon)\|\boldsymbol{x}\|^2_{\ell^2(I)}.$$

As $\varepsilon > 0$ was arbitrary, the claim follows. □

We are now ready to reduce the adaptive m-width to a suitable Gelfand m-width:

Proof of Lemma 5.9 Let $\mathcal{L} : \mathcal{V} \to \mathcal{Y}^m$ be an adaptive sampling operator as in Definition 5.2. Then there exist $Y, \widetilde{Y} \in \mathcal{Y}\setminus\{0\}$ and a normed vector space $\widetilde{\mathcal{V}} \subset L^2_\mu(\mathcal{X})$ such that, if $YF \in \mathcal{V}$ for some $F \in L^2_\mu(\mathcal{X})$, then $F \in \widetilde{\mathcal{V}}$ and $\mathcal{L}(YF) = \widetilde{Y}\widetilde{\mathcal{L}}(F)$, where $\widetilde{\mathcal{L}} : \widetilde{\mathcal{V}} \to \mathbb{R}^m$ is a scalar-valued adaptive sampling operator as in Definition 5.1. Next, let $I \subset \Gamma$ be a finite subset and let us fix $c \in (0,1)$ and $\varepsilon > 0$. By Lemma 5.11 and Lemma 5.12, there exists $R > 0$ sufficiently large such that (5.10) holds and

$$\left\|\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} - H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}\right\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X})} \leq \frac{1-c}{c}|I|^{-1/2}, \quad \forall \boldsymbol{\gamma} \in I. \tag{5.12}$$

We fix some arbitrary sequence $\boldsymbol{x} = (x_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in I} \in \mathbb{R}^I$ with $\|\boldsymbol{x}\|_{\ell^2_{\boldsymbol{u}_I}(I)} \leq c$ and define $F_R, F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ by

$$F_R := \frac{Y}{\|Y\|_{\mathcal{Y}}}\sum_{\boldsymbol{\gamma}\in I} x_{\boldsymbol{\gamma}} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} \quad \text{and} \quad F := \frac{Y}{\|Y\|_{\mathcal{Y}}}\sum_{\boldsymbol{\gamma}\in I} x_{\boldsymbol{\gamma}} H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}.$$

Note that $\|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} = \|\boldsymbol{x}\|_{\ell^2_{\boldsymbol{u}_I}(I)} \leq c$ by Theorem 3.5, and therefore, by the triangle inequality and (5.12),

$$\|F_R - F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} \leq \sum_{\boldsymbol{\gamma}\in I} |x_{\boldsymbol{\gamma}}| \left\| \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} - H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} \right\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X})} \leq \frac{1-c}{c} |I|^{-1/2} \sum_{\boldsymbol{\gamma}\in I} |x_{\boldsymbol{\gamma}}| \leq 1-c.$$

Thus, by the triangle inequality, we have $\|F_R\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} \leq 1$. By Lemma 5.10 and since I is finite, the operator F_R is Lipschitz continuous and we conclude $F_R \in B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$.

By Lemma 5.12, the family of functionals $\{\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}\}_{\boldsymbol{\gamma}\in I}$ is a Riesz basis of $\mathcal{F} := \operatorname{span}\{\widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} : \boldsymbol{\gamma} \in I\}$. Hence, there exists a unique biorthogonal dual basis $\{\widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}\}_{\boldsymbol{\gamma}\in I}$ such that $\langle \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}, \widehat{H}_{\boldsymbol{\gamma'},R,\boldsymbol{\lambda}} \rangle = \delta_{\boldsymbol{\gamma},\boldsymbol{\gamma'}}$ for every $\boldsymbol{\gamma},\boldsymbol{\gamma'} \in I$. The orthogonal projection onto $\mathcal{F}$ is given by

$$P_{\mathcal{F}} : L^2_\mu(\mathcal{X}) \to \mathcal{F}, \quad P_{\mathcal{F}} g := \sum_{\boldsymbol{\gamma}\in I} \left\langle g, \widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} \right\rangle_{L^2_\mu(\mathcal{X})} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}.$$

Let $\{\psi_j\}_{j\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{Y}$. For $G \in L^2_\mu(\mathcal{X};\mathcal{Y})$, we write $g_j := \langle G, \psi_j \rangle_{\mathcal{Y}} \in L^2_\mu(\mathcal{X})$ and find by (5.10) that

$$\begin{aligned} \|G\|^2_{L^2_\mu(\mathcal{X};\mathcal{Y})} = \sum_{j=1}^\infty \|g_j\|^2_{L^2_\mu(\mathcal{X})} \geq \sum_{j=1}^\infty \|P_{\mathcal{F}} g_j\|^2_{L^2_\mu(\mathcal{X})} &\geq \sum_{j=1}^\infty (1-\varepsilon) \sum_{\boldsymbol{\gamma}\in I} \left| \left\langle g_j, \widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} \right\rangle_{L^2_\mu(\mathcal{X})} \right|^2 \\ &= (1-\varepsilon) \sum_{\boldsymbol{\gamma}\in I} \left\| \int_{\mathcal{X}} G \widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} d\mu \right\|^2_{\mathcal{Y}}. \end{aligned} \tag{5.13}$$

We now define the scalar-valued adaptive sampling operator

$$\Xi_R : \mathbb{R}^I \to \mathbb{R}^m, \quad \Xi_R(\boldsymbol{z}) := \widetilde{\mathcal{L}} \left(\frac{1}{\|Y\|_{\mathcal{Y}}} \sum_{\boldsymbol{\gamma}\in I} z_{\boldsymbol{\gamma}} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} \right).$$

We need to show that it is well-defined, that is, $\|Y\|^{-1}_{\mathcal{Y}} \sum_{\boldsymbol{\gamma}\in I} z_{\boldsymbol{\gamma}} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}} \in \widetilde{\mathcal{V}}$ for every $\boldsymbol{z} \in \mathbb{R}^I$. Since $B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}) \subset \mathcal{V}$, it suffices to observe that $Y\|Y\|^{-1}_{\mathcal{Y}} \sum_{\boldsymbol{\gamma}\in I} z_{\boldsymbol{\gamma}} \widetilde{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}$ is Lipschitz continuous as an operator from $\mathcal{X}$ to $\mathcal{Y}$ and it therefore lies in $\mathcal{V}$. Next, let $\mathcal{T} : \mathcal{Y}^m \to L^2_\mu(\mathcal{X};\mathcal{Y})$ be an arbitrary reconstruction map. We define $\widetilde{\mathcal{T}} : \mathbb{R}^m \to L^2_\mu(\mathcal{X};\mathcal{Y})$ by

$$\widetilde{\mathcal{T}} : \mathbb{R}^m \to L^2_\mu(\mathcal{X};\mathcal{Y}), \quad \widetilde{\mathcal{T}}(\boldsymbol{z}) := \mathcal{T}(\widetilde{Y}\boldsymbol{z}),$$

and observe that $\mathcal{T}(\mathcal{L}(F_R)) = \mathcal{T}(\widetilde{Y}\Xi_R(\boldsymbol{x})) = \widetilde{\mathcal{T}}(\Xi_R(\boldsymbol{x}))$. We now set $G := F_R - \mathcal{T}(\mathcal{L}(F_R))$ in (5.13). We use the estimate $\|Z\|_{\mathcal{Y}} \geq \|Y\|^{-1}_{\mathcal{Y}} |\langle Z, Y \rangle_{\mathcal{Y}}|$, which holds for every $Z \in \mathcal{Y}$ by the

Cauchy-Schwarz inequality, and compute

$$\begin{aligned}\|F_R - \mathcal{T}(\mathcal{L}(F_R))\|^2_{L^2_\mu(\mathcal{X};\mathcal{Y})} &\geq (1-\varepsilon)\sum_{\boldsymbol{\gamma}\in I}\left\|\int_{\mathcal{X}}(F_R - \mathcal{T}(\mathcal{L}(F_R)))\,\widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}d\mu\right\|^2_{\mathcal{Y}}\\ &= (1-\varepsilon)\sum_{\boldsymbol{\gamma}\in I}\left\|x_{\boldsymbol{\gamma}}\frac{Y}{\|Y\|_{\mathcal{Y}}} - \int_{\mathcal{X}}\widetilde{\mathcal{T}}(\Xi_R(\boldsymbol{x}))\widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}d\mu\right\|^2_{\mathcal{Y}}\\ &\geq (1-\varepsilon)\sum_{\boldsymbol{\gamma}\in I}\|Y\|^{-2}_{\mathcal{Y}}\left|\left\langle x_{\boldsymbol{\gamma}}\frac{Y}{\|Y\|_{\mathcal{Y}}} - \int_{\mathcal{X}}\widetilde{\mathcal{T}}(\Xi_R(\boldsymbol{x}))\widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}d\mu, Y\right\rangle_{\mathcal{Y}}\right|^2\\ &\geq (1-\varepsilon)\sum_{\boldsymbol{\gamma}\in I}\left|x_{\boldsymbol{\gamma}} - \frac{1}{\|Y\|_{\mathcal{Y}}}\int_{\mathcal{X}}\left\langle\widetilde{\mathcal{T}}(\Xi_R(\boldsymbol{x})), Y\right\rangle_{\mathcal{Y}}\widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}d\mu\right|^2.\end{aligned}$$

Finally, we define the (scalar-valued) reconstruction map

$$\Delta_R : \mathbb{R}^m \to \mathbb{R}^I, \quad \Delta_R(\boldsymbol{z}) := \left(\frac{1}{\|Y\|_{\mathcal{Y}}}\int_{\mathcal{X}}\left\langle\widetilde{\mathcal{T}}(\boldsymbol{z}), Y\right\rangle_{\mathcal{Y}}\widehat{H}_{\boldsymbol{\gamma},R,\boldsymbol{\lambda}}d\mu\right)_{\boldsymbol{\gamma}\in I},$$

and conclude that

$$\|F_R - \mathcal{T}(\mathcal{L}(F_R))\|^2_{L^2_\mu(\mathcal{X};\mathcal{Y})} \geq (1-\varepsilon)\|\boldsymbol{x} - \Delta_R(\Xi_R(\boldsymbol{x}))\|^2_{\ell^2(I)}.$$

We have thus shown that for any pair $(\mathcal{L},\mathcal{T})$ of a Hilbert-valued adaptive sampling operator and a reconstruction map, the error $\|F_R - \mathcal{T}(\mathcal{L}(F_R))\|_{L^2_\mu(\mathcal{X};\mathcal{Y})}$ can be bounded from below by $(1-\varepsilon)\|\boldsymbol{x} - \Delta_R(\Xi_R(\boldsymbol{x}))\|_{\ell^2(I)}$ for some pair (Ξ_R,Δ_R) of a scalar-valued adaptive sampling operator and reconstruction map. Recall that $F_R \in B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and $\boldsymbol{x}\in\mathbb{R}^I$ with $\|\boldsymbol{x}\|_{\ell^2_{\boldsymbol{u}_I}(I)} \leq c$ was arbitrary. Consequently,

$$\begin{aligned}&\Theta_m(B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y});\mathcal{V},L^2_\mu(\mathcal{X};\mathcal{Y}))\\ &= \inf\left\{\sup_{F\in B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})}\|F - \mathcal{T}(\mathcal{L}(F))\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} : \mathcal{L}:\mathcal{V}\to\mathcal{Y}^m \text{ adaptive}, \mathcal{T}:\mathcal{Y}^m\to L^2_\mu(\mathcal{X};\mathcal{Y})\right\}\\ &\geq (1-\varepsilon)^{1/2}\inf\left\{\sup_{\substack{\boldsymbol{x}\in\mathbb{R}^I\\ \|\boldsymbol{x}\|_{\ell^2_{\boldsymbol{u}_I}(I)}\leq c}}\|\boldsymbol{x} - \Delta(\Xi(\boldsymbol{x}))\|_{\ell^2(I)} : \Xi:\mathbb{R}^I\to\mathbb{R}^m \text{ adaptive}, \Delta:\mathbb{R}^m\to\mathbb{R}^I\right\}\\ &= (1-\varepsilon)^{1/2}E^m_{\mathrm{ada}}(cB^2_{\boldsymbol{u}_I}(I),\ell^2(I)) = (1-\varepsilon)^{1/2}E^m_{\mathrm{ada}}(B^2_{(c\boldsymbol{u})_I}(I),\ell^2(I)),\end{aligned}$$

where in the last step we used that $cB^2_{\boldsymbol{u}_I}(I) = B^2_{(c\boldsymbol{u})_I}(I)$. As $\varepsilon > 0$ was arbitrary, we can take the limit $\varepsilon \to 0^+$. The claim now follows by Theorem 5.5. $\square$

We can now finally prove the desired lower bound for the adaptive m-width.

Proof of (5.4) Let $N\in\mathbb{N}$ with $N>m$, $I=\pi([N])=\{\boldsymbol{\pi(1)},\dots,\boldsymbol{\pi(N)}\}\subset\Gamma$ be the index set of the N largest entries of $\boldsymbol{u}$, and fix $c\in(0,1)$. By Theorem 5.7 and Lemma 5.8, we have

$$\begin{aligned} d^m(B^2_{(c\boldsymbol{u})_I}(I),\ell^2(I)) &= d_m(B^2(I),\ell^2_{1/(c\boldsymbol{u})_I}(I)) \\ &= d_m(B^2_{(c\boldsymbol{u})_I}(I),\ell^2(I)) = c\cdot d_m(B^2_{\boldsymbol{u}_I}(I),\ell^2(I)), \end{aligned} \tag{5.14}$$

where the last equality again follows from $B^2_{(c\boldsymbol{u})_I}(I)=cB^2_{\boldsymbol{u}_I}(I)$. For every $p>2$ and $r=r(p):=1/2-1/p$, Hölder's inequality implies $N^{-r}B^p_{\boldsymbol{u}_I}(I)\subset B^2_{\boldsymbol{u}_I}(I)$. Consequently,

$$d_m(B^2_{\boldsymbol{u}_I}(I),\ell^2(I))\geq d_m(N^{-r}B^p_{\boldsymbol{u}_I}(I),\ell^2(I))=N^{-r}d_m(B^p_{\boldsymbol{u}_I}(I),\ell^2(I)). \tag{5.15}$$

Applying Theorem 5.6 with $q=2$ yields

$$d_m(B^p_{\boldsymbol{u}_I}(I),\ell^2(I))=\left(\max_{\substack{i_1,\dots,i_{N-m}\in I\\ i_k\neq i_j}}\left(\sum_{j=1}^{N-m}u_{i_j}^{\frac{2p}{p-2}}\right)^{\frac{1}{p}-\frac{1}{2}}\right)^{-1}.$$

Since $(u_{\boldsymbol{\pi(i)}})_{i\in\mathbb{N}}$ is nonincreasing, it follows with $q=q(p):=\frac{2p}{p-2}\in(2,\infty)$ that

$$\begin{aligned} d_m(B^p_{\boldsymbol{u}_I}(I),\ell^2(I)) = \min_{i_1,\dots,i_{N-m}\in I i_k\neq i_j}\left(\sum_{j=1}^{N-m}u_{i_j}^{\frac{2p}{p-2}}\right)^{\frac{1}{2}-\frac{1}{p}} &= \left(\sum_{j=m+1}^{N}u^q_{\boldsymbol{\pi(j)}}\right)^{1/q} \\ &\geq u_{\boldsymbol{\pi(m+1)}}. \end{aligned}$$

We combine this estimate with (5.14), (5.15), and Lemma 5.9, and conclude

$$u_{\boldsymbol{\pi(m+1)}}\leq c^{-1}N^r\Theta_m(B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y});\mathcal{V},L^2_\mu(\mathcal{X};\mathcal{Y})).$$

Taking the limit $p\to 2^+$ yields $r\to 0^+$ and therefore

$$u_{\boldsymbol{\pi(m+1)}}\leq c^{-1}\Theta_m(B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y});\mathcal{V},L^2_\mu(\mathcal{X};\mathcal{Y})).$$

As $c\in(0,1)$ was arbitrary, we can take the limit $c\to 1^-$, and the claim finally follows. □

5.3.3. Upper bound

Let $\mathcal{K}\in\{B_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y}),B^{\mathrm{Lip}}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})\}$. We now assume that $\mathcal{V}$ is continuously embedded in $L^2_\mu(\mathcal{X};\mathcal{Y})$ and prove the upper bound for the adaptive m-width, namely,

$$\begin{aligned} \Theta_m(\mathcal{K};\mathcal{V},L^2_\mu(\mathcal{X};\mathcal{Y})) &\leq \inf_{S\subset\Gamma,|S|\leq m}\sup_{F\in\mathcal{K}}\|F-F_S\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \\ &\leq \sup_{F\in\mathcal{K}}\left\|F-F_{\{\boldsymbol{\pi(1)},\dots,\boldsymbol{\pi(m)}\}}\right\|_{L^2_\mu(\mathcal{X};\mathcal{Y})}\leq u_{\boldsymbol{\pi(m+1)}}. \end{aligned} \tag{5.16}$$

Proof of (5.16) The second and third inequality hold by (4.1), so we only need to prove the first inequality. We fix $m \in \mathbb{N}$ and $S = \{\boldsymbol{\gamma}^{(\mathbf{1})}, \dots, \boldsymbol{\gamma}^{(\boldsymbol{n})}\} \subset \Gamma$ with $n \leq m$. We define the Hilbert-valued adaptive sampling operator

$$\mathcal{L} : \mathcal{V} \to \mathcal{Y}^m, \quad \mathcal{L}_i(F) := \begin{cases} \int_{\mathcal{X}} F H_{\boldsymbol{\gamma}^{(i)},\boldsymbol{\lambda}} d\mu & \text{if } 1 \leq i \leq n, \\ 0 & \text{if } n+1 \leq i \leq m, \end{cases}$$

and the reconstruction mapping

$$\mathcal{T} : \mathcal{Y}^m \to L^2_\mu(\mathcal{X};\mathcal{Y}), \quad \mathcal{T}(\boldsymbol{Y}) := \sum_{i=1}^m Y_i H_{\boldsymbol{\pi}(\boldsymbol{i}),\boldsymbol{\lambda}}.$$

Since $\mathcal{V}$ is continuously embedded in $L^2_\mu(\mathcal{X};\mathcal{Y})$, it is easy to see that $\mathcal{L}$ is a well-defined bounded linear operator. We need to show that it satisfies the conditions in Definition 5.2. It suffices to show that there exists $Y \in \mathcal{Y} \setminus \{0\}$, a normed vector space $\widetilde{\mathcal{V}} \subset L^2_\mu(\mathcal{X})$, and a scalar-valued adaptive sampling operator $\widetilde{\mathcal{L}} : \widetilde{\mathcal{V}} \to \mathbb{R}^m$ such that, if $YF \in \mathcal{V}$ for some $F \in L^2_\mu(\mathcal{X})$, then $F \in \widetilde{\mathcal{V}}$ and $\mathcal{L}(YF) = Y\widetilde{\mathcal{L}}(F)$. To this end, we choose some $Y \in \mathcal{Y}$ with $\|Y\|_{\mathcal{Y}} = 1$ and define the space $\widetilde{\mathcal{V}} := \left\{F \in L^2_\mu(\mathcal{X}) : YF \in \mathcal{V}\right\}$. It can be readily checked that this defines a normed vector space with norm given by $\|F\|_{\widetilde{\mathcal{V}}} := \|YF\|_{\mathcal{V}}$ for any $F \in \widetilde{\mathcal{V}}$. Moreover, as $\mathcal{V}$ is continuously embedded in $L^2_\mu(\mathcal{X};\mathcal{Y})$, there exists a constant $C > 0$ such that

$$\|F\|_{L^2_\mu(\mathcal{X})} = \|YF\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq C\|YF\|_{\mathcal{V}} = C\|F\|_{\widetilde{\mathcal{V}}}, \quad \forall F \in \widetilde{\mathcal{V}},$$

where in the first step we used the fact that $\|Y\|_{\mathcal{Y}} = 1$. This shows that $\widetilde{\mathcal{V}}$ is continuously embedded in $L^2_\mu(\mathcal{X})$. We now define the operator

$$\widetilde{\mathcal{L}} : \widetilde{\mathcal{V}} \to \mathbb{R}^m, \quad \widetilde{\mathcal{L}}_i(F) := \begin{cases} \int_{\mathcal{X}} F H_{\boldsymbol{\gamma}^{(i)},\boldsymbol{\lambda}} d\mu & \text{if } 1 \leq i \leq n, \\ 0 & \text{if } n+1 \leq i \leq m. \end{cases}$$

Note that $\widetilde{\mathcal{L}}$ is linear and by the continuous embedding of $\widetilde{\mathcal{V}}$ in $L^2_\mu(\mathcal{X})$, it is also bounded. Hence, $\widetilde{\mathcal{L}}$ is a scalar-valued adaptive sampling operator. Moreover, by construction, if $YF \in \mathcal{V}$, then $F \in \widetilde{\mathcal{V}}$ and $\mathcal{L}(YF) = Y\widetilde{\mathcal{L}}(F)$. We conclude that $\mathcal{L}$ is indeed an adaptive Hilbert-valued sampling operator as in Definition 5.2. Consequently,

$$\Theta_m(\mathcal{K};\mathcal{V}, L^2_\mu(\mathcal{X};\mathcal{Y})) \leq \sup_{F \in \mathcal{K}} \|F - \mathcal{T}(\mathcal{L}(F))\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} = \sup_{F \in \mathcal{K}} \|F - F_S\|_{L^2_\mu(\mathcal{X};\mathcal{Y})}.$$

As S was arbitrary, we now take the infimum over all subsets $S \subset \Gamma$ with $|S| \leq m$. $\square$

5.4. *Discussion*

Theorem 5.4 shows that linear Hermite polynomial approximation based on the index set $S = \{\boldsymbol{\pi}(\mathbf{1}), \dots, \boldsymbol{\pi}(\boldsymbol{m})\}$ is optimal for the uniform approximation of $W^{1,2}_{\mu,\boldsymbol{b}}$- and Lipschitz operators with Sobolev norm at most one among all possible recovery strategies which are based on (adaptive) linear information. Moreover, Theorem 5.4 in combination with Theorem 4.6 yields the following **curse of sample complexity**: *No strategy based on finitely many (potentially adaptively chosen) linear samples for the uniform recovery of all operators in the Sobolev unit (Lipschitz) ball can achieve algebraic convergence rates. This holds regardless of the decay rate of the PCA eigenvalues of the covariance operator of the underlying Gaussian measure.*

As already mentioned in the Section 1.1, a related result was previously shown in Kovachki et al. (2024a) by means of the so-called sampling nonlinear m-width $s_m(\mathcal{K})_{L^2_\mu(\mathcal{X})}$ of a set $\mathcal{K} \subset L^2_\mu(\mathcal{X})$, which is based on standard information. More specifically, the sampling operator $\delta_{\boldsymbol{X}} : \mathcal{K} \to \mathcal{Y}^m$ with fixed $\boldsymbol{X} = (X_1, \dots, X_m) \in \mathcal{X}^m$ is given by point evaluation, $\delta_{\boldsymbol{X}}(F) = (F(X_1), \dots, F(X_m)) \in \mathcal{Y}^m$ for every $F \in \mathcal{K}$, and one defines

$$s_m(\mathcal{K})_{L^2_\mu(\mathcal{X})} := \inf\left\{ \sup_{F \in \mathcal{K}} \|F - \mathcal{T}(\delta_{\boldsymbol{X}}(F))\|_{L^2_\mu(\mathcal{X})} : \boldsymbol{X} \in \mathcal{X}^m, \mathcal{T} : \mathcal{Y}^m \to L^2_\mu(\mathcal{X}) \right\}.$$

Theorem 2.12 in Kovachki et al. (2024a) implies the following result (for $p = 2$, $k = 1$), which was termed the *curse of data complexity.*

Theorem 5.13 (Curse of data complexity)**.** *Let μ be a centered Gaussian measure with at most algebraically decreasing (unweighted) PCA eigenvalues $\lambda_i \gtrsim i^{-\alpha}$ for some $\alpha > 0$. Then there exists a constant $C = C(\alpha) > 0$ such that*

$$s_m(\mathrm{Lip}(\mathcal{X}))_{L^2_\mu(\mathcal{X})} \geq C \log(m+1)^{-(\alpha+3)}, \quad \forall m \in \mathbb{N}.$$

Our findings in the present section generalize this result in several directions. First, recall that the adaptive m-width covers recovery strategies based on standard information by choosing $\mathcal{V} = C(\mathcal{X}, \mathcal{Y})$. This implies

$$s_m(\mathrm{Lip}(\mathcal{X}))_{L^2_\mu(\mathcal{X})} \geq \Theta_m(\mathrm{Lip}(\mathcal{X}); C(\mathcal{X}, \mathcal{Y}), L^2_\mu(\mathcal{X}; \mathcal{Y})) = u_{\boldsymbol{\pi}(\boldsymbol{m}+\boldsymbol{1})}.$$

This bound pertains to *general* centered, nondegenerate Gaussian measures and together with Theorem 4.6 we conclude that algebraic decays of the sampling m-width can never be achieved for *any* decay of the PCA eigenvalues. On the other hand, Theorem 4.7 provides upper bounds for the adaptive m-width of the Sobolev unit (Lipschitz) ball in terms of the decay of the PCA eigenvalues $\lambda_{\boldsymbol{b},i}$. In particular, it implies that the curse of sample complexity described above can be overcome asymptotically in the sense that the adaptive m-width can decay with rates which are arbitrarily close to any algebraic rate, provided the decay of the $\lambda_{\boldsymbol{b},i}$ is sufficiently fast, e.g., double exponential.

6. Conclusions

In this article, we analyzed the approximation of Hilbert-valued Lipschitz operators from finite data. We first extended results from infinite-dimensional analysis and showed that all Lipschitz operators lie in a (weighted) Gaussian Sobolev space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X}; \mathcal{Y})$. We then studied Hermite polynomial s-term approximations and proved that they cannot achieve algebraic convergence rates. This *curse of parametric complexity* is independent of the decay of the (weighted) PCA eigenvalues $\lambda_{\boldsymbol{b},i}$ of the covariance operator of the Gaussian measure μ. Next, we analyzed the smallest worst-case error in reconstructing Lipschitz and $W^{1,2}_{\mu,\boldsymbol{b}}$-operators from m arbitrary (potentially adaptively chosen) linear samples in terms of the corresponding adaptive m-width. We showed that from the point of view of information-based complexity there is no difference between the space of all Lipschitz operators, equipped with the $W^{1,2}_{\mu,\boldsymbol{b}}$-norm, and the space of all $W^{1,2}_{\mu,\boldsymbol{b}}$-operators—the adaptive m-widths in both cases coincide. We tightly characterized the dependence of the adaptive m-width on the $\lambda_{\boldsymbol{b},i}$ and identified a *curse of sample complexity*: No recovery strategy based on finite (adaptive) linear information can achieve algebraic convergence rates uniformly for all Lipschitz and $W^{1,2}_{\mu,\boldsymbol{b}}$-operators (of norm at most one). This holds for any (centered, nondegenerate) Gaussian measure independently of its spectral properties. It is an active area of research to identify classes of operators for which efficient learning in the sense of achieving algebraic (or faster) convergence rates is possible. As discussed in Section 1.1, examples include holomorphic operators and solution operators of certain PDEs. On the positive side, we proved that $W^{1,2}_{\mu,\boldsymbol{b}}$-regularity, and Lipschitz regularity in particular, suffices to achieve approximation rates which are arbitrarily close to any algebraic rate, provided that the PCA eigenvalues $\lambda_{\boldsymbol{b},i}$ decay sufficiently fast. As discussed in Section 1.3, there are several avenues for future work, including the construction of practical algorithms that achieve these rates and the investigation into classes to better describe operators of interest in applications.

Acknowledgments

A part of this work was done when GM was visiting the Department of Mathematics at Simon Fraser University, which he thanks for providing a fruitful research environment and working conditions. He also thanks BA for his invitation and hospitality as well as the Hausdorff School for Mathematics (HSM) in Bonn for financial support. The authors would also like to thank Mario Ullrich for useful discussions.

Funding

BA was supported by the the *Natural Sciences and Engineering Research Council of Canada* (NSERC) through grant RGPIN-2021-611675. MG and GM were supported by the *Hausdorff Center for Mathematics* (HCM) in Bonn, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2047/1 – 390685813 and by the CRC 1060 *The Mathematics of Emergent Effects* – 211504053 of the Deutsche Forschungsgemeinschaft.

References

B. Adcock, S. Brugiapaglia, and C. Webster. *Sparse Polynomial Approximation of High-Dimensional Functions.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2022.

B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. *On Efficient Algorithms for Computing Near-Best Polynomial Approximations to High-Dimensional, Hilbert-Valued Functions from Limited Samples*, volume 13 of *Mem. Eur. Math. Soc.* EMS Press, 1 edition, 2024a.

B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. Learning smooth functions in high dimensions. In *Handb. Numer. Anal.*, volume 25, pages 1–52. Elsevier, 2024b.

B. Adcock, N. Dexter, and S. Moraga. Optimal approximation of infinite-dimensional holomorphic functions. *Calcolo*, 61(1):12, 2024c.

B. Adcock, N. Dexter, and S. Moraga. Optimal deep learning of holomorphic operators between Banach spaces. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 27725–27789. Curran Associates, Inc., 2024d.

B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. Near-optimal learning of Banach-valued, high-dimensional functions via deep neural networks. *Neural Netw.*, 181:106761, 2025a.

B. Adcock, N. Dexter, and S. Moraga. Optimal approximation of infinite-dimensional holomorphic functions II: Recovery from i.i.d. pointwise samples. *J. Complexity*, 89:101933, 2025b.

N. Aronszajn. Differentiability of Lipschitzian mappings between Banach spaces. *Studia Math.*, 57(2): 147–190, 1976.

F. Bartel and D. Dũng. Sampling recovery in Bochner spaces and applications to parametric PDEs with random inputs. *Preprint*, 2024. arXiv:2409.05050.

A. Beged-dov. Lower and upper bounds for the number of lattice points in a simplex. *SIAM J. Appl. Math.*, 22(1):106–108, 1972.

J. Berezanskij, Z. Sheftel, and G. Us. *Functional Analysis. Vol. 2*, volume 2 of *Oper. Theory Adv. Appl.* Birkhäuser, Basel, 1996.

K. Bhattacharya, B. Hosseini, N. Kovachki, and A. Stuart. Model reduction and neural networks for parametric PDEs. *SMAI J. Comput. Math.*, 7:121–157, 2021.

V. Bogachev. *Gaussian Measures*, volume 62 of *Math. Surveys Monogr.* American Mathematical Society, Providence, RI, 1998.

N. Boullé and A. Townsend. A mathematical guide to operator learning. In *Handb. Numer. Anal.*, volume 25, pages 83–125. Elsevier, 2024.

N. Boullé, D. Halikias, and A. Townsend. Elliptic PDE learning is provably data-efficient. *Proceedings of the National Academy of Sciences*, 120(39):e2303904120, 2023.

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

T. Chen and H. Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Trans. Neural Netw.*, 6(4):911–917, 1995.

A. Chojnowska-Michalik and B. Goldys. Generalized Ornstein–Uhlenbeck semigroups: Littlewood–Paley–Stein inequalities and the P. A. Meyer equivalence of norms. *J. Funct. Anal.*, 182(2):243–279, 2001.

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

G. Da Prato. *An Introduction to Infinite-Dimensional Analysis.* Universitext. Springer Berlin, Heidelberg, 1 edition, 2006.

G. Da Prato. *Introduction to Stochastic Analysis and Malliavin Calculus*. Scuola Normale Superiore, Pisa, 2014.

G. Da Prato and J. Zabczyk. Regular densities of invariant measures in Hilbert spaces. *J. Funct. Anal.*, 130(2):427–449, 1995.

T. De Ryck and S. Mishra. Generic bounds on the approximation error for physics-informed (and) operator learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 10945–10958. Curran Associates, Inc., 2022.

D. Dũng and M. Griebel. Hyperbolic cross approximation in infinite dimensions. *J. Complexity*, 33: 55–88, 2016.

D. Dũng, V. Nguyen, and D. Pham. Deep ReLU neural network approximation in Bochner spaces and applications to parametric PDEs. *J. Complexity*, 79:101779, 2023.

S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing.* Appl. Numer. Harmon. Anal. Springer New York, New York, NY, 2013.

N. Franco and S. Brugiapaglia. A practical existence theorem for reduced order models based on convolutional autoencoders. *Found. Data Sci.*, 7(1):72–98, 2025.

M. Griebel and J. Oettershagen. On tensor product approximation of analytic functions. *J. Approx. Theory*, 207:348–379, 2016.

M. Griebel and P. Oswald. Stable splittings of Hilbert spaces of functions of infinitely many variables. *J. Complexity*, 41:126–151, 2017.

J. Gwinner, B. Jadamba, A. Khan, and F. Raciti. *Uncertainty Quantification in Variational Inequalities: Theory, Numerics, and Applications.* Chapman and Hall/CRC, New York, 2021.

A.-L. Haji-Ali, H. Harbrecht, M. Peters, and M. Siebenmorgen. Novel results for the anisotropic sparse grid quadrature. *J. Complexity*, 47:62–85, 2018.

L. Herrmann, C. Schwab, and J. Zech. Neural and spectral operator surrogates: Unified construction and expression rate bounds. *Adv. Comput. Math.*, 50(4):72, 2024.

T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces: Volume I: Martingales and Littlewood–Paley Theory.* Number 63 in Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer International Publishing, Cham, 1 edition, 2016.

N. Kovachki, S. Lanthaler, and S. Mishra. On universal approximation and error bounds for Fourier neural operators. *J. Mach. Learn. Res.*, 22(290):1–76, 2021.

N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *J. Mach. Learn. Res.*, 24(89):1–97, 2023.

N. Kovachki, S. Lanthaler, and H. Mhaskar. Data complexity estimates for operator learning. *Preprint*, 2024a. arXiv:2405.15992.

N. Kovachki, S. Lanthaler, and A. Stuart. Operator learning: Algorithms and analysis. In *Handb. Numer. Anal.*, volume 25, pages 419–467. Elsevier, 2024b.

D. Krieg, E. Novak, and M. Ullrich. On the power of adaption and randomization. *Preprint*, 2024. arXiv:2406.07108.

S. Lanthaler. Operator learning with PCA-Net: Upper and lower complexity bounds. *J. Mach. Learn. Res.*, 24(1):318:15014–318:15080, 2023.

S. Lanthaler. Operator learning of Lipschitz operators: An information-theoretic perspective. *Preprint*, 2024. arXiv:2406.18794.

S. Lanthaler and A. Stuart. The parametric complexity of operator learning. *Preprint*, 2024. arXiv:2306.15924.

S. Lanthaler, S. Mishra, and G. Karniadakis. Error estimates for DeepONets: A deep learning framework in infinite dimensions. *Trans. Math. Appl.*, 6(1):tnac001, 2022.

S. Lanthaler, Z. Li, and A. Stuart. Nonlocality and nonlinearity implies universality in operator learning. *Preprint*, 2024. arXiv:2304.13221.

Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *9th International Conference on Learning Representations*, 2021.

H. Liu, H. Yang, M. Chen, T. Zhao, and W. Liao. Deep nonparametric estimation of operators between infinite dimensional spaces. *J. Mach. Learn. Res.*, 25(24):1–67, 2024.

L. Lu, P. Jin, G. Pang, Z. Zhang, and G. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.*, 3(3):218–229, 2021.

A. Lunardi, M. Miranda, and D. Pallara. Infinite dimensional analysis. *19th Internet Seminar*, 2015/2016. URL https://www.mathematik.tu-darmstadt.de/media/analysis/lehrmaterial_anapde/hallerd/Lectures.pdf. Accessed: 2024-10-25.

H. Luo, X. Zhang, and S. Zhao. Sobolev embeddings in infinite dimensions. *Sci. China Math.*, 66(10): 2157–2178, 2023.

C. Milbradt and M. Wahl. High-probability bounds for the reconstruction error of PCA. *Stat. Probab. Lett.*, 161:108741, 2020.
E. Novak. On the power of adaption. *J. Complexity*, 12(3):199–237, 1996.
E. Novak and H. Woźniakowski. *Tractability of Multivariate Problems. Volume I, Linear information.* European Mathematical Society, Zurich, 2008.
J. Opschoor, C. Schwab, and J. Zech. Exponential ReLU DNN expression of holomorphic maps in high dimension. *Constr. Approx.*, 55(1):537–582, 2022.
A. Pietsch. s-Numbers of operators in Banach spaces. *Studia Math.*, 51(3):201–223, 1974.
M. Reiß and M. Wahl. Nonasymptotic upper bounds for the reconstruction error of PCA. *Ann. Statist.*, 48(2):1098–1123, 2020.
J. Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.*, 48(4):1875–1897, 2020.
C. Schwab and J. Zech. Deep learning in high dimension: Neural network expression rates for analytic functions in $L^2(\mathbb{R}^d,\gamma_d)$. *SIAM/ASA J. Uncertain. Quantif.*, 11(1):199–234, 2023.
C. Schwab, A. Stein, and J. Zech. Deep operator network approximation rates for Lipschitz operators. *Preprint*, 2023. arXiv:2307.09835.
I. Shigekawa. Sobolev spaces over the Wiener space based on an Ornstein–Uhlenbeck operator. *Kyoto J. Math.*, 32(4), 1992.
M. Stesin. Aleksandrov diameters of finite-dimensional sets and classes of smooth functions. *Dokl. Akad. Nauk SSSR*, 220(6):1278–1281, 1975.

A. Notions of differentiability

We recall several notions of differentiability which we use throughout the paper, loosely following Bogachev (1998, Chpt. 5.1). The constructions in this section hold for general Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. As usual, we denote by $L(\mathcal{X},\mathcal{Y})$ the space of all bounded linear operators F from $\mathcal{X}$ to $\mathcal{Y}$ with finite operator norm $\|F\|_{L(\mathcal{X},\mathcal{Y})} := \sup_{X\in\mathcal{X},X\neq 0}\|F(X)\|_{\mathcal{Y}}/\|X\|_{\mathcal{X}}$. We write $L(\mathcal{X}) := L(\mathcal{X},\mathbb{R})$.

Definition A.1 (Differentiability)**.** Let $\mathcal{M}$ be a collection of non-empty subsets of $\mathcal{X}$, $X\in\mathcal{X}$, and Ω be an open neighborhood of X. A mapping $F:\Omega\to\mathcal{Y}$ is said to be differentiable with respect to $\mathcal{M}$ at the point X if there exists a continuous linear mapping $\ell\in L(\mathcal{X},\mathcal{Y})$ such that for every fixed set $M\in\mathcal{M}$, we have

$$\lim_{t\to 0}\sup_{Z\in M}\left\|\frac{F(X+tZ)-F(X)}{t}-\ell(Z)\right\|_{\mathcal{Y}}=0.$$

In this case, ℓ is unique and we write $D^{\mathcal{M}}F(X) := \ell$ for the derivative of F at X.

If $\mathcal{M}$ is the class of all *finite*, *compact*, or *bounded* subsets of $\mathcal{X}$, then we say that F is *Gâteaux*, *Hadamard*, or *Fréchet differentiable at* X, respectively. In the latter case, we also often just call F *differentiable* at X. If $\mathcal{X}$ is finite-dimensional, it is easy to see that Hadamard and Fréchet differentiability at X are equivalent and the corresponding derivatives of F at X coincide. This will become important in the proof of Theorem 3.10 in Appendix D. We henceforth drop the superscript $\mathcal{M}$ in the notation of the derivative and write $DF(X)$ instead of $D^{\mathcal{M}}F(X)$. In the following, we will focus on Fréchet differentiability, unless stated otherwise.

We call F *Fréchet differentiable* (or just *differentiable*) if it is Fréchet differentiable at every point $X \in \mathcal{X}$. The resulting derivative DF is a mapping from $\mathcal{X}$ to $L(\mathcal{X},\mathcal{Y})$.

If $\mathcal{E}$ is a linear subspace of $\mathcal{X}$ (possibly with a stronger norm), then we say that F is *differentiable along* $\mathcal{E}$ *at the point* X if the mapping $Z \mapsto F(X+Z)$ is differentiable from $\mathcal{E}$ to $\mathcal{Y}$ at $Z=0$. We call F *differentiable along* $\mathcal{E}$ if it is differentiable along $\mathcal{E}$ at every point $X \in \mathcal{X}$. The resulting derivative $D_{\mathcal{E}}F$ is a mapping from $\mathcal{X}$ to $L(\mathcal{E},\mathcal{Y})$. Moreover, the differential operator $D_{\mathcal{E}}$, mapping F to its derivative $D_{\mathcal{E}}F$, is linear in F.

Example A.2. If $\mathcal{E}=\mathcal{X}$, then $D_{\mathcal{E}}F = DF$. If F is Fréchet differentiable and we choose $\mathcal{E}$ to be the Cameron-Martin space $\mathcal{H}$ of a Gaussian measure on $\mathcal{X}$ (see Appendix C.1), then $D_{\mathcal{E}}F$ is the $\mathcal{H}$-derivative of F which is commonly used in infinite-dimensional analysis (see Lunardi et al., 2015/2016, Sect. 9).

If $\mathcal{E} = \mathrm{span}\{Z\}$, $Z \in \mathcal{X} \setminus \{0\}$, is one-dimensional, we obtain the usual directional derivative

$$\frac{\partial}{\partial Z}F(X) := D_{\mathrm{span}\{Z\}}F(X)(Z) = \lim_{t\to 0}\frac{F(X+tZ)-F(X)}{t} \in \mathcal{Y}.$$

In this case, the Gâteaux, Hadamard, and Fréchet derivatives along $\mathcal{E}$ at X coincide. For any subspaces $\mathcal{E}' \subset \mathcal{E} \subset \mathcal{X}$, it can be readily seen that if $F:\mathcal{X}\to\mathcal{Y}$ is differentiable along $\mathcal{E}$ at a point $X \in \mathcal{X}$, then F is also differentiable along $\mathcal{E}'$ at X and $D_{\mathcal{E}}F(X)|_{\mathcal{E}'} = D_{\mathcal{E}'}F(X)$. In particular, if F is differentiable along $\mathcal{E}$ at X, then, for any $Z \in \mathcal{E}\setminus\{0\}$, the directional derivative $\frac{\partial}{\partial Z}F(X)$ at X exists and

$$\frac{\partial}{\partial Z}F(X) = D_{\mathcal{E}}F(X)(Z).$$

If $\mathcal{X} = \mathbb{R}^n$ and $Z = \boldsymbol{e_i}$ is the ith standard unit vector, we use the standard notation $\partial_i F(\boldsymbol{x}) := \frac{\partial}{\partial \boldsymbol{e_i}}F(\boldsymbol{x})$ for $\boldsymbol{x} = (x_1,\ldots,x_n)\in\mathbb{R}^n$. If $\mathcal{E}$ is a Hilbert subspace of $\mathcal{X}$ and if $F:\mathcal{X}\to\mathbb{R}$ is differentiable along $\mathcal{E}$ at X, then, by the Riesz representation theorem, there exists a unique $Z\in\mathcal{E}$ such that $D_{\mathcal{E}}F(X)(X') = \langle Z, X'\rangle_{\mathcal{E}}$ for every $X'\in\mathcal{E}$. In this case, we call Z the $\mathcal{E}$*-gradient of* F *at* X and write $\nabla_{\mathcal{E}}F(X) := Z$.

Definition A.3 (The space $C_b^1(\mathcal{X})$)**.** We denote by $C_b^1(\mathcal{X})$ the space of all *boundedly (Fréchet) differentiable functionals on* $\mathcal{X}$, that is, the set of all Fréchet differentiable mappings $F:\mathcal{X}\to\mathbb{R}$ which are bounded on $\mathcal{X}$ and whose derivative DF is bounded in $L(\mathcal{X})$. The corresponding norm is given by

$$\|F\|_{C_b^1(\mathcal{X})} := \sup_{X\in\mathcal{X}}|F(X)| + \|DF\|_{L(\mathcal{X})}.$$

B. Closability and closure of operators

We define closability and the closure of operators between Hilbert spaces and state standard properties. For further details we refer to Berezanskij et al. (1996, Chpt. 12). Let $\mathcal{H}_1,\mathcal{H}_2$ be two Hilbert spaces. A linear $\mathcal{H}_2$-valued operator (not necessarily bounded) acting on $\mathcal{H}_1$ is a linear mapping $A:\mathrm{dom}(A)\to\mathcal{H}_2$ from a linear subspace $\mathrm{dom}(A)\subset\mathcal{H}_1$ to $\mathcal{H}_2$. The set $\mathrm{dom}(A)$

is called the *domain* of A. The *graph* of A is defined as the set

$$\Gamma_A := \{(H, AH) \in \mathcal{H}_1 \oplus \mathcal{H}_2 : H \in \operatorname{dom}(A)\}.$$

Considered as a subspace of the direct sum $\mathcal{H}_1 \oplus \mathcal{H}_2$ and equipped with the *graph inner product*,

$$\langle H, K\rangle_{\Gamma_A} := \langle H, K\rangle_{\mathcal{H}_1} + \langle AH, AK\rangle_{\mathcal{H}_2}, \quad H, K \in \operatorname{dom}(A),$$

this becomes a Hilbert space with *graph norm*

$$\|(H, AH)\|_{\Gamma_A} := \sqrt{\langle H, H\rangle_{\Gamma_A}} = \left(\|H\|^2_{\mathcal{H}_1} + \|AH\|^2_{\mathcal{H}_2}\right)^{1/2}.$$

Definition B.1 (Closability and closure)**.** A linear operator $A : \operatorname{dom}(A) \to \mathcal{H}_2$ is called *closable* (in $\mathcal{H}_1$) if the closure of its graph $\overline{\Gamma}_A$ in $\mathcal{H}_1 \oplus \mathcal{H}_2$ is the graph of some (necessarily unique) linear operator, that is, there exists a linear operator $\overline{A} : \operatorname{dom}(\overline{A}) \to \mathcal{H}_2$ such that $\overline{\Gamma}_A = \Gamma_{\overline{A}}$. In this case, we call $\overline{A}$ the *closure* of A.

If A is closable, then the domain of its closure is given by

$$\operatorname{dom}(\overline{A}) = \left\{H \in \mathcal{H}_1 : \exists (H_n)_{n\in\mathbb{N}} \subset \operatorname{dom}(A) : \lim_{n\to\infty} H_n = H,\ (AH_n)_{n\in\mathbb{N}} \text{ converges in } \mathcal{H}_2\right\}.$$

For $H \in \operatorname{dom}(\overline{A})$, we have $AH_n \to \overline{A}H$ in $\mathcal{H}_2$ for every sequence $(H_n)_{n\in\mathbb{N}} \subset \operatorname{dom}(\overline{A})$ with $H_n \to H$ in $\mathcal{H}_1$, and the limit $\lim_{n\to\infty} AH_n$ is independent of the sequence $(H_n)_{n\in\mathbb{N}}$ (cf. Berezanskij et al. (1996, Thm. 2.1)). If A is closable, we equip $\operatorname{dom}(\overline{A})$ with the graph inner product, which turns $(\operatorname{dom}(\overline{A}), \langle\cdot,\cdot\rangle_{\Gamma_{\overline{A}}})$ into a Hilbert space.

C. Results from infinite-dimensional analysis

We recall some well-known results from infinite-dimensional analysis which are used to define the Gaussian Sobolev space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ (see Definition 3.1) and to prove Theorem 3.10 in Appendix D. We first define the Cameron-Martin space $\mathcal{H}$ of μ in $\mathcal{X}$ in Appendix C.1 and then discuss the construction of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ as well as some of its important properties in Appendix C.2. We mainly follow Lunardi et al. (2015/2016) and Da Prato (2006), which consider $\mathcal{Y} = \mathbb{R}$ and the cases $\boldsymbol{b} = \sqrt{\boldsymbol{\lambda}}$ and $\boldsymbol{b} = \mathbf{1}$, respectively, and generalize the proofs therein to general Hilbert-valued operators in the case $\mathbf{0} < \boldsymbol{b} \leq \mathbf{1}$. More information can also be found in Bogachev (1998). Throughout this section, we use notation as introduced in Sections 2 and 3.

C.1. *The Cameron-Martin space*

Theorem C.1 (Fernique's theorem (Bogachev, 1998, Thm. 2.8.5))**.** *There exists* $\alpha > 0$ *such that*

$$\int_{\mathcal{X}} \exp(\alpha\|X\|^2_{\mathcal{X}})d\mu(X) < \infty.$$

Fernique's theorem implies that any mapping $\mathcal{X} \to \mathcal{Y}$ which grows at most polynomially at infinity belongs to $L^2_\mu(\mathcal{X};\mathcal{Y})$. In particular, the mapping

$$j : \mathcal{X}^* \to L^2_\mu(\mathcal{X}), \quad F \mapsto j(F) = F,$$

is well-defined. We use it to define the Cameron-Martin space:

Definition C.2 (Cameron-Martin space)**.** The *Cameron-Martin space of μ (in $\mathcal{X}$)* is the set $\mathcal{H}$ of all $X \in \mathcal{X}$ whose $\mathcal{H}$-norm is finite, where

$$\|X\|_{\mathcal{H}} := \sup\left\{F(X) : F \in \mathcal{X}^*, \|j(F)\|_{L^2_\mu(\mathcal{X})} \leq 1\right\}.$$

To further describe the structure of $\mathcal{H}$, we define the *reproducing kernel Hilbert space* $\mathcal{X}^*_\mu$ of μ as the closure of $j(\mathcal{X}^*)$ in $L^2_\mu(\mathcal{X})$ as well as the mapping

$$R_\mu : \mathcal{X}^*_\mu \to \mathcal{X}, \; F \mapsto \int_{\mathcal{X}} XF(X) d\mu(X),$$

where the integral is in the sense of Bochner. Notice that R_μ is well-defined by Fernique's theorem.

Proposition C.3 (Relation between $\mathcal{H}$ and $\mathcal{X}^*_\mu$ (Lunardi et al., 2015/2016, Prop. 3.1.2))**.** *An element $H \in \mathcal{X}$ belongs to $\mathcal{H}$ if and only if there exists $\hat{H} \in \mathcal{X}^*_\mu$ such that $H = R_\mu(\hat{H})$. In this case, we have $\|H\|_{\mathcal{H}} = \|\hat{H}\|_{L^2_\mu(\mathcal{X})}$. Hence, $R_\mu : \mathcal{X}^*_\mu \to \mathcal{H}$ is an isometric isomorphism that turns $\mathcal{H}$ into a Hilbert space with inner product $\langle H, K\rangle_{\mathcal{H}} := \langle \hat{H}, \hat{K}\rangle_{L^2_\mu(\mathcal{X})}$ whenever $H = R_\mu(\hat{H})$ and $K = R_\mu(\hat{K})$.*

In our case, where $\mathcal{X}$ is a separable Hilbert space, the Cameron-Martin space has a particularly simple structure. Indeed, it follows from Lunardi et al. (2015/2016, Thm. 4.2.7) that the family of vectors $\{\xi_i\}_{i\in\mathbb{N}}$ with

$$\xi_i := \sqrt{\lambda_i}\phi_i, \quad \forall i \in \mathbb{N}, \tag{C.1}$$

is an orthonormal basis of $\mathcal{H}$. The corresponding elements $\hat{\xi}_i \in \mathcal{X}^*_\mu$ are given by

$$\hat{\xi}_i(\cdot) = \lambda_i^{-1/2} \langle \cdot, \phi_i\rangle_{\mathcal{X}}, \quad \forall i \in \mathbb{N}. \tag{C.2}$$

The right-hand side in (C.2) is an element of $\mathcal{X}^*$ and we henceforth identify each $\hat{\xi}_i$ with its version in $\mathcal{X}^*$.

C.2. *The Gaussian Sobolev space $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$*

Recall from Section 2.3 the weighted space $\mathcal{X}_{\boldsymbol{b}}$ with weight sequence $\boldsymbol{b} = (b_i)_{i\in\mathbb{N}}$, $\mathbf{0} < \boldsymbol{b} \leq \mathbf{1}$, and with orthonormal basis $\{\eta_i\}_{i\in\mathbb{N}}$, $\eta_i := b_i\phi_i$, as defined in (2.1).

C.2.1. Construction

The construction of Gaussian Sobolev spaces is based on so-called *cylindrical functionals*:

Definition C.4 (Cylindrical functionals and operators)**.** A functional $\varphi : \mathcal{X} \to \mathbb{R}$ is called a *cylindrical functional* if there exist $n \in \mathbb{N}$, $\ell_1, \dots, \ell_n \in \mathcal{X}^*$, and a function $\omega : \mathbb{R}^n \to \mathbb{R}$ such that

$$\varphi(X) = \omega(\ell_1(X), \dots, \ell_n(X)), \quad \forall X \in \mathcal{X}.$$

We call φ a *cylindrical boundedly Fréchet differentiable functional* if, with the above notation, $\omega \in C_b^1(\mathbb{R}^n)$. The space of all such functionals is denoted by $\mathcal{F}C_b^1(\mathcal{X})$. Moreover, we define the set of all *cylindrical boundedly Fréchet differentiable* $\mathcal{Y}$*-valued operators* by

$$\mathcal{F}C_b^1(\mathcal{X},\mathcal{Y}) := \operatorname{span}\left\{\mathcal{X} \ni X \mapsto \varphi(X)Y \in \mathcal{Y} : \varphi \in \mathcal{F}C_b^1(\mathcal{X}), Y \in \mathcal{Y}\right\}$$

so that every $F \in \mathcal{F}C_b^1(\mathcal{X},\mathcal{Y})$ can be written as $F = \sum_{i=1}^n \varphi_i Y_i$ for some $\varphi_i \in \mathcal{F}C_b^1(\mathcal{X})$, $Y_i \in \mathcal{Y}$, and $n \in \mathbb{N}$.

Proposition C.5 (Closability of $D_{\mathcal{X}_{\boldsymbol{b}}}$)**.** *The (Fréchet) differential operator along* $\mathcal{X}_{\boldsymbol{b}}$*,* $D_{\mathcal{X}_{\boldsymbol{b}}} : \mathcal{F}C_b^1(\mathcal{X};\mathcal{Y}) \to L_\mu^2(\mathcal{X}; HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))$*, is closable in* $L_\mu^2(\mathcal{X};\mathcal{Y})$*.*

Proof First, a short computation shows that for every $F \in \mathcal{F}C_b^1(\mathcal{X},\mathcal{Y})$, the derivative $D_{\mathcal{X}_{\boldsymbol{b}}}F$ lies in $L_\mu^2(\mathcal{X}; HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))$. Hence, the mapping $D_{\mathcal{X}_{\boldsymbol{b}}} : \mathcal{F}C_b^1(\mathcal{X};\mathcal{Y}) \to L_\mu^2(\mathcal{X}; HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))$ is well-defined. The proof of closability is a straight-forward modification of the proof of Lunardi et al. (2015/2016, Lem. 10.2.4), replacing the derivative along the Cameron-Martin space by the derivative along the space $\mathcal{X}_{\boldsymbol{b}}$. □

C.2.2. Properties

We commence with an important criterion for an operator to belong to $W_{\mu,\boldsymbol{b}}^{1,2}(\mathcal{X};\mathcal{Y})$:

Lemma C.6. *If* $F_n \to F$ *in* $L_\mu^2(\mathcal{X};\mathcal{Y})$ *as* $n \to \infty$ *and* $\sup_{n\in\mathbb{N}} \|F_n\|_{W_{\mu,\boldsymbol{b}}^{1,2}(\mathcal{X};\mathcal{Y})} < \infty$*, then* $F \in W_{\mu,\boldsymbol{b}}^{1,2}(\mathcal{X};\mathcal{Y})$*.*

Proof Since $W_{\mu,\boldsymbol{b}}^{1,2}(\mathcal{X};\mathcal{Y})$ is a Hilbert space, it is reflexive. As $(F_n)_{n\in\mathbb{N}}$ is bounded in $W_{\mu,\boldsymbol{b}}^{1,2}(\mathcal{X};\mathcal{Y})$ by assumption, there exists a subsequence $(F_{n_k})_{k\in\mathbb{N}}$ which converges weakly in $W_{\mu,\boldsymbol{b}}^{1,2}(\mathcal{X};\mathcal{Y})$ to some G as $k \to \infty$. Since $F_{n_k} \to F$ in $L_\mu^2(\mathcal{X};\mathcal{Y})$ by assumption, we conclude that $F = G$, and the claim follows. □

Next, we consider the ℓ^2-characterization of $W_{\mu,\boldsymbol{b}}^{1,2}(\mathcal{X};\mathcal{Y})$, see Theorem 3.5. For $\boldsymbol{\gamma} \in \Gamma$ and $i \in \mathbb{N}$, we define $\boldsymbol{\gamma}^{(\boldsymbol{i})} = (\gamma_k^{(i)})_{k\in\mathbb{N}} \in \Gamma$ as follows: If $\gamma_i = 0$, we set $\boldsymbol{\gamma}^{(\boldsymbol{i})} := \mathbf{0}$, and if $\gamma_i > 0$, we set

$$\gamma_k^{(i)} := \begin{cases} \gamma_k - 1 & \text{if } k = i, \\ \gamma_k & \text{if } k \neq i. \end{cases}$$

Proposition C.7. *Let $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. Then,*

$$\frac{\partial}{\partial \eta_i} F = \sum_{\boldsymbol{\gamma}\in\Gamma} b_i \sqrt{\frac{\gamma_i}{\lambda_i}} \left(\int_{\mathcal{X}} F H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu \right) H_{\boldsymbol{\gamma}^{(i)},\boldsymbol{\lambda}}, \quad \forall i \in \mathbb{N}, \tag{C.3}$$

and

$$\|F\|^2_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} = \sum_{\boldsymbol{\gamma}\in\Gamma} \left(1 + \sum_{i=1}^{\infty} b_i^2 \frac{\gamma_i}{\lambda_i} \right) \left\| \int_{\mathcal{X}} F H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu \right\|^2_{\mathcal{Y}}. \tag{C.4}$$

Conversely, if for a family of vectors $(Y_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma} \subset \mathcal{Y}$ one has

$$\sum_{\boldsymbol{\gamma}\in\Gamma} \left(1 + \sum_{i=1}^{\infty} b_i^2 \frac{\gamma_i}{\lambda_i} \right) \|Y_{\boldsymbol{\gamma}}\|^2_{\mathcal{Y}} < \infty, \tag{C.5}$$

then $F := \sum_{\boldsymbol{\gamma}\in\Gamma} Y_{\boldsymbol{\gamma}} H_{\boldsymbol{\gamma},\boldsymbol{\lambda}}$ lies in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$.

Proof First, let us fix a cylindrical functional $\varphi \in \mathcal{F}C^1_b(\mathcal{X})$. By Da Prato (2006, Lemma 10.14), the partial derivatives of φ satisfy

$$\frac{\partial}{\partial \phi_i} \varphi = \sum_{\boldsymbol{\gamma}\in\Gamma} \sqrt{\frac{\gamma_i}{\lambda_i}} \left(\int_{\mathcal{X}} \varphi H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu \right) H_{\boldsymbol{\gamma}^{(i)},\boldsymbol{\lambda}}, \quad \forall i \in \mathbb{N}. \tag{C.6}$$

Since $\frac{\partial}{\partial \eta_i}\varphi = b_i \frac{\partial}{\partial \phi_i}\varphi$, it follows from (C.6) together with orthonormality of the Hermite polynomials that

$$\int_{\mathcal{X}} \left(\frac{\partial}{\partial \eta_i} \varphi \right) H_{\boldsymbol{\gamma}^{(i)},\boldsymbol{\lambda}} d\mu = b_i \sqrt{\frac{\gamma_i}{\lambda_i}} \int_{\mathcal{X}} \varphi H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu, \quad \forall i \in \mathbb{N}. \tag{C.7}$$

By linearity, the identities (C.6) and (C.7) hold, in fact, for every cylindrical *operator* $\varphi \in \mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$.

Now suppose that $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. It suffices to prove (C.3) as (C.4) then follows by Parseval's identity (2.5). For this, it is enough to show that

$$\int_{\mathcal{X}} \left(\frac{\partial}{\partial \eta_i} F \right) H_{\boldsymbol{\gamma}^{(i)},\boldsymbol{\lambda}} d\mu = b_i \sqrt{\frac{\gamma_i}{\lambda_i}} \int_{\mathcal{X}} F H_{\boldsymbol{\gamma},\boldsymbol{\lambda}} d\mu, \quad \forall i \in \mathbb{N}. \tag{C.8}$$

By definition of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, there exists a sequence $(F_n)_{n\in\mathbb{N}} \subset \mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$ such that $\lim_{n\to\infty} F_n = F$ and $\lim_{n\to\infty} \frac{\partial}{\partial \eta_i} F_n = \frac{\partial}{\partial \eta_i} F$ in $L^2_\mu(\mathcal{X};\mathcal{Y})$. We set $\varphi = F_n$ in (C.7) and take the limit $n \to \infty$ to obtain (C.8).

Conversely, suppose that (C.5) holds for a sequence $(Y_{\boldsymbol{\gamma}})_{\boldsymbol{\gamma}\in\Gamma} \subset \mathcal{Y}$. We fix an enumeration $\tau : \mathbb{N} \to \Gamma$ of Γ and define

$$F_n := \sum_{j=1}^{n} Y_{\boldsymbol{\tau}(\boldsymbol{j})} H_{\boldsymbol{\tau}(\boldsymbol{j}),\boldsymbol{\lambda}}, \quad \forall n \in \mathbb{N}.$$

By Parseval's identity, we can bound the $L^2_\mu(\mathcal{X};\mathcal{Y})$-norm of the F_n by

$$\|F_n\|^2_{L^2_\mu(\mathcal{X};\mathcal{Y})} = \sum_{j=1}^{n} \left\|Y_{\boldsymbol{\tau}(\boldsymbol{j})}\right\|^2_{\mathcal{Y}} \le \sum_{\boldsymbol{\gamma}\in\Gamma} \|Y_{\boldsymbol{\gamma}}\|^2_{\mathcal{Y}} \tag{C.9}$$

and the right-hand side is finite by (C.5). This implies, in particular, that $(F_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2_\mu(\mathcal{X};\mathcal{Y})$. Hence, there exists some $F \in L^2_\mu(\mathcal{X};\mathcal{Y})$ such that $F_n \to F$ in $L^2_\mu(\mathcal{X};\mathcal{Y})$ as $n \to \infty$. Since $F_n \in \mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$, we can set $\varphi = F_n$ in (C.6) and obtain

$$\frac{\partial}{\partial \eta_i} F_n = \sum_{j=1}^{n} b_i \sqrt{\frac{\tau(j)_i}{\lambda_i}} Y_{\boldsymbol{\tau}(\boldsymbol{j})} H_{\boldsymbol{\tau}(\boldsymbol{j})^{(i)},\boldsymbol{\lambda}}, \quad \forall i \in \mathbb{N}.$$

Consequently, we can bound the $L^2_\mu(\mathcal{X};HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))$-norm of the derivatives $D_{\mathcal{X}_{\boldsymbol{b}}} F_n$ by

$$\begin{aligned}
&\int_{\mathcal{X}} \|D_{\mathcal{X}_{\boldsymbol{b}}} F_n(X)\|^2_{HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y})} d\mu(X) = \int_{\mathcal{X}} \sum_{i=1}^{\infty} \left\| \frac{\partial}{\partial \eta_i} F_n(X) \right\|^2_{\mathcal{Y}} d\mu(X) \\
&= \sum_{i=1}^{\infty} \int_{\mathcal{X}} \left\| \sum_{j=1}^{n} b_i \sqrt{\frac{\tau(j)_i}{\lambda_i}} Y_{\boldsymbol{\tau}(\boldsymbol{j})} H_{\boldsymbol{\tau}(\boldsymbol{j})^{(i)},\boldsymbol{\lambda}}(X) \right\|^2_{\mathcal{Y}} d\mu(X) = \sum_{i=1}^{\infty} \sum_{j=1}^{n} b_i^2 \frac{\tau(j)_i}{\lambda_i} \left\|Y_{\boldsymbol{\tau}(\boldsymbol{j})}\right\|^2_{\mathcal{Y}} \\
&\le \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} b_i^2 \frac{\tau(j)_i}{\lambda_i} \left\|Y_{\boldsymbol{\tau}(\boldsymbol{j})}\right\|^2_{\mathcal{Y}},
\end{aligned} \tag{C.10}$$

and the right-hand side is again finite by (C.5). Combining (C.9) and (C.10), we conclude that the F_n are uniformly bounded in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. By Lemma C.6, it then follows that $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. □

We now consider operators with a special structure. They will become important in the proof of Theorem 3.10 in Appendix D. To this end, recall from (C.1), (C.2) the orthonormal basis $\{\xi_i\}_{i\in\mathbb{N}}$ of $\mathcal{H}$ with $\xi_i = \sqrt{\lambda_i}\phi_i$, $R_\mu(\hat{\xi}_i) = \xi_i$, and $\hat{\xi}_i \in \mathcal{X}^*$. For every $F \in L^2_\mu(\mathcal{X};\mathcal{Y})$ and $n \in \mathbb{N}$, we define

$$\mathbb{E}_n F : \mathcal{X} \to \mathcal{Y}, \quad \mathbb{E}_n F := \mathbb{E}[F \mid \hat{\xi}_1, \dots, \hat{\xi}_n], \tag{C.11}$$

to be the conditional expectation of F with respect to the σ-algebra generated by the random variables $\hat{\xi}_1, \dots, \hat{\xi}_n$. Furthermore, for $n \in \mathbb{N}$, we define the mapping

$$P_n : \mathcal{X} \to \operatorname{span}\{\hat{\xi}_i : i \in [n]\}, \quad P_n X := \sum_{i=1}^{n} \hat{\xi}_i(X) \xi_i. \tag{C.12}$$

Proposition C.8 (Properties of $\mathbb{E}_n F$ in L^2_μ)**.** *For $F \in L^2_\mu(\mathcal{X};\mathcal{Y})$, let $\mathbb{E}_n F$ and P_n be given as in (C.11) and (C.12), respectively. Then the following properties hold true:*

(i) *For every $F \in L^2_\mu(\mathcal{X};\mathcal{Y})$ and $n \in \mathbb{N}$, we have*

$$\mathbb{E}_n F(X) = \int_{\mathcal{X}} F(P_n X + (I - P_n)Z) d\mu(Z) \quad \text{for } \mu\text{-a.e. } X \in \mathcal{X}.$$

In particular, $\mathbb{E}_n F$ can be identified with an operator F_n on $P_n(\mathcal{X})$ by setting $F_n(Z) := \mathbb{E}_n F(X)$ for $Z = P_n X$.

(ii) *For every $F \in L^2_\mu(\mathcal{X};\mathcal{Y})$, the sequence $(\mathbb{E}_n F)_{n\in\mathbb{N}}$ converges to F in $L^2_\mu(\mathcal{X};\mathcal{Y})$.*

Proof We refer to the proofs of Proposition 7.4.1 and Proposition 7.4.4, respectively, in Lunardi et al. (2015/2016), which can be adopted almost verbatim, only changing the Lebesgue integrals to Bochner integrals. □

Proposition C.9 (Properties of $\mathbb{E}_n F$ in $W^{1,2}_{\mu,\boldsymbol{b}}$)**.** *Let $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and $\mathbb{E}_n F$ be defined as in (C.11). Then, $\mathbb{E}_n F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} \mathbb{E}_n F = F$ in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$.*

The case of $\mathcal{H}$-differentiable functionals, that is, $\boldsymbol{b} = \sqrt{\boldsymbol{\lambda}}$ (see Remark 2.1) and $\mathcal{Y} = \mathbb{R}$, is covered by Lunardi et al. (2015/2016, Prop. 10.1.2) which (for $p = 2$) reads as follows:

Proposition C.10. *Let $F \in W^{1,2}_{\mu,\sqrt{\boldsymbol{\lambda}}}(\mathcal{X})$. Then, for every $n \in \mathbb{N}$, we have $\mathbb{E}_n F \in W^{1,2}_{\mu,\sqrt{\boldsymbol{\lambda}}}(\mathcal{X})$ and the following properties hold true:*

(i) *For every $i \in \mathbb{N}$, the ith partial derivative of $\mathbb{E}_n F$ is given by*

$$\frac{\partial}{\partial \xi_i} \mathbb{E}_n F = \begin{cases} \mathbb{E}_n(\frac{\partial}{\partial \xi_i} F) & \text{if } i \leq n, \\ 0 & \text{if } i > n. \end{cases}$$

(ii) *We have $\lim_{n\to\infty} \mathbb{E}_n F = F$ in $W^{1,2}_{\mu,\sqrt{\boldsymbol{\lambda}}}(\mathcal{X})$.*

Proof of Proposition C.9 We first prove the claim for $F \in \mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$ and then for general operators in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ by a density argument. To this end, let us fix $F \in \mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$ and $n \in \mathbb{N}$. By linearity, it suffices to consider cylindrical operators of the form $F(\cdot) = \varphi(\cdot)Y$ with $\varphi \in \mathcal{F}C^1_b(\mathcal{X})$ and $Y \in \mathcal{Y}$. Since $\eta_i = b_i \lambda_i^{-1/2} \xi_i$, we have

$$\frac{\partial}{\partial \eta_i} \mathbb{E}_n F(X) = b_i \lambda_i^{-1/2} \frac{\partial}{\partial \xi_i} \mathbb{E}_n F(X) = b_i \lambda_i^{-1/2} Y \frac{\partial}{\partial \xi_i} \mathbb{E}_n \varphi(X), \quad \forall X \in \mathcal{X},\ \forall i \in \mathbb{N}.$$

Using Parseval's identity, Proposition C.10(i), and the contraction property of the conditional expectation in L^p-spaces, we compute

$$\begin{aligned}
&\left\|D_{\mathcal{X}_{\boldsymbol{b}}}\mathbb{E}_n F\right\|^2_{L^2_\mu(\mathcal{X};HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))} = \int_{\mathcal{X}} \sum_{i=1}^{\infty} \left\| \frac{\partial}{\partial \eta_i} \mathbb{E}_n F(X) \right\|^2_{\mathcal{Y}} d\mu(X) \\
&= \int_{\mathcal{X}} \sum_{i=1}^{\infty} b_i^2 \lambda_i^{-1} \|Y\|^2_{\mathcal{Y}} \left| \frac{\partial}{\partial \xi_i} \mathbb{E}_n \varphi(X) \right|^2 d\mu(X) = \sum_{i=1}^{n} b_i^2 \lambda_i^{-1} \|Y\|^2_{\mathcal{Y}} \left\| \mathbb{E}_n \left(\frac{\partial}{\partial \xi_i} \varphi \right) \right\|^2_{L^2_\mu(\mathcal{X})} \\
&\leq \sum_{i=1}^{n} b_i^2 \lambda_i^{-1} \|Y\|^2_{\mathcal{Y}} \left\| \frac{\partial}{\partial \xi_i} \varphi \right\|^2_{L^2_\mu(\mathcal{X})} \leq \int_{\mathcal{X}} \sum_{i=1}^{\infty} \left\| \frac{\partial}{\partial \eta_i} F(X) \right\|^2_{\mathcal{Y}} d\mu(X) \\
&= \left\|D_{\mathcal{X}_{\boldsymbol{b}}} F\right\|^2_{L^2_\mu(\mathcal{X};HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))}.
\end{aligned} \tag{C.13}$$

Since, in addition, $\|\mathbb{E}_n F\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq \|F\|_{L^2_\mu(\mathcal{X};\mathcal{Y})}$ by the contraction property, we conclude that $\|\mathbb{E}_n F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} \leq \|F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})}$. In particular, this shows that $\mathbb{E}_n F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ for every $n \in \mathbb{N}$. To prove convergence in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, we first note that by Proposition C.8(ii), $\mathbb{E}_n F$ converges to F in $L^2_\mu(\mathcal{X};\mathcal{Y})$ as $n \to \infty$. Since φ is cylindrical, it can be written as $\varphi(X) = \omega(\hat{\xi}_1(X), \ldots, \hat{\xi}_k(X))$ for some $k \in \mathbb{N}$ and $\omega \in C^1_b(\mathbb{R}^k)$. This implies

$$\left\|D_{\mathcal{X}_{\boldsymbol{b}}}\mathbb{E}_n F - D_{\mathcal{X}_{\boldsymbol{b}}} F\right\|_{L^2_\mu(\mathcal{X};HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))} \leq \big(\max_{i\in[k]} b_i^2 \lambda_i^{-1}\big) \|Y\|_{\mathcal{Y}} \left\|\nabla_{\mathcal{X}_{\sqrt{\boldsymbol{\lambda}}}} \mathbb{E}_n \varphi - \nabla_{\mathcal{X}_{\sqrt{\boldsymbol{\lambda}}}} \varphi\right\|_{L^2_\mu(\mathcal{X};\mathcal{X}_{\sqrt{\boldsymbol{\lambda}}})},$$

and the right-hand side converges to 0 as $n \to \infty$ by Proposition C.10(ii). Altogether, we conclude that $\lim_{n\to\infty} \mathbb{E}_n F = F$ in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ for all $F \in \mathcal{F}C^1_b(\mathcal{X},\mathcal{Y})$. The claim for general $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ then follows from a density argument and similar estimates as above. □

We provide one more technical lemma which asserts that operators of a certain form lie in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. To this end, recall that $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n;\mathcal{Y})$ denotes the Sobolev space of all weakly differentiable mappings $\mathbb{R}^n \to \mathcal{Y}$ which are up to their first derivative locally integrable (with respect to Lebesgue measure $\mathcal{L}^n$). As usual, we set $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n;\mathbb{R}) = W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$. Moreover, for $X \in \mathcal{X}$ and $i \in \mathbb{N}$, we use the notation $x_i := \langle X, \phi_i \rangle_{\mathcal{X}}$, where $\{\phi_i\}_{i\in\mathbb{N}}$ is the PCA basis of $\mathcal{X}$. In particular, for $\hat{\xi}_i \in \mathcal{X}^*$ given by (C.2) we have $\lambda_i^{-1/2} x_i = \hat{\xi}_i(X)$.

Lemma C.11. *Let $F : \mathcal{X} \to \mathcal{Y}$ be an operator of the form $F = \omega \circ (\hat{\xi}_1, \ldots, \hat{\xi}_n)$ with $\omega \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n;\mathcal{Y})$ and $n \in \mathbb{N}$. Then, $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ and*

$$\frac{\partial}{\partial \eta_i} F(X) = b_i \lambda_i^{-1/2} \partial_i \omega(\lambda_1^{-1/2} x_1, \ldots, \lambda_n^{-1/2} x_n)$$

for μ-a.e. $X \in \mathcal{X}$ and every $i \in [n]$.

Proof For brevity, we write $\hat{\boldsymbol{\xi}} := (\hat{\xi}_1, \ldots, \hat{\xi}_n)$. For $R > 0$, let $\varphi_R : \mathbb{R}^m \to \mathbb{R}$ be a smooth bump function such that $\varphi_R = 1$ on $B_R := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_{\mathbb{R}^n} < R\}$ and $\mathrm{supp}(\varphi_R) \subset B_{2R}$. We set $\omega_R := \omega\varphi_R$. By a standard approximation argument (see, e.g., Brezis (2011, Thm 9.2) adapted to the Hilbert-valued case by replacing Lebesgue integrals by Bochner integrals), there exists a sequence $(\omega_{R,j})_{j\in\mathbb{N}} \subset C_c^\infty(\mathbb{R}^n;\mathcal{Y})$ of smooth, compactly supported operators such that $\lim_{j\to\infty}\omega_{R,j} = \omega_R$ and $\lim_{j\to\infty}\partial_i\omega_{R,j} = \partial_i\omega_R$ in $L^2(\mathbb{R}^n;\mathcal{Y})$ for every $i\in[n]$. We define $F_{R,j} := \omega_{R,j}\circ\hat{\boldsymbol{\xi}}$ and note that $F_{R,j}\in\mathcal{F}C_b^1(\mathcal{X},\mathcal{Y})$. By construction, taking the limit $j\to\infty$ yields

$$F_{R,j}\to\omega_R\circ\hat{\boldsymbol{\xi}} \quad \text{in } L^2_\mu(\mathcal{X};\mathcal{Y}) \tag{C.14}$$

as well as

$$\frac{\partial}{\partial\eta_i}F_{R,j}\to b_i\lambda_i^{-1/2}(\partial_i\omega_R)\circ\hat{\boldsymbol{\xi}} \quad \text{in } L^2_\mu(\mathcal{X};\mathcal{Y}) \text{ for every } i\in[n]. \tag{C.15}$$

The right hand-sides in (C.14) and (C.15) converge for $R\to\infty$ in $L^2_\mu(\mathcal{X})$ to ω and $b_i\lambda_i^{-1/2}(\partial_i\omega\circ\hat{\boldsymbol{\xi}})$, respectively. Hence, taking a suitable subsequence of radii $R(j)$ shows that for $j\to\infty$,

$$F_{R(j),j}\to\omega\circ\hat{\boldsymbol{\xi}} \quad \text{in } L^2_\mu(\mathcal{X};\mathcal{Y})$$

as well as

$$\frac{\partial}{\partial\eta_i}F_{R(j),j}\to b_i\lambda_i^{-1/2}\partial_i\omega\circ\hat{\boldsymbol{\xi}} \quad \text{in } L^2_\mu(\mathcal{X};\mathcal{Y}) \text{ for every } i\in[n].$$

Moreover, we have $\frac{\partial}{\partial\eta_i}F_{R(j),j} = 0$ for $i>n$ and every $j\in\mathbb{N}$. This yields the claim by definition of $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$, see Definition 3.1 and recall Appendix B. □

D. Proof of Theorem 3.10

The argument follows along the lines of the proof of Lunardi et al. (2015/2016, Prop. 10.1.4) with several important modifications. Let $F:\mathcal{X}\to\mathcal{Y}$ be Lipschitz continuous with $L := [F]_{\mathrm{Lip}(\mathcal{X},\mathcal{Y})}$. The idea is to use Lemma C.6 to show that $F\in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. By Lipschitz continuity, we have $\|F(X)\|_{\mathcal{Y}}\le\|F(0)\|_{\mathcal{Y}}+L\|X\|_{\mathcal{X}}$ for every $X\in\mathcal{X}$ and therefore, by the Fernique theorem (Theorem C.1), $F\in L^2_\mu(\mathcal{X};\mathcal{Y})$. As approximating sequence to F we take $F_n := \mathbb{E}_nF$, as defined in (C.11). By Proposition C.8, we have $\lim_{n\to\infty}\mathbb{E}_nF = F$ in $L^2_\mu(\mathcal{X};\mathcal{Y})$ and we can write $\mathbb{E}_nF(X) = V_n(T_n(X))$ for some $V_n:\mathbb{R}^n\to\mathcal{Y}$ and $T_n:\mathcal{X}\to\mathbb{R}^n$, $T_n(X) := (\hat{\xi}_1(X),\dots,\hat{\xi}_n(X))$ with $\xi_i,\hat{\xi}_i$ defined in (C.1), (C.2). Note that V_n inherits Lipschitz continuity from F. Indeed, using Proposition C.8(i) and the fact that $\hat{\xi}_i(\xi_j)=\delta_{i,j}$, we find for $\boldsymbol{x},\bar{\boldsymbol{x}}\in\mathbb{R}^n$ that

$$\begin{aligned}
\|V_n(\boldsymbol{x}+\bar{\boldsymbol{x}})-V_n(\boldsymbol{x})\|_{\mathcal{Y}} &= \left\|\mathbb{E}_nF\left(\sum_{i=1}^n x_i\xi_i+\sum_{i=1}^n\bar{x}_i\xi_i\right)-\mathbb{E}_nF\left(\sum_{i=1}^n x_i\xi_i\right)\right\|_{\mathcal{Y}}\\
&\le\int_{\mathcal{X}}\left\|F\left(\sum_{i=1}^n x_i\xi_i+\sum_{i=1}^n\bar{x}_i\xi_i+(I-P_n)Z\right)-F\left(\sum_{i=1}^n x_i\xi_i+(I-P_n)Z\right)\right\|_{\mathcal{Y}}d\mu(Z)\\
&\le L\left\|\sum_{i=1}^n\bar{x}_i\xi_i\right\|_{\mathcal{X}} = L\left\|\sum_{i=1}^n\sqrt{\lambda_i}\bar{x}_i\phi_i\right\|_{\mathcal{X}} = L\left(\sum_{i=1}^n\lambda_i\bar{x}_i^2\right)^{1/2} &\text{(D.1)}\\
&\le L\left(\max_{i\in[n]}\sqrt{\lambda_i}\right)\|\bar{\boldsymbol{x}}\|_{\mathbb{R}^n}. &\text{(D.2)}
\end{aligned}$$

Proof of (i). Suppose that $\dim(\mathcal{Y}) = m \in \mathbb{N}$ and let $\{\psi_j\}_{j\in[m]}$ be an orthonormal basis of $\mathcal{Y}$. For $j \in [m]$, we define the function(al)s

$$V_n^{(j)} : \mathbb{R}^n \to \mathbb{R}, \quad \boldsymbol{x} \mapsto \langle V_n(\boldsymbol{x}), \psi_j \rangle_{\mathcal{Y}}$$
$$\mathbb{E}_n^{(j)} F : \mathcal{X} \to \mathbb{R}, \quad X \mapsto \langle \mathbb{E}_n F(X), \psi_j \rangle_{\mathcal{Y}} = V_n^{(j)}(T_n(X)).$$

$V_n^{(j)}$ is Lipschitz continuous due to (D.2). By Rademacher's theorem, it is therefore Fréchet differentiable $\mathcal{L}^n$-almost everywhere and, in particular, it belongs to $W^{1,1}_{\text{loc}}(\mathbb{R}^n)$. By Lemma C.11, we then obtain

$$D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n^{(j)} F)(X)(\eta_i) = \begin{cases} b_i \lambda_i^{-1/2} \partial_i V_n^{(j)}(T_n(X)) & \text{if } 1 \le i \le n, \\ 0 & \text{otherwise.} \end{cases}$$

at any point $X \in \mathcal{X}$ such that $V_n^{(j)}$ is differentiable at $T_n(X)$. To derive an upper bound for $|\partial_i V_n^{(j)}(\boldsymbol{x})|$, $\boldsymbol{x} \in \mathbb{R}^n$, we equip $\mathbb{R}^n$ with a rescaled Euclidean inner product,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbb{R}^n_{\boldsymbol{\lambda}}} := \sum_{i=1}^n \sqrt{\lambda_i} x_i y_i$$

with induced norm $\|\boldsymbol{x}\|_{\mathbb{R}^n_{\boldsymbol{\lambda}}} := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathbb{R}^n_{\boldsymbol{\lambda}}}}$. We denote the hereby defined space as $\mathbb{R}^n_{\boldsymbol{\lambda}}$ and observe that $\{\lambda_i^{-1/2} \boldsymbol{e_i}\}_{i\in\mathbb{N}}$ is an orthonormal basis. Now note that (D.1) implies that $V_n^{(j)}$ is L-Lipschitz as a function from $\mathbb{R}^n_{\boldsymbol{\lambda}}$ to $\mathbb{R}$. Hence, for $\mathcal{L}^n$-almost every $\boldsymbol{x} \in \mathbb{R}^n$ and $1 \le i \le n$, it follows

$$\begin{aligned} \sum_{i=1}^n \lambda_i^{-1} |\partial_i V_n^{(j)}(\boldsymbol{x})|^2 &= \sum_{i=1}^n |D V_n^{(j)}(\boldsymbol{x})(\lambda_i^{-1/2} \boldsymbol{e_i})|^2 \\ &= \|D V_n^{(j)}(\boldsymbol{x})\|^2_{HS(\mathbb{R}^n_{\boldsymbol{\lambda}}, \mathbb{R})} = \|D V_n^{(j)}(\boldsymbol{x})\|^2_{L(\mathbb{R}^n_{\boldsymbol{\lambda}}, \mathbb{R})} \le L^2. \end{aligned} \tag{D.3}$$

Consequently, whenever $V_n^{(j)}$ is differentiable at $T_n(X)$, we have for every $j \in [m]$,

$$\begin{aligned} &\left\| D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n F)(X) \right\|^2_{HS(\mathcal{X}_{\boldsymbol{b}}, \mathcal{Y})} = \left\| \sum_{j=1}^m D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n^{(j)} F)(X) \psi_j \right\|^2_{HS(\mathcal{X}_{\boldsymbol{b}}, \mathcal{Y})} \\ &= \sum_{i=1}^\infty \left\| \sum_{j=1}^m D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n^{(j)} F)(X)(\eta_i) \psi_j \right\|^2_{\mathcal{Y}} = \sum_{i=1}^\infty \sum_{j=1}^m |D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n^{(j)} F)(X)(\eta_i)|^2 \\ &= \sum_{i=1}^n \sum_{j=1}^m b_i^2 \lambda_i^{-1} |\partial_i V_n^{(j)}(T_n(X))|^2 \le m L^2, \end{aligned} \tag{D.4}$$

where in the last step we used (D.3) and the fact that $b_i \le 1$ for every $i \in \mathbb{N}$.

Next, we show that for μ-a.e. $X \in \mathcal{X}$ and every $j \in [m]$, $V_n^{(j)}$ is Fréchet differentiable at $T_n(X)$. To this end, let $A \subset \mathbb{R}^n$ be the set such that for every j, $V_n^{(j)}$ is differentiable at every point $\boldsymbol{x} \in \mathbb{R}^n \setminus A$. Then, each $V_n^{(j)}$ is Fréchet differentiable at any point $T_n(X)$ with $X \in \mathcal{X} \setminus T_n^{-1}(A)$. It thus suffices to show that $\mu(T_n^{-1}(A)) = 0$. We know that $\mathcal{L}^n(A) = 0$, and since μ_n is absolutely continuous with respect to $\mathcal{L}^n$, it follows that $\mu_n(A) = 0$. Moreover, it is easy to see that μ_n is equal to the push-forward measure $(T_n)_\sharp\mu$, and thus $\mu(T_n^{-1}(A)) = 0$. Hence, (D.4) holds for μ-a.e. $X \in \mathcal{X}$, which implies

$$\int_{\mathcal{X}} \left\| D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n F)(X) \right\|^2_{HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y})} d\mu(X) \leq mL^2, \quad \forall n \in \mathbb{N}.$$

As $\mathbb{E}_n F \to F$ in $L^2_\mu(\mathcal{X};\mathcal{Y})$, we can now use Lemma C.6 to conclude that $F \in W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. This shows that $\mathrm{Lip}(\mathcal{X},\mathcal{Y}) \subset W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$.

Finally, we show that $C^{0,1}(\mathcal{X},\mathcal{Y})$ is continuously embedded in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$. By (D.4), we know that $\left\| D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n F)(X) \right\|_{L^2_\mu(\mathcal{X};HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y}))} \leq \sqrt{m}L$. Moreover, from Proposition C.8(i) it follows that $\|\mathbb{E}_n F\|_{L^2_\mu(\mathcal{X};\mathcal{Y})} \leq \sup_{X \in \mathcal{X}} \|F(X)\|_{\mathcal{Y}}$. In total, we have

$$\|\mathbb{E}_n F\|_{W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})} \leq \left(\sup_{X \in \mathcal{X}} \|F(X)\|^2_{\mathcal{Y}} + mL^2 \right)^{1/2} \leq \sqrt{m}\|F\|_{C^{0,1}(\mathcal{X},\mathcal{Y})}.$$

As $\lim_{n\to\infty} \mathbb{E}_n F = F$ in $W^{1,2}_{\mu,\boldsymbol{b}}(\mathcal{X};\mathcal{Y})$ by Proposition C.9, the claim follows.

Proof of (ii). Suppose that $\dim(\mathcal{Y}) = \infty$. Notice that in this case we cannot use the same argument as in the proof of (i) for two reasons. First, (D.4) does not give a meaningful bound for $m = \infty$. Second, and more subtly, a similar estimate as in (D.3) does not hold because equality of the operator norm and the Hilbert-Schmidt norm is only true for rank-one-operators. We thus have to argue differently.

To this end, assume that $\boldsymbol{b} \in \ell^2(\mathbb{N})$. We derive bounds for the partial derivatives $\partial_i V_n$ and use square-summability of $\boldsymbol{b}$ to ensure finiteness even if $\mathcal{Y}$ has infinite dimension. First, as $V_n : \mathbb{R}^n \to \mathcal{Y}$ is Lipschitz continuous, we can apply the generalized Rademacher theorem to conclude that V_n is Hadamard differentiable $\mathcal{L}^n$-almost everywhere in $\mathbb{R}^n$, see Aronszajn (1976, Thm. 1 in Chpt. 2 & Rmk. 2 in Chpt. 1). Since $\mathbb{R}^n$ is finite-dimensional, V_n is, in fact, Fréchet differentiable $\mathcal{L}^n$-almost everywhere—recall Appendix A. In particular, V_n belongs to $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n;\mathcal{Y})$. As in the proof of (i), we can thus apply Lemma C.11 to get

$$D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n F)(X)(\eta_i) = \begin{cases} b_i \lambda_i^{-1/2} \partial_i V_n(T_n(X)) & \text{if } 1 \leq i \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

at any point $X \in \mathcal{X}$ such that V_n is differentiable at $T_n(X)$. Note that setting $\boldsymbol{z} = \boldsymbol{e_i}$ (i.e., the ith standard unit vector in $\mathbb{R}^n$) in (D.1) implies that $\|\partial_i V_n(\boldsymbol{x})\|_{\mathcal{Y}} \leq L\sqrt{\lambda_i}$ for $\mathcal{L}^n$-a.e. $\boldsymbol{x} \in \mathbb{R}^n$ and every $1 \leq i \leq n$. Consequently, whenever V_n is differentiable at $T_n(X)$, it follows that

$$\begin{aligned} \left\| D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n F)(X) \right\|^2_{HS(\mathcal{X}_{\boldsymbol{b}},\mathcal{Y})} &= \sum_{i=1}^{\infty} \left\| D_{\mathcal{X}_{\boldsymbol{b}}}(\mathbb{E}_n F)(X)(\eta_i) \right\|^2_{\mathcal{Y}} \\ &= \sum_{i=1}^{n} b_i^2 \lambda_i^{-1} \|\partial_i V_n(X)\|^2_{\mathcal{Y}} \leq L^2 \|\boldsymbol{b}\|^2_{\ell^2(\mathbb{N})}. \end{aligned}$$

We can now proceed as in the proof of (i) to conclude the proof. □