



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. Operator approximations 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 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. Limitations and future work are discussed in Section 1.3. Basic notation is recalled in Section 1.4.

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), general neural operators (Kovachki et al., 2023), and PCANet (Bhattacharya et al., 2021), have been proposed and their efficiency has been demonstrated in various practical applications. We refer to the recent reviews by Kovachki et al. (2024b) and Boullé and Townsend (2024) and references therein. However, the empirical success of these neural operators has to a large extent not yet been supported 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 a 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., 2025; Lanthaler, 2023). Albeit 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, Ch. 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 Düng, 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 (Düng and Griebel, 2016).

1.1.2. Analysis of Lipschitz operators

However, many operators fail to be holomorphic. Among 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 generically do not exhibit any regularity beyond Lipschitz continuity. We refer to Schwab et al. (2026) and Gwinner et al. (2021, Ch. 6) and references therein for further details.

As a specific example, consider the following obstacle problem on the unit square $\Omega = (0, 1)^2$: Given an obstacle $v \in H_0^1(\Omega)$ and an external force $f \in L^2(\Omega)$, find a function $u \in H_0^1(\Omega)$ with minimal energy $\frac{1}{2} \int_{\Omega} \|\nabla u\|^2 d\mathbf{x} - \int_{\Omega} f u d\mathbf{x}$ which stays above v . This problem asks for the resulting shape of an elastic membrane u that is spanned over an obstacle v under an external load f and vanishing Dirichlet boundary conditions (Rodrigues, 1987, Ch. 1). The associated obstacle-to-solution operator

$$F : H_0^1(\Omega) \rightarrow H_0^1(\Omega),$$

$$v \mapsto F(v) = \operatorname{argmin} \left\{ \frac{1}{2} \int_{\Omega} \|\nabla u\|^2 d\mathbf{x} - \int_{\Omega} f u d\mathbf{x} : u \in H_0^1(\Omega), u \geq v \text{ a.e. in } \Omega \right\} \quad (1.1)$$

maps any obstacle v to the corresponding solution $u = F(v)$ of the obstacle problem. It is known that this operator is Lipschitz continuous but not Fréchet differentiable (Mignot, 1976).

Obstacle-to-solution operators similar to (1.1) appear in practice in, for example, contact problems in elasticity, fluid filtration through a porous medium, and ingot solidification in continuous casting. We refer to Rodrigues (1987) for further details. Other practical applications of obstacle problems include, among many others, ice sheet modeling (Jouvet and Bueler, 2012) as well as option pricing in mathematical finance (Jaillet et al., 1990).

The study of approximating Lipschitz operators has recently attracted increased research attention, as they exhibit fundamentally different properties from those of holomorphic operators in terms of learnability. It was shown in Lanthaler (2023) that bounded (with respect to the uniform norm) Lipschitz (and C^k -Fréchet differentiable) operators cannot be approximated with algebraic parametric complexity using PCA-Net. More specifically, the number of 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 (2026). It can seemingly be overcome by neural operators which use hyper-expressive activation functions or non-standard NN architectures (Schwab et al., 2026). 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 all parameters 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 all NN parameters, in fact, still scales exponentially with the inverse of the approximation error.

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 L^2 -error 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. For example, it is immediate to see that the obstacle-to-solution operator in (1.1) is not bounded.

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 of the (covariance operator of the) underlying Gaussian measure. 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 of 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 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) and 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 let μ be a Gaussian measure on $\mathcal{X}$ with PCA eigenvalues $\boldsymbol{\lambda} = (\lambda_i)_{i=1}^\infty$. We now give an (informal) overview of our three main contributions.

1. In Section 2, we extend standard results from infinite-dimensional analysis and define the (weighted) Gaussian Sobolev space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ by means of a sequence $\mathbf{b} = (b_i)_{i=1}^\infty$ of positive bounded weights. As our first main contribution, we show that this space contains the set $\text{Lip}(\mathcal{X}; \mathcal{Y})$ of (potentially unbounded) Lipschitz operators which map from $\mathcal{X}$ to $\mathcal{Y}$:

Result 1 (Lipschitz operators are Gaussian Sobolev operators (Theorem 2.11)). *If $\mathcal{Y}$ is finite-dimensional, then $\text{Lip}(\mathcal{X}; \mathcal{Y}) \subset W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. If $\mathcal{Y}$ is infinite-dimensional and if $\mathbf{b} \in \ell^2(\mathbb{N})$, then $\text{Lip}(\mathcal{X}; \mathcal{Y}) \subset W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. In both cases, the space $C^{0,1}(\mathcal{X}; \mathcal{Y})$ of bounded Lipschitz operators is continuously embedded in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$.*

This result is known in the unweighted case ($b_i = 1$ for all i) if $\mathcal{Y} = \mathbb{R}$ and can be easily generalized to the vector-valued case $\mathcal{Y} = \mathbb{R}^n$. For $\dim(\mathcal{Y}) = \infty$, the weighted setting provides the correct structure for our analysis: The condition $\mathbf{b} \in \ell^2(\mathbb{N})$ gives control of the degree of (weak) differentiability of the operators in all (infinitely many) coordinate directions.

Result 1 is crucial in our subsequent analysis. Elements in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ are characterized as operators whose polynomial expansion coefficients with respect to the (infinite-dimensional, λ -rescaled) Hermite polynomials $\{H_{\gamma, \lambda}\}_{\gamma \in \Gamma}$, with countable index set Γ , are weighted ℓ^2 -summable. The corresponding weights $\mathbf{u} = (u_\gamma)_{\gamma \in \Gamma}$ are given in terms of the ($\mathbf{b}$ -weighted) PCA eigenvalues $\lambda_{\mathbf{b}, i} = \lambda_i / b_i^2$. As a result, we can study the approximation of a given Lipschitz operator by considering its Hermite polynomial expansions.

2. In Section 3, 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 $s \rightarrow \infty$. 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 $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ by Hermite polynomials $H_{\gamma, \lambda}$ with $\gamma \in S$. Moreover, let $\pi : \mathbb{N} \rightarrow \Gamma$ be a bijection such that $(u_{\pi(i)})_{i=1}^\infty$ is a nonincreasing rearrangement of $\mathbf{u}$. Our second main contribution is the following result:

Result 2 (Curse of parametric complexity (Theorems 3.1, 3.5, 3.6)). *For every $s \in \mathbb{N}$, we have*

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

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

As we will discuss in Section 3.4, 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 4, we adopt a more general view and consider learning Lipschitz and $W_{\mu, \mathbf{b}}^{1,2}$ -operators from finitely many arbitrary linear samples. Using tools from *information-based complexity* (IBC) (see Novak and Woźniakowski (2008)), we define the adaptive m -width $\Theta_m(\mathcal{K})$ of a set $\mathcal{K}$ of operators. It quantifies the smallest 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 mapping $\mathcal{T} : \mathcal{Y}^m \rightarrow L_\mu^2(\mathcal{X}; \mathcal{Y})$. Specifically,

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

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

Result 3 (Characterization of the adaptive m -width (Theorems 3.5, 3.6, 4.4)). *For any number of samples $m \in \mathbb{N}$, we have*

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

where $\mathcal{K}$ is either the set of all Lipschitz or all $W_{\mu, \mathbf{b}}^{1,2}$ -operators of at most unit $W_{\mu, \mathbf{b}}^{1,2}$ -norm in both cases. Again, $u_{\pi(m+1)}$ cannot decay with algebraic order as $m \rightarrow \infty$, regardless of the spectral properties of μ . But its decay can become arbitrarily close to any algebraic rate if the PCA eigenvalues $\lambda_{\mathbf{b}, i}$ 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 smallest worst-case L_μ^2 -approximation error. This holds for general (centered, nondegenerate) Gaussian measures μ . In light of Result 3, notice that Result 2 shows that Hermite polynomial approximation is optimal among all possible (adaptive) sampling and reconstruction operators for approximating Lipschitz and $W_{\mu, \mathbf{b}}^{1,2}$ -operators. This sharpens, in particular, the “adaptivity does not help” principle from IBC (Novak, 1996), in that it shows that in this case adaptivity in the $\mathcal{Y}$ -valued samples does not even help to improve any constants in the error bounds. Finally, we emphasize that Result 3 implies that from an IBC point of view, there is no difference between the space $\text{Lip}(\mathcal{X}; \mathcal{Y})$, equipped with the $W_{\mu, \mathbf{b}}^{1,2}$ -norm, and the whole space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$, 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, it is of key interest to design algorithms with (near-)optimal sample complexity which achieve an optimal approximation error with the minimum number 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) and Franco and Brugiapaglia (2025), or with techniques from statistical learning theory (see, e.g., Schmidt-Hieber (2020) and 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. Extending and corroborating previous studies (e.g., Kovachki et al. (2024a); Lanthaler and Stuart (2026)), 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.

1.4. 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 assume that $\dim(\mathcal{X}) = \infty$, unless stated otherwise. 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 lowercase letters at times. We equip $\mathcal{X}$ with a centered, nondegenerate Gaussian measure μ with covariance operator $\Sigma := \int_{\mathcal{X}} X \otimes X d\mu(X)$. We recall that $\Sigma : \mathcal{X} \rightarrow \mathcal{X}$ is positive definite, self-adjoint, and trace class (Da Prato, 2006, Prop. 1.8). As such, there exists an orthonormal eigenbasis (PCA basis) $\{\phi_i\}_{i=1}^{\infty}$ of $\mathcal{X}$ and a sequence of corresponding PCA eigenvalues $\lambda = (\lambda_i)_{i=1}^{\infty}$, which satisfy $\sum_{i=1}^{\infty} \lambda_i < \infty$.

2. Gaussian Sobolev and Lipschitz operators

The results in this section are the basis for the approximation theoretical analysis carried out in the remainder of this paper. We state the definition of the weighted Gaussian Sobolev space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ and its characterization as a weighted ℓ^2 -space in Section 2.2. We recall necessary notions in Section 2.1 and provide technical details in Appendix C.2. In Section 2.3, we prove that all Lipschitz operators $\mathcal{X} \rightarrow \mathcal{Y}$ lie in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$.

2.1. Preliminaries

The space $L_{\mu}^2(\mathcal{X}; \mathcal{Y})$ denotes the Lebesgue-Bochner space of (equivalence classes of) strongly measurable operators $F : \mathcal{X} \rightarrow \mathcal{Y}$ with finite Bochner norm

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

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

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

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

The family $\{H_n\}_{n=0}^{\infty}$ is an orthonormal basis of $L_{\mu_1}^2(\mathbb{R})$ with $\mu_1 = \mathcal{N}(0, 1)$ (Da Prato, 2006, Prop. 9.4). The higher-dimensional Hermite polynomials are defined as products of the one-dimensional ones. To this end, we introduce the set of all sequences of nonnegative integers with finite support,

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

with $\text{supp}(\gamma) := \{i \in \mathbb{N} : \gamma_i \neq 0\}$. It is easy to see that Γ is countable. For $\gamma \in \Gamma$, $d \in \mathbb{N}$, we set

$$H_{\gamma, d} : \mathbb{R}^d \rightarrow \mathbb{R}, \quad H_{\gamma, d}(\mathbf{x}) := \prod_{i=1}^d H_{\gamma_i}(x_i). \quad (2.1)$$

We define Hermite polynomials on the infinite-dimensional space $\mathcal{X}$ by means of the PCA basis $\{\phi_i\}_{i=1}^\infty$ and the PCA eigenvalues $\boldsymbol{\lambda} = (\lambda_i)_{i=1}^\infty$:

$$H_{\gamma, \boldsymbol{\lambda}} : \mathcal{X} \rightarrow \mathbb{R}, \quad H_{\gamma, \boldsymbol{\lambda}}(X) := \prod_{i=1}^\infty H_{\gamma_i} \left(\frac{\langle X, \phi_i \rangle_{\mathcal{X}}}{\sqrt{\lambda_i}} \right). \quad (2.2)$$

The family $\{H_{\gamma, \boldsymbol{\lambda}}\}_{\gamma \in \Gamma}$ is an orthonormal basis of $L_\mu^2(\mathcal{X})$ (Da Prato, 2006, Thm. 9.7). Recalling that $L_\mu^2(\mathcal{X}; \mathcal{Y}) = L_\mu^2(\mathcal{X}) \otimes \mathcal{Y}$ with Hilbertian tensor product, any $F \in L_\mu^2(\mathcal{X}; \mathcal{Y})$ can thus be written as an unconditionally $L_\mu^2(\mathcal{X}; \mathcal{Y})$ -convergent expansion in Hermite polynomials, also called the *Wiener-Hermite Polynomial Chaos (PC) expansion*,

$$F = \sum_{\gamma \in \Gamma} Y_\gamma H_{\gamma, \boldsymbol{\lambda}} \quad \text{with} \quad Y_\gamma := \int_{\mathcal{X}} F(X) H_{\gamma, \boldsymbol{\lambda}}(X) d\mu(X) \in \mathcal{Y}. \quad (2.3)$$

The Y_γ are called the *Wiener-Hermite PC coefficients* of F . Moreover, *Parseval's identity* holds, that is,

$$\|F\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})}^2 = \sum_{\gamma \in \Gamma} \|Y_\gamma\|_{\mathcal{Y}}^2. \quad (2.4)$$

2.2. The space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$

The definition of the space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ is based on the Fréchet differential operator $D_{\mathcal{X}_{\mathbf{b}}}$ along a particular weighted space $\mathcal{X}_{\mathbf{b}}$ (see Appendix A for background on notions of differentiability). We commence with the definition of the latter. Let $\mathbf{b} = (b_i)_{i=1}^\infty \in \ell^\infty(\mathbb{N})$, $\mathbf{b} > \mathbf{0}$ be a sequence of positive bounded weights. Here and throughout the text, $\mathbf{0}$ denotes the constant zero sequence and inequalities between sequences are understood componentwise. By means of the PCA basis $\{\phi_i\}_{i=1}^\infty$ of $\mathcal{X}$ we define

$$\mathcal{X}_{\mathbf{b}} := \left\{ X \in \mathcal{X} : \sum_{i=1}^\infty b_i^{-2} |\langle X, \phi_i \rangle_{\mathcal{X}}|^2 < \infty \right\}. \quad (2.5)$$

Note that $\mathcal{X}_{\mathbf{b}}$ is a Hilbert subspace of $\mathcal{X}$ with inner product and induced norm given by, respectively,

$$\langle X, Z \rangle_{\mathcal{X}_{\mathbf{b}}} := \sum_{i=1}^\infty b_i^{-2} \langle X, \phi_i \rangle_{\mathcal{X}} \langle Z, \phi_i \rangle_{\mathcal{X}} \quad \text{and} \quad \|X\|_{\mathcal{X}_{\mathbf{b}}} := \sqrt{\langle X, X \rangle_{\mathcal{X}_{\mathbf{b}}}}$$

for all $X, Z \in \mathcal{X}_{\mathbf{b}}$. It is easy to see that $\mathcal{X}_{\mathbf{b}}$ has an orthonormal basis $\{\eta_i\}_{i=1}^\infty$, given by

$$\eta_i := b_i \phi_i, \quad i \in \mathbb{N}. \quad (2.6)$$

The role of $\mathcal{X}_{\mathbf{b}}$ will become clear in light of Theorem 2.11 below. See also Remark 2.4. It allows us to control the L_μ^2 -norm of the weak derivatives of a Sobolev operator $\mathcal{X} \rightarrow \mathcal{Y}$ in the case of infinite-dimensional output spaces $\mathcal{Y}$.

Example 2.1. If $\mathbf{b} = \mathbf{1}$, we recover the full space $\mathcal{X}_1 = \mathcal{X}$. If $\mathbf{b} = \sqrt{\lambda} := (\sqrt{\lambda_i})_{i=1}^\infty$, we obtain the Cameron-Martin space $\mathcal{X}_{\sqrt{\lambda}} = \mathcal{H}$ of μ , see (C.1). We remark, however, that the choice $\mathbf{b} = \sqrt{\lambda}$ is excluded by the assumptions we put on $\mathbf{b}$, see Assumption 2.7 and Example 2.8. While the choice $\mathbf{b} = \sqrt{\lambda}$ is not needed for our analysis, the connection to the Cameron-Martin space of μ is nevertheless essential as it allows us to use results from infinite-dimensional analysis. See Appendix C for details.

Let us now turn to the definition of the space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. We start with defining $D_{\mathcal{X}_b}$ as an operator which maps from the set $\mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$ of boundedly differentiable cylindrical operators to $L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$. Here and throughout the text, $\text{HS}(\mathcal{X}_b; \mathcal{Y})$ denotes the space of all Hilbert-Schmidt operators from $\mathcal{X}_b$ to $\mathcal{Y}$, and we write $\text{HS}(\mathcal{X}_b) := \text{HS}(\mathcal{X}_b; \mathbb{R})$. The theory of the Cameron-Martin space $\mathcal{H}$ of μ allows us to show that $D_{\mathcal{X}_b}$ is closable in $L_\mu^2(\mathcal{X}; \mathcal{Y})$. Details are provided in Appendix C. See, in particular, Proposition C.5. Relevant background on the closability and closure of operators is recalled in Appendix B.

Definition 2.2 (The Sobolev space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$). We define the space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ as the domain of the closure of the operator $D_{\mathcal{X}_b} : \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y}) \rightarrow L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ (still denoted by $D_{\mathcal{X}_b}$).

The space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ is a Hilbert space with norm

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

which is induced by the inner product

$$\langle F, G \rangle_{W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})} := \int_{\mathcal{X}} \langle F(X), G(X) \rangle_{\mathcal{Y}} d\mu(X) + \int_{\mathcal{X}} \langle D_{\mathcal{X}_b} F(X), D_{\mathcal{X}_b} G(X) \rangle_{\text{HS}(\mathcal{X}_b; \mathcal{Y})} d\mu(X).$$

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

Remark 2.3. Defining Gaussian Sobolev spaces as the domain of the closure of a suitable differential operator is standard. See, e.g., Chojnowska-Michalik and Goldys (2001); Da Prato (2006, 2014); Lunardi et al. (2015/2016). For an equivalent definition via the completion of $\mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$ under a suitable Sobolev norm, see Bogachev (1998, Ch. 5.2).

Remark 2.4. Weighted Gaussian Sobolev spaces with $\mathcal{Y} = \mathbb{R}$ 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). Recently, Luo et al. (2023) defined weighted Gaussian Sobolev spaces of functionals on $\ell^r(\mathbb{N})$, $r \geq 1$, by weighting the partial derivatives with weights $\mathbf{b} = (b_i)_{i=1}^\infty \in \ell^\infty(\mathbb{N})$. They remarked that their construction unifies various definitions of Gaussian Sobolev spaces via different choices of $\mathbf{b}$. Our definition of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X})$ is equivalent to this construction in the case $r = 2$. However, our approach via the differential operator along the space $\mathcal{X}_b$ highlights the role of the latter as the underlying differential structure of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X})$. For $\mathbf{b} = \sqrt{\lambda}$, we obtain the same space as defined in Bogachev

(1998); *Lunardi et al. (2015/2016)*; *Da Prato (2014)*, which is commonly used in Malliavin calculus. For $\mathbf{b} = \mathbf{1}$, we obtain a space as defined in *Da Prato (2006)*, which is strictly smaller.

Remark 2.5. In the case of finite-dimensional domains $\mathcal{X} = \mathbb{R}^n$, $n \in \mathbb{N}$, we have $\mathcal{X}_{\mathbf{b}} = \mathbb{R}^n$ for any weights $\mathbf{b} = (b_i)_{i=1}^n > \mathbf{0}$. If additionally $\mathcal{Y} = \mathbb{R}$, there is a strong relation between Gaussian and classical Sobolev spaces. In fact,

$$W_{\mu, \mathbf{b}}^{1,2}(\mathbb{R}^n) = \left\{ f \in W_{\text{loc}}^{1,2}(\mathbb{R}^n) : f \in L_{\mu}^2(\mathbb{R}^n), \nabla f \in L_{\mu}^2(\mathbb{R}^n; \mathbb{R}^n) \right\},$$

and the Gaussian and classical weak derivatives coincide (*Bogachev, 1998, Prop. 1.5.2*).

Next, we use the Wiener-Hermite PC expansion (2.3) of $L_{\mu}^2(\mathcal{X}; \mathcal{Y})$ -operators to characterize the space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ by a weighted ℓ^2 -space. More specifically, Gaussian Sobolev regularity of an operator translates to weighted summability of its Wiener-Hermite PC coefficients and vice versa. We start with some notation. Define the ***b-weighted PCA eigenvalues***

$$\lambda_{\mathbf{b}} = (\lambda_{\mathbf{b}, i})_{i=1}^{\infty}, \quad \lambda_{\mathbf{b}, i} := \lambda_i / b_i^2, \quad (2.7)$$

and the weights

$$\mathbf{u} = (u_{\gamma})_{\gamma \in \Gamma}, \quad u_{\gamma} = u_{\gamma}(\lambda_{\mathbf{b}}) := \left(1 + \sum_{i=1}^{\infty} \frac{\gamma_i}{\lambda_{\mathbf{b}, i}} \right)^{-1/2}. \quad (2.8)$$

We denote by $\ell_{\mathbf{u}}^2(\Gamma; \mathcal{Y})$ the space of all $\mathcal{Y}$ -valued sequences $\mathbf{Y} = (Y_{\gamma})_{\gamma \in \Gamma}$ whose norm

$$\|\mathbf{Y}\|_{\ell_{\mathbf{u}}^2(\Gamma; \mathcal{Y})} := \left(\sum_{\gamma \in \Gamma} u_{\gamma}^{-2} \|Y_{\gamma}\|_{\mathcal{Y}}^2 \right)^{1/2}$$

is finite. The following result is an immediate consequence of Proposition C.8.

Theorem 2.6 (ℓ^2 -characterization of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$). *The mapping*

$$\ell_{\mathbf{u}}^2(\Gamma; \mathcal{Y}) \rightarrow W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y}), \quad \mathbf{Y} = (Y_{\gamma})_{\gamma \in \Gamma} \mapsto \sum_{\gamma \in \Gamma} Y_{\gamma} H_{\gamma, \lambda}$$

is an isometric isomorphism. In particular, using the representation (2.3), we have

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

Note that by setting $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}$, and $\lambda_1 = \dots = \lambda_n = 1$, we immediately recover the classical result for the isotropic Gaussian Sobolev space (*López, 2018*), namely,

$$\|F\|_{W_{\mu, \mathbf{1}}^{1,2}(\mathbb{R}^n)}^2 = \sum_{\gamma \in \mathbb{N}_0^n} \left(1 + \sum_{i=1}^n \gamma_i \right) \left| \int_{\mathbb{R}^n} F H_{\gamma, \mathbf{1}} d\mu \right|^2.$$

Assumption 2.7 (Properties of $\mathbf{b}$). *The sequence $\mathbf{b} = (b_i)_{i=1}^\infty \in \ell^\infty(\mathbb{N})$, $\mathbf{b} > \mathbf{0}$, is a sequence of positive bounded real numbers such that the sequence of weighted PCA eigenvalues $\lambda_{\mathbf{b}} = (\lambda_{\mathbf{b},i})_{i=1}^\infty$, defined in (2.7), is nonincreasing and satisfies $\lim_{i \rightarrow \infty} \lambda_{\mathbf{b},i} = 0$. If $\dim(\mathcal{Y}) = \infty$, we assume in addition that $\mathbf{b} \in \ell^2(\mathbb{N})$.*

The weights u_γ in (2.8) are key to our analysis. By Assumption 2.7, every superlevel set $\{\gamma \in \Gamma : u_\gamma \geq t\}$, $t > 0$, is finite. We can thus order the weights u_γ in a nonincreasing way, that is, there exists a bijection $\pi : \mathbb{N} \rightarrow \Gamma$ that gives a *nonincreasing rearrangement* of $\mathbf{u}$ so that

$$u_{\pi(1)} \geq u_{\pi(2)} \geq \cdots > 0. \quad (2.9)$$

The mapping π is unique up to permutations of weights of the same value. The additional requirement in Assumption 2.7 that $\mathbf{b} \in \ell^2(\mathbb{N})$ in the case $\dim(\mathcal{Y}) = \infty$ is necessary in our analysis to show that the set of Lipschitz operators is a subset of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$, see Theorem 2.11.

Example 2.8. If $\mathcal{Y}$ is finite-dimensional and the λ_i are assumed to be nonincreasing, we can choose $\mathbf{b} = \mathbf{1}$, so that $\lambda_{\mathbf{b}} = \lambda$. This is not possible if $\mathcal{Y}$ is infinite-dimensional, as $\mathbf{1} \notin \ell^2(\mathbb{N})$. Note that $\mathbf{b} = \sqrt{\lambda}$ is not a valid choice either, as it yields $\lambda_{\mathbf{b}} = \mathbf{1}$ and thus violates the limit condition in Assumption 2.7. However, since $\lambda \in \ell^1(\mathbb{N})$, a valid choice in any case is given, for example, by $b_i := \lambda_i^{1/2} / (\sum_{j=i}^\infty \lambda_j)^{1/4}$, $i \in \mathbb{N}$. In this case, $\mathbf{b} > \mathbf{0}$ and $\lambda_{\mathbf{b},i} = (\sum_{j=i}^\infty \lambda_j)^{1/2}$ so that $\lim_{i \rightarrow \infty} \lambda_{\mathbf{b},i} = 0$. It is a short exercise to show that $\mathbf{b} \in \ell^2(\mathbb{N})$. For this, set $s_i := \sum_{j=i}^\infty \lambda_j$ and compute

$$b_i^2 = \frac{s_i - s_{i+1}}{\sqrt{s_i}} = \frac{(\sqrt{s_i} + \sqrt{s_{i+1}})(\sqrt{s_i} - \sqrt{s_{i+1}})}{\sqrt{s_i}} \leq \frac{2\sqrt{s_i}(\sqrt{s_i} - \sqrt{s_{i+1}})}{\sqrt{s_i}} = 2(\sqrt{s_i} - \sqrt{s_{i+1}}).$$

Consequently, $\sum_{i=1}^\infty b_i^2 \leq 2\sqrt{s_1} < \infty$.

Remark 2.9. *The condition $\lim_{i \rightarrow \infty} \lambda_{\mathbf{b},i} = 0$ is equivalent to the compact embedding of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X})$ in $L_\mu^2(\mathcal{X})$ (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).*

2.3. Lipschitz operators

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

Definition 2.10 (Lipschitz operators). An operator $F : \mathcal{X} \rightarrow \mathcal{Y}$ 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

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

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

of all Lipschitz operators $F \in \text{Lip}(\mathcal{X}; \mathcal{Y})$ with finite norm

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

Notice that $C^{0,1}(\mathcal{X}; \mathcal{Y})$ is a strict subset of $\text{Lip}(\mathcal{X}; \mathcal{Y})$ as operators in $\text{Lip}(\mathcal{X}; \mathcal{Y})$ are not necessarily bounded.

The next result is the main result of this section. It shows that Lipschitz operators are Gaussian Sobolev operators. This motivates Gaussian Sobolev spaces as a natural setting for the study of Lipschitz operators. Recall that on finite-dimensional domains, the classical Rademacher theorem states that Lipschitz functions $\mathbb{R}^n \rightarrow \mathbb{R}$ are differentiable $\mathcal{L}^n$ -almost everywhere (here and throughout, $\mathcal{L}^n$ denotes the Lebesgue measure). This can be generalized to Banach-valued operators: Any Lipschitz operator $F : \mathbb{R}^n \rightarrow \mathcal{Y}$ is $\mathcal{L}^n$ -almost everywhere Fréchet differentiable if $\mathcal{Y}$ is a Banach space which has the Radon-Nikodym property (which is satisfied by any Hilbert space). See [Benyamini and Lindenstrauss \(1999, Prop. 6.41\)](#). Theorem 2.11 can be interpreted as a further generalization of Rademacher's theorem to operators on infinite-dimensional domains with the Lebesgue measure replaced by a Gaussian measure. In fact, the proof is based on a reduction of the infinite-dimensional setting to the finite-dimensional one by means of a suitable approximating sequence and a subsequent application of the classical Rademacher theorem. We present a sketch of the proof below and provide full details in Appendix D.

Theorem 2.11 (Lipschitz operators are Gaussian Sobolev operators). *Let $\mathbf{b} = (b_i)_{i=1}^\infty \in \ell^\infty(\mathbb{N})$, $\mathbf{b} > \mathbf{0}$, be a sequence of positive bounded numbers.*

(i) *If $\mathcal{Y}$ is finite-dimensional, then $\text{Lip}(\mathcal{X}; \mathcal{Y}) \subset W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ and*

$$\|D\mathcal{X}_{\mathbf{b}}F\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_{\mathbf{b}}; \mathcal{Y}))} \leq \sqrt{\dim(\mathcal{Y})} \cdot \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})} \cdot [F]_{\text{Lip}}, \quad \forall F \in \text{Lip}(\mathcal{X}; \mathcal{Y}).$$

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

$$\|F\|_{W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})} \leq \max \left\{ 1, \sqrt{\dim(\mathcal{Y})} \cdot \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})} \right\} \cdot \|F\|_{C^{0,1}(\mathcal{X}; \mathcal{Y})}, \quad \forall F \in C^{0,1}(\mathcal{X}; \mathcal{Y}).$$

(ii) *If $\mathcal{Y}$ is infinite-dimensional and if $\mathbf{b} \in \ell^2(\mathbb{N})$, then $\text{Lip}(\mathcal{X}; \mathcal{Y}) \subset W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ and*

$$\|D\mathcal{X}_{\mathbf{b}}F\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_{\mathbf{b}}; \mathcal{Y}))} \leq \|\mathbf{b}\|_{\ell^2(\mathbb{N})} \cdot [F]_{\text{Lip}}, \quad \forall F \in \text{Lip}(\mathcal{X}; \mathcal{Y}).$$

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

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

Proof (Sketch) Let $F \in \text{Lip}(\mathcal{X}; \mathcal{Y})$. In order to show that F lies in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$, it suffices to find a sequence $(F_n)_{n=1}^\infty$ of operators which converges to F in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ with the F_n uniformly

bounded in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ (see Lemma C.7). To this end, we construct a specific sequence of operators of the form $F_n = V_n \circ T_n$ with $V_n : \mathbb{R}^n \rightarrow \mathcal{Y}$ and $T_n : \mathcal{X} \rightarrow \mathbb{R}^n$ which converges to F in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ and, in fact, in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. 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. It remains to show that F_n is differentiable μ -almost everywhere and that there exists a constant $C > 0$ such that

$$\|D_{\mathcal{X}_b} F_n(X)\|_{\text{HS}(\mathcal{X}_b; \mathcal{Y})} \leq C \quad (2.10)$$

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 $\{\psi_j\}_{j=1}^m$ of $\mathcal{Y}$ and consider the (Lipschitz continuous) coordinate functions $V_n^{(j)} := \langle V_n, \psi_j \rangle_{\mathcal{Y}} : \mathbb{R}^n \rightarrow \mathbb{R}$. Unwinding definitions, a straightforward calculation then leads to (2.10) for any bounded weight sequence $\mathbf{b} > \mathbf{0}$. The resulting constant C is given by $\sqrt{\dim(\mathcal{Y})} \cdot \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})} \cdot [F]_{\text{Lip}}$. The same argument does not work in case (ii) where $\mathcal{Y}$ has infinite dimension. At this point, we additionally require that $\mathbf{b} \in \ell^2(\mathbb{N})$. A slight modification of the argument in (i) then yields (2.10) with C given by $\|\mathbf{b}\|_{\ell^2(\mathbb{N})} \cdot [F]_{\text{Lip}}$. The continuous embedding of $C^{0,1}(\mathcal{X}; \mathcal{Y})$ in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ follows in both cases from (2.10) and the fact that $F_n \rightarrow F$ in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. $\square$

In light of Theorem 2.11, note that Assumption 2.7 implies that $\text{Lip}(\mathcal{X}; \mathcal{Y}) \subset W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$, regardless of whether $\mathcal{Y}$ is finite- or infinite-dimensional.

Remark 2.12. For functionals (i.e., $\mathcal{Y} = \mathbb{R}$), 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 $\text{Lip}(\mathcal{X}) \subset W_{\mu, \mathbf{1}}^{1,2}(\mathcal{X})$ and $\text{Lip}(\mathcal{X}) \subset W_{\mu, \sqrt{\lambda}}^{1,2}(\mathcal{X})$, respectively.

Remark 2.13. One can define Gaussian Sobolev spaces $W_{\mu, \mathbf{b}}^{1,p}(\mathcal{X}; \mathcal{Y})$ for any order $1 \leq p < \infty$ and a proof analogous to the one of Theorem 2.11 shows that they contain $\text{Lip}(\mathcal{X}; \mathcal{Y})$ 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 $p = 2$ is there a simple characterization of $W_{\mu, \mathbf{b}}^{1,p}(\mathcal{X}; \mathcal{Y})$ in terms of a weighted ℓ^2 -space as given by Theorem 2.6.

3. Polynomial s -term approximation

The ℓ^2 -characterization of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ via Wiener-Hermite PC expansions (see Theorem 2.6) motivates studying polynomial s -term approximations of $W_{\mu, \mathbf{b}}^{1,2}$ -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

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

and the corresponding orthogonal L_μ^2 -projection

$$(\cdot)_S : L_\mu^2(\mathcal{X}; \mathcal{Y}) \rightarrow \mathcal{P}_{S; \mathcal{Y}}, \quad F \mapsto F_S := \sum_{\gamma \in S} \left(\int_{\mathcal{X}} F H_{\gamma, \lambda} d\mu \right) H_{\gamma, \lambda}. \quad (3.1)$$

We further define the *Sobolev unit ball* and the *Sobolev unit Lipschitz ball* as, respectively,

$$\begin{aligned} B_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y}) &:= \left\{ F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y}) : \|F\|_{W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})} \leq 1 \right\}, \\ B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y}) &:= \left\{ F \in \text{Lip}(\mathcal{X}; \mathcal{Y}) : \|F\|_{W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})} \leq 1 \right\}. \end{aligned} \quad (3.2)$$

As usual, when $\mathcal{Y} = \mathbb{R}$, we drop it from the notation.

Next, let $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ and let $S \subset \Gamma$ be finite with $|S| \leq s$. Setting $Y_\gamma := \int_{\mathcal{X}} F H_{\gamma, \lambda} d\mu$, it follows from Parseval's identity (2.4) and Theorem 2.6 that

$$\|F - F_S\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})}^2 = \sum_{\gamma \in \Gamma \setminus S} \|Y_\gamma\|_{\mathcal{Y}}^2 \leq \left(\max_{\gamma \in \Gamma \setminus S} u_\gamma^2 \right) \sum_{\gamma \in \Gamma} u_\gamma^{-2} \|Y_\gamma\|_{\mathcal{Y}}^2 = \left(\max_{\gamma \in \Gamma \setminus S} u_\gamma^2 \right) \|F\|_{W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})}^2.$$

Let us recall from (2.9) the mapping $\pi : \mathbb{N} \rightarrow \Gamma$ which gives a nonincreasing rearrangement of $\mathbf{u} = (u_\gamma)_{\gamma \in \Gamma}$. For $s \in \mathbb{N}$, we set $S = \pi([s]) = \{\pi(\mathbf{1}), \dots, \pi(\mathbf{s})\}$ and conclude for $\mathcal{K} \in \{B_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y}), B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y})\}$ that

$$\inf_{S \subset \Gamma, |S| \leq s} \sup_{F \in \mathcal{K}} \|F - F_S\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})} \leq \sup_{F \in \mathcal{K}} \|F - F_{\pi([s])}\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})} \leq \max_{\gamma \in \Gamma \setminus \pi([s])} u_\gamma = u_{\pi(\mathbf{s}+1)}. \quad (3.3)$$

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 4.4 in the special case $\mathcal{V} = L_\mu^2(\mathcal{X}; \mathcal{Y})$.

Theorem 3.1 (Best polynomial s -term error). *For $\mathcal{K} \in \{B_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y}), B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y})\}$ and every $s \in \mathbb{N}$, we have*

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

Motivated by Theorem 3.1, we study in the rest of this section the decay of $u_{\pi(\mathbf{s}+1)}$ as $s \rightarrow \infty$. The proofs are based on the relation of the set $\pi([s])$ to an anisotropic total degree index set, which we discuss in Section 3.1. Subsequently, we prove lower and upper bounds for $u_{\pi(\mathbf{s}+1)}$ in Sections 3.2 and 3.3.

3.1. 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 3.2 (Anisotropic TD index set). For $d \in \mathbb{N}$ and $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$, $\mathbf{a} > \mathbf{0}$, we define the anisotropic TD index set in d dimensions with weight $\mathbf{a}$ by

$$\Lambda_{d,\mathbf{a}}^{\text{TD}} := \left\{ \nu \in \mathbb{N}_0^d : \sum_{i=1}^d a_i \nu_i \leq 1 \right\}.$$

Lemma 3.3 (Lower and upper size bounds for anisotropic TD index sets). *Let $d \in \mathbb{N}$ and $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$ with $0 < a_1 \leq \dots \leq a_d$. We have*

$$\prod_{i=1}^d \frac{1}{a_i i} \leq |\Lambda_{d,\mathbf{a}}^{\text{TD}}| \leq \prod_{i=1}^d \left(\frac{1}{a_i i} + 1 \right). \quad (3.4)$$

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. $\square$

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

$$S(\varepsilon) := \left\{ \gamma \in \Gamma : u_\gamma^{-2} \leq 1 + \frac{1}{\varepsilon^2} \right\} = \left\{ \gamma \in \Gamma : \sum_{i=1}^{\infty} \frac{\gamma_i}{\lambda_{\mathbf{b},i}} \leq \frac{1}{\varepsilon^2} \right\} \quad (3.5)$$

as well as the quantity

$$d(\varepsilon) := \min \{ l \in \mathbb{N}_0 : \lambda_{\mathbf{b},l+1} < \varepsilon^2 \} \in \mathbb{N}_0. \quad (3.6)$$

By definition, we have $S(\varepsilon) = \pi(|S(\varepsilon)|)$. If $d(\varepsilon) > 0$, then $S(\varepsilon)$ is isomorphic to an anisotropic TD index set via the mapping

$$S(\varepsilon) \xrightarrow{\cong} \Lambda_{d(\varepsilon),\mathbf{a}}^{\text{TD}}, \quad \gamma \mapsto (\gamma_1, \dots, \gamma_{d(\varepsilon)}), \quad (3.7)$$

with weights $a_i := \varepsilon^2 / \lambda_{\mathbf{b},i}$, $i \in [d(\varepsilon)]$. Observe that by the monotonicity condition in Assumption 2.7 we have $0 < a_1 \leq a_2 \leq \dots \leq a_{d(\varepsilon)}$.

Remark 3.4 (Effective dimension). *The number $d(\varepsilon)$, defined in (3.6), can be interpreted as the effective dimension of the approximation problem in the following sense: For $i > d(\varepsilon)$, we have $\lambda_i \leq \lambda_{\mathbf{b},i} \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})}^2 < \varepsilon^2 \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})}^2$. That is, the variance of μ in the i th coordinate direction (with respect to the PCA basis $\{\phi_i\}_{i=1}^\infty$) essentially vanishes for very small values of ε . Consequently, $\mathcal{N}(0, \lambda_i) \approx \delta_0$ for $i > d(\varepsilon)$, where 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 $\bigotimes_{i=1}^{d(\varepsilon)} \mathcal{N}(0, \lambda_i)$.*

3.2. Lower bound

We now prove the second main result of this section which is a lower bound for $u_{\pi(s+1)}$. It states that, regardless of the (unweighted) PCA eigenvalues $\boldsymbol{\lambda}$ and choice of $\mathbf{b}$, one cannot achieve an algebraic decay of $u_{\pi(s+1)}$ as $s \rightarrow \infty$. The proof is based on relation (3.7) and the lower size bound in Lemma 3.3.

Theorem 3.5 (Impossibility of algebraic decay of $u_{\pi(s+1)}$). *For any $p \in \mathbb{N}$, there exists $\bar{s} \in \mathbb{N}$, depending on $\lambda_{\mathbf{b},1}, \dots, \lambda_{\mathbf{b},p}$, and p , such that*

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

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

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

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

We first 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 (3.5) and (3.6), 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 $\Lambda_{n,\mathbf{a}}^{\text{TD}}$ with weight $\mathbf{a} = (a_1, \dots, a_n)$, $a_i := \varepsilon^2 / \lambda_{\mathbf{b},i}$, via the mapping $\gamma \mapsto (\gamma_1, \dots, \gamma_n)$. We now combine these observations with the lower size bound in (3.4) to conclude

$$|S(\varepsilon)| = |S(\varepsilon, d(\varepsilon))| \geq |S(\varepsilon, d')| \geq \prod_{i=1}^{d'} \frac{\lambda_{\mathbf{b},i}}{\varepsilon^2 i}, \quad \forall 1 \leq d' \leq d(\varepsilon). \quad (3.8)$$

Analogously to the definition of $d(\varepsilon)$ in (3.6), we set

$$\tilde{d}(\varepsilon) := \min \left\{ l \in \mathbb{N}_0 : \frac{\lambda_{\mathbf{b},l+1}}{l+1} < \varepsilon^2 \right\} \in \mathbb{N}_0.$$

Note that we have $\tilde{d}(\varepsilon) \leq d(\varepsilon)$ and $\tilde{d}(\varepsilon) \rightarrow \infty$ as $\varepsilon \rightarrow 0$.

Next, let $p \in \mathbb{N}$ be arbitrary. We fix some $0 < \bar{\varepsilon} = \bar{\varepsilon}(\lambda_{\mathbf{b},p}, p) \leq \min\{\sqrt{\lambda_{\mathbf{b},p}/p}, 1\}$. By definition of $\tilde{d}(\varepsilon)$, we have $\tilde{d}(\varepsilon) \geq p$ for every $0 < \varepsilon \leq \bar{\varepsilon}$. Since $\lambda_{\mathbf{b},i}/(\varepsilon^2 i) \geq 1$ for every $1 \leq i \leq \tilde{d}(\varepsilon)$, it follows from (3.8) that

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

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

$$\tilde{\varepsilon}^2 = \tilde{\varepsilon}^2(\lambda_{\mathbf{b},1}, \dots, \lambda_{\mathbf{b},p}, p, s) = \tilde{C}^{1/p} (s+1)^{-1/p}.$$

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

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

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

3.3. Upper bounds

We have now seen that $u_{\pi(s+1)}$ cannot decay algebraically as $s \rightarrow \infty$, regardless of the decay of the $\mathbf{b}$ -weighted PCA eigenvalues $\lambda_{\mathbf{b},i}$. We now study the decay of $u_{\pi(s+1)}$ 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 principle, the proof of the following result can be adapted to any other decay of the $\lambda_{\mathbf{b},i}$ as long as $(\lambda_{\mathbf{b},i})_{i=1}^\infty$ is nonincreasing and $\lim_{i \rightarrow \infty} \lambda_{\mathbf{b},i} = 0$. It is based on relation (3.7) and the upper size bound in Lemma 3.3.

Theorem 3.6 (Typical decays of $u_{\pi(s+1)}$). *Let $\alpha, \beta > 0$.*

- (a) Algebraic spectral decay: *Let $\lambda_{\mathbf{b},i} = i^{-\alpha}$ 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 $s \geq \bar{s}$,*

$$u_{\pi(s+1)} \leq \eta \log(s)^{-\frac{1}{2(1/\alpha + \delta)}}. \quad (3.9)$$

- (b) Exponential spectral decay: *Let $\lambda_{\mathbf{b},i} = e^{-\alpha i^\beta}$ 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 $s \geq \bar{s}$,*

$$u_{\pi(s+1)} \leq e^{-\frac{1}{2}\alpha^{\frac{1}{\beta+1}} \left(\frac{\beta+1}{\beta+\delta}\right)^{\frac{1}{1+1/\beta}} \log(s)^{\frac{1}{1+1/\beta}}}. \quad (3.10)$$

- (c) Double exponential spectral decay: *Let $\lambda_{\mathbf{b},i} = e^{-e^{\alpha i}}$ 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 $s \geq \bar{s}$,*

$$u_{\pi(s+1)} \leq e^{-\eta \log(s)^{\frac{1}{1+\delta}}}. \quad (3.11)$$

Proof All rates can be derived by suitably choosing the parameter ε in the set $S(\varepsilon)$, defined in (3.5). Suppose that $\varepsilon \leq \sqrt{\lambda_{\mathbf{b},1}}$, so that $\lambda_{\mathbf{b},i} \geq \varepsilon^2$ for every $i \in [d(\varepsilon)]$ by definition of $d(\varepsilon)$ (see (3.6)). By the relation (3.7), we may use the upper size bound in (3.4) to compute

$$|S(\varepsilon)| \leq \prod_{i=1}^{d(\varepsilon)} \left(\frac{\lambda_{\mathbf{b},i}}{i\varepsilon^2} + 1 \right) \leq \prod_{i=1}^{d(\varepsilon)} \frac{\lambda_{\mathbf{b},i}}{\varepsilon^2} \left(\frac{1}{i} + 1 \right) \leq (2\varepsilon^{-2})^{d(\varepsilon)} \prod_{i=1}^{d(\varepsilon)} \lambda_{\mathbf{b},i}. \quad (3.12)$$

For brevity, we write in the following $d = d(\varepsilon)$.

Case (a). Using (3.12) as well as the Stirling type estimate $d^d \leq e^d d!$, we obtain

$$|S(\varepsilon)| \leq (2\varepsilon^{-2})^d \prod_{i=1}^d i^{-\alpha} = (2\varepsilon^{-2})^d (d!)^{-\alpha} \leq (2\varepsilon^{-2})^d e^{\alpha d} d^{-\alpha d}. \quad (3.13)$$

Let $0 < \varepsilon \leq 2^{-\alpha/2}$. By definition of d , we have $d \leq \varepsilon^{-2/\alpha} < d+1 \leq 2d$. We take the logarithm on both sides in (3.13) 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 $\alpha + \frac{1}{2} \log(2^{\alpha+1}) \leq \frac{1}{2} \eta^{2(1/\alpha+\delta)} \varepsilon^{-2\delta}$. It follows that

$$\log(|S(\varepsilon)|) \leq \eta^{2(1/\alpha+\delta)} \varepsilon^{-2/\alpha-2\delta}, \quad \forall \varepsilon \leq \bar{\varepsilon}. \quad (3.14)$$

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

$$\varepsilon^2 = \varepsilon(s)^2 = \eta^2 \log(s)^{-\frac{1}{1/\alpha+\delta}}. \quad (3.15)$$

We can now choose $\bar{s} = \bar{s}(\alpha, \delta, \eta) \in \mathbb{N}$ sufficiently large such that the right-hand side in (3.15) is smaller than $\bar{\varepsilon}^2$ for every $s \geq \bar{s}$. Then, (3.14) 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 $\pi(\mathbf{s} + \mathbf{1}) \notin S(\varepsilon)$ and consequently,

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

Case (b). By (3.12), we find

$$|S(\varepsilon)| \leq (2\varepsilon^{-2})^d \prod_{i=1}^d e^{-\alpha i^\beta} = (2\varepsilon^{-2})^d e^{-\alpha \sum_{i=1}^d i^\beta}. \quad (3.16)$$

Let $0 < \varepsilon \leq e^{-\alpha/2}$. By definition of d , we have $d \leq \alpha^{-1/\beta} \log(\varepsilon^{-2})^{1/\beta} < d+1$. We take the logarithm on both sides of (3.16) and compute

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

Next, let $\tilde{\delta} > 0$ and $0 < \eta < 1$ be arbitrary. There exists $0 < \bar{\varepsilon} = \bar{\varepsilon}(\alpha, \beta, \tilde{\delta}, \eta) \leq e^{-\alpha/2}$ such that $\log(2) \leq \tilde{\delta} \log(\varepsilon^{-2})$ 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,

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

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

Case (c). By (3.12), we find that

$$|S(\varepsilon)| \leq (2\varepsilon^{-2})^d \prod_{i=1}^d e^{-e^{\alpha i}} = (2\varepsilon^{-2})^d e^{-\sum_{i=1}^d e^{\alpha i}}. \quad (3.17)$$

Let $0 < \varepsilon \leq e^{-e^\alpha/2}$. Then, by definition of d , we have $d \leq \log(\log(\varepsilon^{-2}))/\alpha < d+1$. We take the logarithm on both sides in (3.17) and compute

$$\log(|S(\varepsilon)|) \leq d \log(2\varepsilon^{-2}) - \sum_{i=1}^d e^{\alpha i} \leq \frac{1}{\alpha} \log(\log(\varepsilon^{-2})) \log(2\varepsilon^{-2}) - \frac{\log(\varepsilon^{-2}) - 1}{e^\alpha - 1} + 1. \quad (3.18)$$

Next, let $\delta, \eta > 0$ be arbitrary. There exists $0 < \bar{\varepsilon} = \bar{\varepsilon}(\alpha, \delta, \eta) \leq e^{-e^\alpha/2}$ such that for every $0 < \varepsilon \leq \bar{\varepsilon}$, the right-hand side in (3.18) 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 in (3.4) show that the bounds in Theorem 3.6 are asymptotically sharp in the limit $s \rightarrow \infty$ —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

$$u_{\pi(s+1)} \gtrsim e^{-\frac{1}{2}\alpha^{\frac{1}{\beta+1}} \left(\frac{\beta+1}{\beta-\delta}\right)^{\frac{1}{1+1/\beta}} \log(s+1)^{\frac{1}{1+1/\beta}}}, \quad \forall s \geq \bar{s}. \quad (3.19)$$

Here and henceforth, we write $x \gtrsim y$ for $x, y \in \mathbb{R}$ if there exists a global constant $C > 0$, independent of any parameters, such that $x \geq Cy$. In case (c), for every $\eta > 0$, there exists $\bar{s} = \bar{s}(\alpha, \eta) \in \mathbb{N}$ such that

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

In case (a) with algebraic spectral decay, the lower size bound (3.4) is too weak and leads to inconclusive results which do not asymptotically match the upper bound (3.9).

3.4. Discussion

Theorem 3.5 shows that $u_{\pi(s+1)}$ decays subalgebraically as $s \rightarrow \infty$. 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 3.1, 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 $s \rightarrow \infty$. This holds regardless of the decay rate of the PCA eigenvalues.*

We have now seen that approximating Lipschitz operators by truncating Wiener-Hermite PC expansions cannot be done efficiently with algebraic convergence rates. On the other hand, Theorem 3.6 shows the connection between the decay of the eigenvalues $\lambda_{\mathbf{b},i}$ in i and the decay of $u_{\pi(s+1)}$ 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 $\lambda_{\mathbf{b},i}$ decay sufficiently fast. Since the lower bound (3.19) asymptotically matches the upper

bound (3.10), we see, however, that to this end, the eigenvalue decay has to be faster than exponential. Equation (3.11) shows that a double exponential decay is sufficient.

In the context of the above-stated curse of parametric 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 3.7 (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*

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

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

Remark 3.8. *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 true PCA eigenvalues, which puts Theorem 3.7 in the setting of the present paper.*

We now argue that the curse of parametric complexity described by Theorems 3.1 and 3.5 is consistent with the curse of parametric PCA-Net complexity described by Theorem 3.7. 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 3.9 (Algebraic s -term convergence for Lipschitz operators). *There exists $\beta > 0$ such that for every $F \in B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y})$, there are constants $c(F) > 0$ and $\bar{s}(F) \in \mathbb{N}$ such that*

$$\|F - F_{\pi([s])}\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})} \leq c(F) s^{-\beta}, \quad \forall s \geq \bar{s}(F). \quad (3.20)$$

Proposition 3.10. *Grant Assumption 3.9. Then there exists $\alpha > 0$ with the following property: For all $F \in C^{0,1}(\mathcal{X}; \mathcal{Y})$, there exist constants $C = C(F, \alpha) > 0$ and $\bar{\epsilon} = \bar{\epsilon}(F, \alpha) > 0$ such that for all $0 < \epsilon \leq \bar{\epsilon}$, there is a PCA-Net $\Psi = \tilde{\mathcal{D}}_\mathcal{Y} \circ \psi \circ \tilde{\mathcal{E}}_\mathcal{X}$ such that*

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

With the notion introduced in Lanthaler (2023), Proposition 3.10 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 3.7, hence implying Assumption 3.9 to be false. The proof of Proposition 3.10 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. It is given in Appendix E.1.

4. Optimal sampling and (adaptive) m -widths

Up to this point, we have studied the best approximation of $W_{\mu, \mathbf{b}}^{1,2}$ - and Lipschitz operators by Hermite polynomials. This is a certain type of what in IBC is referred to as *linear information* (Novak and Woźniakowski, 2008, Ch. 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_γ . We follow in parts ideas from Adcock et al. (2024c), who studied holomorphic operators.

4.1. Adaptive sampling operators

Definition 4.1 (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

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

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

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

$$\mathcal{L}(F) := \left(\int_{\mathcal{X}} F H_{\gamma^{(i)}, \lambda} d\mu \right)_{i=1}^m \in \mathbb{R}^m, \quad \forall F \in L_\mu^2(\mathcal{X}). \quad (4.1)$$

Throughout the text, we denote by $C_b(\mathcal{X}; \mathcal{Y})$ the space of all bounded continuous operators $\mathcal{X} \rightarrow \mathcal{Y}$, equipped with the uniform norm. We write $C_b(\mathcal{X}) := C_b(\mathcal{X}; \mathbb{R})$. If $\mathcal{V} = C_b(\mathcal{X})$ and $X_1, \dots, X_m \in \mathcal{X}$, we can define a pointwise sampling operator, generating (non-adaptive) *standard information* (Novak and Woźniakowski, 2008, Ch. 4.1.1), by

$$\mathcal{L}(F) := (F(X_i))_{i=1}^m \in \mathbb{R}^m, \quad \forall F \in C_b(\mathcal{X}). \quad (4.2)$$

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

Next, we consider the Hilbert-valued case. For any $Y \in \mathcal{Y}$, $\mathbf{v} = (v_i)_{i=1}^m \in \mathbb{R}^m$, and $F \in L_\mu^2(\mathcal{X})$, we write $Y\mathbf{v}$ for the vector $(Yv_i)_{i=1}^m \in \mathcal{Y}^m$ and YF for the mapping $X \mapsto YF(X)$.

Definition 4.2 (Adaptive sampling operator; Hilbert-valued case). Let $\mathcal{V} \subset L_\mu^2(\mathcal{X}; \mathcal{Y})$ be a vector subspace with norm $\|\cdot\|_{\mathcal{V}}$ and consider a mapping

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

where $\mathcal{L}_1 : \mathcal{V} \rightarrow \mathcal{Y}$ is a bounded linear operator and $\mathcal{L}_i : \mathcal{V} \times \mathcal{Y}^{i-1} \rightarrow \mathcal{Y}$ 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 $Y, \tilde{Y} \in \mathcal{Y} \setminus \{0\}$, a normed vector space $\tilde{\mathcal{V}} \subset L_\mu^2(\mathcal{X})$, and a scalar-valued adaptive sampling operator $\tilde{\mathcal{L}} : \tilde{\mathcal{V}} \rightarrow \mathbb{R}^m$ such that, if $YF \in \mathcal{V}$ for some $F \in L_\mu^2(\mathcal{X})$, then $F \in \tilde{\mathcal{V}}$ and $\mathcal{L}(YF) = Y\tilde{\mathcal{L}}(F)$.

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 $\mathcal{V} = C_b(\mathcal{X}; \mathcal{Y})$ (seen as a subspace of $L_\mu^2(\mathcal{X}; \mathcal{Y})$ and equipped with the L^∞ -norm) and define

$$\mathcal{L}(F) := (F(X_i))_{i=1}^m \in \mathcal{Y}^m, \quad \forall F \in \mathcal{V},$$

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

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

where $\tilde{\mathcal{L}} : \tilde{\mathcal{V}} \rightarrow \mathbb{R}^m$ is the adaptive pointwise sampling operator in (4.2). 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 (4.1), is a Hilbert-valued sampling operator in the sense of Definition 4.2 (under a mild condition on $\mathcal{V}$), see Section 4.3.3.

4.2. Adaptive m -widths and main result

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

Definition 4.3 (Adaptive m -width). Let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be a normed vector subspace of $L_\mu^2(\mathcal{X}; \mathcal{Y})$ and let $\mathcal{K} \subset \mathcal{V}$ be a subset. The adaptive m -width of $\mathcal{K}$ in $\mathcal{V}$ is given by

$$\begin{aligned} & \Theta_m(\mathcal{K}; \mathcal{V}, L_\mu^2(\mathcal{X}; \mathcal{Y})) \\ & := \inf \left\{ \sup_{F \in \mathcal{K}} \|F - \mathcal{T}(\mathcal{L}(F))\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})} : \mathcal{L} : \mathcal{V} \rightarrow \mathcal{Y}^m \text{ adaptive, } \mathcal{T} : \mathcal{Y}^m \rightarrow L_\mu^2(\mathcal{X}; \mathcal{Y}) \right\}. \end{aligned} \quad (4.3)$$

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 $\mathcal{L}$. Note that in (4.3) we allow

for any (possibly nonlinear) reconstruction mappings. The choice of $\mathcal{V}$, however, determines which sampling operators are allowed. If $\mathcal{V} = C_b(\mathcal{X}; \mathcal{Y})$, we can use pointwise sampling, whereas, if $\mathcal{V} = L_\mu^2(\mathcal{X}; \mathcal{Y})$, this is not possible.

We remark that the term “adaptive m -width” is not standard terminology in the literature. For better understanding, we put it into the context of other widths in nonlinear approximation. The adaptive m -width is the Hilbert-valued version of the adaptive compressive m -width (Foucart and Rauhut, 2013), which we introduce in Section 4.3.1 where we also discuss its relation to the Gelfand and Kolmogorov width. If we set $\mathcal{Y} = \mathbb{R}$ in (4.3) and assume $\mathcal{L}$ to be a general continuous mapping $\mathcal{K} \rightarrow \mathbb{R}^m$ instead of a (scalar-valued) adaptive sampling operator on $\mathcal{V}$, then we obtain the *continuous nonlinear m -width* as introduced in DeVore et al. (1989). If we assume in addition that $\mathcal{T}$ is continuous (and $\mathcal{K}$ is compact), then we obtain the *nonlinear m -width* (or *manifold width*) in DeVore (1998). If $\mathcal{L} : \mathcal{K} \rightarrow \mathbb{R}^m$ and $\mathcal{T} : \mathbb{R}^m \rightarrow L_\mu^2(\mathcal{X})$ are instead required to be bounded linear mappings, the resulting quantity is the *linear m -width* (Pinkus, 1985, Def. 4.1). The various m -width notions therefore differ in two aspects: the type of information available about the target operator, and how that information may be used to reconstruct it. In this work, we study the adaptive m -width, since (adaptively chosen) linear samples combined with general reconstruction strategies are of high importance in practical applications.

Recently, Kovachki et al. (2024a) studied the so-called *sampling nonlinear m -width* $s_m(\mathcal{K})_{L_\mu^2(\mathcal{X})}$ of a set $\mathcal{K} \subset C_b(\mathcal{X})$, which is based on standard information. In principle, it results from (4.3) in the case $\mathcal{Y} = \mathbb{R}$ by replacing the adaptive sampling operator $\mathcal{L} : \mathcal{V} \rightarrow \mathbb{R}^m$ by a pointwise sampling operator $\delta_{\mathbf{X}} : \mathcal{K} \rightarrow \mathbb{R}^m$. More specifically, for fixed $\mathbf{X} = (X_1, \dots, X_m) \in \mathcal{X}^m$, define $\delta_{\mathbf{X}}(F) = (F(X_1), \dots, F(X_m)) \in \mathbb{R}^m$ for every $F \in \mathcal{K}$, and set

$$s_m(\mathcal{K})_{L_\mu^2(\mathcal{X})} := \inf \left\{ \sup_{F \in \mathcal{K}} \|F - \mathcal{T}(\delta_{\mathbf{X}}(F))\|_{L_\mu^2(\mathcal{X})} : \mathbf{X} \in \mathcal{X}^m, \mathcal{T} : \mathbb{R}^m \rightarrow L_\mu^2(\mathcal{X}) \right\}. \quad (4.4)$$

The definition in Kovachki et al. (2024a) is slightly more general and we refer to the original article for more details. Since the adaptive m -width includes recovery strategies based on standard information if we choose $\mathcal{V} = C_b(\mathcal{X})$, it constitutes a lower bound for the sampling nonlinear m -width: $s_m(\mathcal{K})_{L_\mu^2(\mathcal{X})} \geq \Theta_m(\mathcal{K}; C_b(\mathcal{X}), L_\mu^2(\mathcal{X}))$. This relation allows us to use lower bounds on the adaptive m -width (Theorems 4.4 and 4.13) to generalize some results in Kovachki et al. (2024a). Further details are discussed in Section 4.4.

Our main result in this section is the tight characterization of the adaptive m -width of the Sobolev unit (Lipschitz) ball, defined in (3.2), in terms of the weights u_γ , given by (2.8). The lower bound for the adaptive m -width pertains to arbitrary $\mathcal{V}$ which contain the Sobolev unit (Lipschitz) ball. For the upper bound, we require the additional assumption that $\mathcal{V}$ is continuously embedded in $L_\mu^2(\mathcal{X}; \mathcal{Y})$.

Theorem 4.4 (Tight characterization of adaptive m -width). *Let $\mathcal{K} \in \{B_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y}), B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y})\}$ and let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be a normed vector subspace of $L_\mu^2(\mathcal{X}; \mathcal{Y})$ containing $\mathcal{K}$. For every $m \in \mathbb{N}$, we have the lower bound*

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

where $\pi : \mathbb{N} \rightarrow \Gamma$ is a bijection which gives a nonincreasing rearrangement of $\mathbf{u} = (u_\gamma)_{\gamma \in \Gamma}$, see (2.9). If, in addition, $\mathcal{V}$ is continuously embedded in $L_\mu^2(\mathcal{X}; \mathcal{Y})$, we have for every $m \in \mathbb{N}$ the

matching upper bound

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

Suppose in the following that $\mathcal{V}$ is continuously embedded in $L_\mu^2(\mathcal{X}; \mathcal{Y})$. It is worth noting that in this case the adaptive m -width of the Sobolev unit (Lipschitz) ball is not only bounded from above and below by $u_{\pi(\mathbf{m}+1)}$ up to constants, as proven in [Adcock et al. \(2024c\)](#) for holomorphic operators, but it is exactly equal to $u_{\pi(\mathbf{m}+1)}$. A similar characterization holds for the linear m -width of unit balls of weighted Hilbert spaces ([Pinkus, 1985](#), Thm. 2.2). See also ([Pietsch, 1980](#), Thm. 11.11.3) for an equivalent formulation in terms of approximation numbers of diagonal operators between ℓ^p -spaces. Hence, Theorem 4.4 can be interpreted as an extension of this result to the setting of adaptive sampling and general (possibly nonlinear) reconstruction strategies. We highlight, however, two aspects that show that, besides this extension, the reach of Theorem 4.4 is considerably broader. First, it covers the recovery of operators from $\mathcal{Y}$ -valued samples (not only scalar-valued ones). Second, it also applies to the Sobolev unit Lipschitz ball, which is the unit ball of a pre-Hilbert space, but not of a Hilbert space (as required by the linear m -width characterization).

Theorem 4.4 has important implications for the role of adaptivity in our approximation problem. It is well-known in IBC that “adaptivity does not help”, that is, adaptive methods in approximation problems can be better than non-adaptive methods in terms of the smallest worst-case error only by a factor of at most 2. We refer to [Novak \(1996\)](#) and [Krieg et al. \(2025\)](#) and references therein. For non-empty, convex, symmetric subsets of pre-Hilbert spaces, the factor 2 can in fact be reduced to 1 ([Novak and Woźniakowski, 2008](#), Thm. 4.5). That is, adaptivity does not even help to improve the constants in the error bounds. We emphasize that these “adaptivity does not help” results hold for recovery strategies based on scalar-valued information. Theorem 4.4 extends this principle to the recovery of operators from $\mathcal{Y}$ -valued information which is generated by Hilbert-valued adaptive sampling operators, as defined in Definition 4.2. Indeed, it shows that the smallest worst-case error in our approximation problem is equal to $u_{\pi(\mathbf{m}+1)}$, and it is attained by the non-adaptive linear projection $F \mapsto F_S$, see (3.1), with $S = \{\pi(\mathbf{1}), \dots, \pi(\mathbf{m})\}$.

Finally, we remark that Theorem 4.4 implies that the adaptive m -width of both the Sobolev unit ball and the Sobolev unit Lipschitz ball coincide. From an IBC point of view, there is thus no difference between the space $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ and the space $\text{Lip}(\mathcal{X}; \mathcal{Y})$, equipped with the $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ -norm, in terms of how well elements in these spaces can be approximated by (adaptively chosen) linear samples.

4.3. Proof of Theorem 4.4

We briefly outline the idea of the proof of Theorem 4.4. The upper bound follows standard computations. It suffices to show that the L_μ^2 -projection $F \mapsto F_S$, defined in (3.1), can be written as the composition of a (non-adaptive) sampling operator and a suitable reconstruction mapping. Details are provided in Section 4.3.3.

The lower bound is less obvious. We first reduce the continuous problem to a discrete one in N dimensions by considering a suitable Lipschitz operator. The adaptive m -width of the

continuous problem can be bounded from below by the adaptive compressive m -width of the discrete problem. The latter, in turn, is lower bounded by the Gelfand width of certain weighted balls in $\mathbb{R}^N$ with weights given by the u_γ . We then use known duality relations between the Gelfand and Kolmogorov width and an explicit characterization of the latter in terms of the u_γ to conclude the proof. We recall relevant notions and results about widths in Section 4.3.1. Further information can be found in [Foucart and Rauhut \(2013, Ch. 10\)](#). The proof of the lower bound is provided in Section 4.3.2.

4.3.1. 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

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

An equivalent characterization is given by

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

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

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

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

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

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

Theorem 4.5 ([Foucart and Rauhut \(2013, Thm. 10.4\)](#)). *If $\mathcal{K}$ is symmetric with respect to the origin, i.e., $-\mathcal{K} = \mathcal{K}$, then $d^m(\mathcal{K}, \mathcal{Z}) \leq E_{\text{ada}}^m(\mathcal{K}, \mathcal{Z})$.*

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 introduce the following notation. For $1 \leq p \leq \infty$, an index set I , and positive weights $\mathbf{w} = (w_i)_{i \in I} > \mathbf{0}$, we define the space $\ell_{\mathbf{w}}^p(I)$ as the set of all sequences of real numbers $\mathbf{x} = (x_i)_{i \in I}$ whose norm $\|\mathbf{x}\|_{\ell_{\mathbf{w}}^p(I)}$ is finite, where

$$\|\mathbf{x}\|_{\ell_{\mathbf{w}}^p(I)} := \begin{cases} \left(\sum_{i \in I} w_i^{-p} |x_i|^p \right)^{1/p} & \text{if } 1 \leq p < \infty, \\ \sup_{i \in I} \{w_i^{-1} |x_i|\} & \text{if } p = \infty. \end{cases}$$

We denote the closed unit ball in $\ell_{\mathbf{w}}^p(I)$ by

$$B_{\mathbf{w}}^p(I) := \left\{ \mathbf{x} \in \ell_{\mathbf{w}}^p(I) : \|\mathbf{x}\|_{\ell_{\mathbf{w}}^p(I)} \leq 1 \right\}.$$

If $\mathbf{w} = \mathbf{1}$, we write $(\ell^p(I), \|\cdot\|_{\ell^p(I)})$ and $B^p(I)$.

Theorem 4.6 (Characterization of Kolmogorov width). *Let $N \in \mathbb{N}$ with $N > m$, $1 \leq q < p \leq \infty$, and $\mathbf{w} \in \mathbb{R}^N$ be a vector of positive weights. Then,*

$$d_m(B_{\mathbf{w}}^p([N]), \ell^q([N])) = \left(\max_{\substack{i_1, \dots, i_{N-m} \in [N] \\ i_k \neq i_j}} \left(\sum_{j=1}^{N-m} w_{i_j}^{\frac{pq}{p-q}} \right)^{\frac{1}{p} - \frac{1}{q}} \right)^{-1}.$$

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). For a (possibly finite) sequence $\mathbf{w} = (w_i)_{i \in I}$ of nonvanishing real numbers, we write $1/\mathbf{w} := (1/w_i)_{i \in I}$ for the sequence of reciprocals.

Theorem 4.7 (Duality of Kolmogorov width and Gelfand width). *For $1 \leq p, q \leq \infty$, let $\mathbf{w} \in \mathbb{R}^N$ be a vector of positive weights and let $1 \leq p^*, q^* \leq \infty$ be such that $1/p + 1/p^* = 1$ and $1/q + 1/q^* = 1$. Then $d_m(B^p([N]), \ell^q([N])) = d^m(B_{1/\mathbf{w}}^{q^*}([N]), \ell^{p^*}([N]))$.*

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

Lemma 4.8 (Adcock et al. (2024c, Lem. B.4)). *Let $\mathbf{w} \in \mathbb{R}^N$ be a vector of positive weights and $1 \leq p, q \leq \infty$. Then $d_m(B^p([N]), \ell^q([N])) = d_m(B_{1/\mathbf{w}}^p([N]), \ell^q([N]))$.*

4.3.2. Lower bound

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

$$\Theta_m(B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y}); \mathcal{V}, L_{\mu}^2(\mathcal{X}; \mathcal{Y})) \geq u_{\pi(\mathbf{m}+1)}, \quad \forall m \in \mathbb{N}. \quad (4.5)$$

With the notions and results from Section 4.3.1, we can concretize the proof strategy that we outlined previously. We proceed in two main steps. We first reduce the continuous problem to a discrete one which can be analyzed by the adaptive compressive m -width of the unit ball in a space of suitably weighted finite sequences, see the proof of Lemma 4.9. In the discrete setting, we apply Theorems 4.5 and 4.7 as well as Lemma 4.8 to relate the adaptive compressive m -width to the Kolmogorov m -width. In the last step, we use the characterization of the latter, provided by Theorem 4.6, in combination with a limiting argument to conclude the claim.

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

Lemma 4.9 (Reduction to discrete problem). *Let $I \subset \Gamma$ be a finite index set. Then, for every constant $c \in (0, 1)$, we have*

$$\Theta_m(B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y}); \mathcal{V}, L_{\mu}^2(\mathcal{X}; \mathcal{Y})) \geq d^m(B_{(c\mathbf{u})_I}^2(I), \ell^2(I)).$$

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

$$\tilde{H}_{n,R}(x) := \begin{cases} H_n(x) & \text{if } -R \leq x \leq R, \\ H_n(R) & \text{if } x > R, \\ H_n(-R) & \text{if } x < -R. \end{cases} \quad (4.6)$$

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

$$\tilde{H}_{\gamma,R,d} : \mathbb{R}^d \rightarrow \mathbb{R}, \quad \tilde{H}_{\gamma,R,d}(\mathbf{x}) := \prod_{i=1}^d \tilde{H}_{\gamma_i,R}(x_i), \quad (4.7)$$

as well as

$$\tilde{H}_{\gamma,R,\lambda} : \mathcal{X} \rightarrow \mathbb{R}, \quad \tilde{H}_{\gamma,R,\lambda}(X) := \prod_{i=1}^{\infty} \tilde{H}_{\gamma_i,R} \left(\frac{\langle X, \phi_i \rangle_{\mathcal{X}}}{\sqrt{\lambda_i}} \right). \quad (4.8)$$

Before proving Lemma 4.9, we need a couple of preliminary results.

Lemma 4.10 (Lipschitz continuity). *For every $R > 0$ and every $\gamma \in \Gamma$, the functional $\tilde{H}_{\gamma,R,\lambda} : \mathcal{X} \rightarrow \mathbb{R}$, defined in (4.8), is Lipschitz continuous.*

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

$$S_{\lambda,d} : \mathcal{X} \rightarrow \mathbb{R}^d, \quad S_{\lambda,d}(X) := \left(\frac{\langle X, \phi_i \rangle_{\mathcal{X}}}{\sqrt{\lambda_i}} \right)_{i=1}^d,$$

and note that $\tilde{H}_{\gamma,R,\lambda} = \tilde{H}_{\gamma,R,d} \circ S_{\lambda,d}$, with $\tilde{H}_{\gamma,R,d}$ given by (4.7). As $S_{\lambda,d}$ is Lipschitz continuous, it suffices to show that $\tilde{H}_{\gamma,R,d}$ is Lipschitz continuous. This, in turn, follows by a simple induction argument on d . $\square$

By Lemma 4.10 and Theorem 2.11, we have $\tilde{H}_{\gamma,R,\lambda} \in W_{\mu,\mathbf{b}}^{1,2}(\mathcal{X})$ for every $R > 0$. By Lemma C.6, we also know that $H_{\gamma,\lambda} \in W_{\mu,\mathbf{b}}^{1,2}(\mathcal{X})$. The next result establishes the connection between $\tilde{H}_{\gamma,R,\lambda}$ and $H_{\gamma,\lambda}$ in the limit $R \rightarrow \infty$. The proof is given in Appendix E.2.

Lemma 4.11 (Convergence in $W_{\mu,\mathbf{b}}^{1,2}(\mathcal{X})$). *For every $\gamma \in \Gamma$, we have*

$$\lim_{R \rightarrow \infty} \tilde{H}_{\gamma,R,\lambda} = H_{\gamma,\lambda} \quad \text{in } W_{\mu,\mathbf{b}}^{1,2}(\mathcal{X}).$$

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 $\tilde{H}_{\gamma,R,\lambda}$. The proof is given in Appendix E.3.

Lemma 4.12 (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*

$$(1 - \varepsilon) \|\mathbf{x}\|_{\ell^2(I)}^2 \leq \left\| \sum_{\gamma \in I} x_\gamma \tilde{H}_{\gamma, R, \lambda} \right\|_{L_\mu^2(\mathcal{X})}^2 \leq (1 + \varepsilon) \|\mathbf{x}\|_{\ell^2(I)}^2, \quad \forall \mathbf{x} = (x_\gamma)_{\gamma \in I} \in \mathbb{R}^I. \quad (4.9)$$

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

We are now ready to reduce the adaptive m -width to a suitable Gelfand m -width. For a sequence $\mathbf{u} = (u_\gamma)_{\gamma \in \Gamma}$ and an index set $I \subset \Gamma$, we henceforth denote by $\mathbf{u}_I$ the subsequence $(u_\gamma)_{\gamma \in I}$.

Proof of Lemma 4.9 Let $\mathcal{L} : \mathcal{V} \rightarrow \mathcal{Y}^m$ be an adaptive sampling operator as in Definition 4.2. Then there exist $Y, \tilde{Y} \in \mathcal{Y} \setminus \{0\}$ and a normed vector space $\tilde{\mathcal{V}} \subset L_\mu^2(\mathcal{X})$ such that, if $YF \in \mathcal{V}$ for some $F \in L_\mu^2(\mathcal{X})$, then $F \in \tilde{\mathcal{V}}$ and $\mathcal{L}(YF) = \tilde{Y}\tilde{\mathcal{L}}(F)$, where $\tilde{\mathcal{L}} : \tilde{\mathcal{V}} \rightarrow \mathbb{R}^m$ is a scalar-valued adaptive sampling operator as in Definition 4.1. Next, let $I \subset \Gamma$ be a finite subset and let us fix $c \in (0, 1)$ and $\varepsilon > 0$. By Lemma 4.11 and Lemma 4.12, there exists $R > 0$ sufficiently large such that (4.9) holds and

$$\left\| \tilde{H}_{\gamma, R, \lambda} - H_{\gamma, \lambda} \right\|_{W_{\mu, b}^{1,2}(\mathcal{X})} \leq \frac{1-c}{c} |I|^{-1/2}, \quad \forall \gamma \in I. \quad (4.10)$$

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

$$F_R := \frac{Y}{\|Y\|_{\mathcal{Y}}} \sum_{\gamma \in I} x_\gamma \tilde{H}_{\gamma, R, \lambda} \quad \text{and} \quad F := \frac{Y}{\|Y\|_{\mathcal{Y}}} \sum_{\gamma \in I} x_\gamma H_{\gamma, \lambda}. \quad (4.11)$$

Note that $\|F\|_{W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})} = \|\mathbf{x}\|_{\ell_{\mathbf{u}_I}^2(I)} \leq c$ by Theorem 2.6. Therefore, by the triangle inequality and (4.10), we have

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

The last step follows from Hölder's inequality and $\|\mathbf{x}\|_{\ell^2(I)} \leq \|\mathbf{x}\|_{\ell_{\mathbf{u}_I}^2(I)}$. Thus, by the triangle inequality, we have $\|F_R\|_{W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})} \leq 1$. By Lemma 4.10 and since I is finite, the operator F_R is Lipschitz continuous and we conclude $F_R \in B_{\mu, b}^{\text{Lip}}(\mathcal{X}; \mathcal{Y})$.

By Lemma 4.12, the family of functionals $\{\tilde{H}_{\gamma, R, \lambda}\}_{\gamma \in I}$ is a Riesz basis of $\mathcal{F} := \text{span}\{\tilde{H}_{\gamma, R, \lambda} : \gamma \in I\}$. Hence, there exists a unique biorthogonal dual basis $\{\hat{H}_{\gamma, R, \lambda}\}_{\gamma \in I}$ such

that $\langle \tilde{H}_{\gamma,R,\lambda}, \hat{H}_{\gamma',R,\lambda} \rangle_{L_\mu^2(\mathcal{X})} = \delta_{\gamma,\gamma'}$ for every $\gamma, \gamma' \in I$. Here, $\delta_{\gamma,\gamma'} = \prod_{i=1}^\infty \delta_{\gamma_i, \gamma'_i}$ is the product of Kronecker deltas. The orthogonal projection onto $\mathcal{F}$ is given by

$$P_{\mathcal{F}} : L_\mu^2(\mathcal{X}) \rightarrow \mathcal{F}, \quad P_{\mathcal{F}} g := \sum_{\gamma \in I} \left\langle g, \hat{H}_{\gamma,R,\lambda} \right\rangle_{L_\mu^2(\mathcal{X})} \tilde{H}_{\gamma,R,\lambda}.$$

Let $\{\psi_j\}_{j=1}^\infty$ be an orthonormal basis of $\mathcal{Y}$. For $G \in L_\mu^2(\mathcal{X}; \mathcal{Y})$, we write $g_j := \langle G, \psi_j \rangle_{\mathcal{Y}} \in L_\mu^2(\mathcal{X})$ and find by (4.9) that

$$\begin{aligned} \|G\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})}^2 &= \sum_{j=1}^\infty \|g_j\|_{L_\mu^2(\mathcal{X})}^2 \geq \sum_{j=1}^\infty \|P_{\mathcal{F}} g_j\|_{L_\mu^2(\mathcal{X})}^2 \geq \sum_{j=1}^\infty (1-\varepsilon) \sum_{\gamma \in I} \left| \left\langle g_j, \hat{H}_{\gamma,R,\lambda} \right\rangle_{L_\mu^2(\mathcal{X})} \right|^2 \\ &= (1-\varepsilon) \sum_{\gamma \in I} \left\| \int_{\mathcal{X}} G \hat{H}_{\gamma,R,\lambda} d\mu \right\|_{\mathcal{Y}}^2. \end{aligned} \quad (4.12)$$

We now define the scalar-valued adaptive sampling operator

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

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

$$\tilde{\mathcal{T}} : \mathbb{R}^m \rightarrow L_\mu^2(\mathcal{X}; \mathcal{Y}), \quad \tilde{\mathcal{T}}(\mathbf{z}) := \mathcal{T}(\tilde{Y}\mathbf{z}),$$

and observe that $\mathcal{T}(\mathcal{L}(F_R)) = \mathcal{T}(\tilde{Y}\Xi_R(\mathbf{x})) = \tilde{\mathcal{T}}(\Xi_R(\mathbf{x}))$. We now set $G := F_R - \mathcal{T}(\mathcal{L}(F_R))$ in (4.12). We use the estimate $\|Z\|_{\mathcal{Y}} \geq \|Y\|_{\mathcal{Y}}^{-1} |\langle Z, Y \rangle_{\mathcal{Y}}|$, which holds for every $Z \in \mathcal{Y}$ by the Cauchy-Schwarz inequality, and compute

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

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

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

and conclude that

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

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

$$\begin{aligned} & \Theta_m(B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y}); \mathcal{V}, L_{\mu}^2(\mathcal{X}; \mathcal{Y})) \\ &= \inf \left\{ \sup_{F \in B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y})} \|F - \mathcal{T}(\mathcal{L}(F))\|_{L_{\mu}^2(\mathcal{X}; \mathcal{Y})} : \mathcal{L} : \mathcal{V} \rightarrow \mathcal{Y}^m \text{ adaptive, } \mathcal{T} : \mathcal{Y}^m \rightarrow L_{\mu}^2(\mathcal{X}; \mathcal{Y}) \right\} \\ &\geq (1 - \varepsilon)^{1/2} \inf \left\{ \sup_{\substack{\mathbf{x} \in \mathbb{R}^I \\ \|\mathbf{x}\|_{\ell_{\mathbf{u}_I}^2(I)} \leq c}} \|\mathbf{x} - \Delta(\Xi(\mathbf{x}))\|_{\ell^2(I)} : \Xi : \mathbb{R}^I \rightarrow \mathbb{R}^m \text{ adaptive, } \Delta : \mathbb{R}^m \rightarrow \mathbb{R}^I \right\} \\ &= (1 - \varepsilon)^{1/2} E_{\text{ada}}^m(cB_{\mathbf{u}_I}^2(I), \ell^2(I)) = (1 - \varepsilon)^{1/2} E_{\text{ada}}^m(B_{(\mathbf{c}\mathbf{u})_I}^2(I), \ell^2(I)), \end{aligned}$$

where in the last step we used that $cB_{\mathbf{u}_I}^2(I) = B_{(\mathbf{c}\mathbf{u})_I}^2(I)$. As $\varepsilon > 0$ was arbitrary, we can take the limit $\varepsilon \rightarrow 0^+$. The claim now follows by Theorem 4.5. $\square$

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

Proof of (4.5) Let $N \in \mathbb{N}$ with $N > m$, $I = \pi([N]) = \{\pi(\mathbf{1}), \dots, \pi(\mathbf{N})\} \subset \Gamma$ be the index set of the N largest entries of $\mathbf{u}$, and fix $c \in (0, 1)$. By Theorem 4.7 and Lemma 4.8, we have

$$\begin{aligned} d^m(B_{(\mathbf{c}\mathbf{u})_I}^2(I), \ell^2(I)) &= d_m(B^2(I), \ell_{1/(\mathbf{c}\mathbf{u})_I}^2(I)) \\ &= d_m(B_{(\mathbf{c}\mathbf{u})_I}^2(I), \ell^2(I)) = c \cdot d_m(B_{\mathbf{u}_I}^2(I), \ell^2(I)), \end{aligned} \tag{4.13}$$

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

$$d_m(B_{\mathbf{u}_I}^2(I), \ell^2(I)) \geq d_m(N^{-r}B_{\mathbf{u}_I}^p(I), \ell^2(I)) = N^{-r}d_m(B_{\mathbf{u}_I}^p(I), \ell^2(I)). \quad (4.14)$$

Applying Theorem 4.6 with $q = 2$ yields

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

Since $(u_{\boldsymbol{\pi}(\mathbf{i})})_{i=1}^\infty$ is nonincreasing, it follows with $q = q(p) := \frac{2p}{p-2} \in (2, \infty)$ that

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

We combine this estimate with (4.13), (4.14), and Lemma 4.9, and conclude

$$u_{\boldsymbol{\pi}(\mathbf{m}+\mathbf{1})} \leq c^{-1} N^r \Theta_m(B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y}); \mathcal{V}, L_\mu^2(\mathcal{X}; \mathcal{Y})).$$

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

$$u_{\boldsymbol{\pi}(\mathbf{m}+\mathbf{1})} \leq c^{-1} \Theta_m(B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y}); \mathcal{V}, L_\mu^2(\mathcal{X}; \mathcal{Y})).$$

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

4.3.3. Upper bound

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

$$\begin{aligned} \Theta_m(\mathcal{K}; \mathcal{V}, L_\mu^2(\mathcal{X}; \mathcal{Y})) &\leq \inf_{S \subset \Gamma, |S| \leq m} \sup_{F \in \mathcal{K}} \|F - F_S\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})} \\ &\leq \sup_{F \in \mathcal{K}} \|F - F_{\{\boldsymbol{\pi}(\mathbf{1}), \dots, \boldsymbol{\pi}(\mathbf{m})\}}\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})} \leq u_{\boldsymbol{\pi}(\mathbf{m}+\mathbf{1})}. \end{aligned} \quad (4.15)$$

Proof of (4.15) The second and third inequality hold by (3.3), so we only need to prove the first inequality. We fix $m \in \mathbb{N}$ and $S = \{\gamma^{(\mathbf{1})}, \dots, \gamma^{(\mathbf{n})}\} \subset \Gamma$ with $n \leq m$. We define the Hilbert-valued

adaptive sampling operator

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

and the reconstruction mapping

$$\mathcal{T} : \mathcal{Y}^m \rightarrow L_{\mu}^2(\mathcal{X}; \mathcal{Y}), \quad \mathcal{T}(Y) := \sum_{i=1}^n Y_i H_{\gamma^{(i)}, \lambda}.$$

Since $\mathcal{V}$ is continuously embedded in $L_{\mu}^2(\mathcal{X}; \mathcal{Y})$, 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 4.2. It suffices to show that there exists $Y \in \mathcal{Y} \setminus \{0\}$, a normed vector space $\tilde{\mathcal{V}} \subset L_{\mu}^2(\mathcal{X})$, and a scalar-valued adaptive sampling operator $\tilde{\mathcal{L}} : \tilde{\mathcal{V}} \rightarrow \mathbb{R}^m$ such that, if $YF \in \mathcal{V}$ for some $F \in L_{\mu}^2(\mathcal{X})$, then $F \in \tilde{\mathcal{V}}$ and $\mathcal{L}(YF) = Y\tilde{\mathcal{L}}(F)$. To this end, we choose some $Y \in \mathcal{Y}$ with $\|Y\|_{\mathcal{Y}} = 1$ and define the space $\tilde{\mathcal{V}} := \{F \in L_{\mu}^2(\mathcal{X}) : YF \in \mathcal{V}\}$. It can be readily checked that this defines a normed vector space with norm given by $\|F\|_{\tilde{\mathcal{V}}} := \|YF\|_{\mathcal{V}}$ for any $F \in \tilde{\mathcal{V}}$. Moreover, as $\mathcal{V}$ is continuously embedded in $L_{\mu}^2(\mathcal{X}; \mathcal{Y})$, there exists a constant $C > 0$ such that

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

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

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

Note that $\tilde{\mathcal{L}}$ is linear and by the continuous embedding of $\tilde{\mathcal{V}}$ in $L_{\mu}^2(\mathcal{X})$, it is also bounded. Hence, $\tilde{\mathcal{L}}$ is a scalar-valued adaptive sampling operator. Moreover, by construction, if $YF \in \mathcal{V}$, then $F \in \tilde{\mathcal{V}}$ and $\mathcal{L}(YF) = Y\tilde{\mathcal{L}}(F)$. We conclude that $\mathcal{L}$ is indeed an adaptive Hilbert-valued sampling operator as in Definition 4.2. Consequently,

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

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

4.4. Discussion

Theorem 4.4 shows that linear Hermite polynomial approximation based on the index set $S = \{\pi(\mathbf{1}), \dots, \pi(\mathbf{m})\}$ is optimal for the uniform approximation of $W_{\mu, \mathbf{b}}^{1,2}$ - and Lipschitz operators with Sobolev norm at most one among all possible recovery strategies which are based on (adaptive) linear information. Moreover, Theorem 4.4 in combination with Theorem 3.5 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 Section 1.1, a related *curse of data complexity* was shown in Kovachki et al. (2024a) by means of the sampling nonlinear m -width $s_m(\mathcal{K})_{L_\mu^2(\mathcal{X})}$, defined in (4.4). Let $B_{\mu,\mathbf{b}}^{C^{0,1}}(\mathcal{X};\mathcal{Y}) = B_{\mu,\mathbf{b}}^{\text{Lip}}(\mathcal{X};\mathcal{Y}) \cap C^{0,1}(\mathcal{X};\mathcal{Y})$ be the Sobolev unit $C^{0,1}$ -ball. Theorem 2.12 in Kovachki et al. (2024a) implies the following result.

Theorem 4.13 (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*

$$s_m(B_{\mu,\mathbf{1}}^{C^{0,1}}(\mathcal{X}))_{L_\mu^2(\mathcal{X})} \geq C \log(m+1)^{-(\alpha+3)}, \quad \forall m \in \mathbb{N}.$$

Our findings in the present section generalize this result in several directions. Notice that the operator F_R in (4.11), which we use to prove the lower bound of the adaptive m -width, is not only Lipschitz continuous, but in fact also bounded. As such, it lies in $B_{\mu,\mathbf{b}}^{C^{0,1}}(\mathcal{X};\mathcal{Y})$. Following the proof of Lemma 4.9, we deduce that $\Theta_m(B_{\mu,\mathbf{b}}^{C^{0,1}}(\mathcal{X};\mathcal{Y}); \mathcal{V}, L_\mu^2(\mathcal{X};\mathcal{Y})) \geq u_{\pi(m+1)}$ for any vector subspace $\mathcal{V} \subset L_\mu^2(\mathcal{X};\mathcal{Y})$ which contains $B_{\mu,\mathbf{b}}^{C^{0,1}}(\mathcal{X};\mathcal{Y})$. As already noted in Section 4.2, by setting $\mathcal{Y} = \mathbb{R}$, $\mathcal{V} = C_b(\mathcal{X})$, and $\mathbf{b} = \mathbf{1}$, we can thus conclude that

$$s_m(B_{\mu,\mathbf{1}}^{C^{0,1}}(\mathcal{X}))_{L_\mu^2(\mathcal{X})} \geq \Theta_m(B_{\mu,\mathbf{1}}^{C^{0,1}}(\mathcal{X}); C_b(\mathcal{X}), L_\mu^2(\mathcal{X})) \geq u_{\pi(m+1)}.$$

This bound pertains to *general* centered, nondegenerate Gaussian measures and together with Theorem 3.5 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 3.6 provides upper bounds for the adaptive m -width of the Sobolev unit (Lipschitz) ball in terms of the decay of the PCA eigenvalues $\lambda_{\mathbf{b},i}$. 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 $\lambda_{\mathbf{b},i}$ is sufficiently fast, e.g., double exponential.

5. 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 $W_{\mu,\mathbf{b}}^{1,2}(\mathcal{X};\mathcal{Y})$. 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 $\lambda_{\mathbf{b},i}$ of the covariance operator of the Gaussian measure μ .

Next, we analyzed the smallest worst-case error in reconstructing Lipschitz and $W_{\mu,\mathbf{b}}^{1,2}$ -operators from m arbitrary (potentially adaptively chosen) linear samples in terms of the corresponding adaptive m -width. We extended the IBC principle that “adaptivity does not help” in reducing the approximation error from scalar- to $\mathcal{Y}$ -valued samples and showed that

adaptivity does not even help to improve any constants in the error bounds. Furthermore, we proved that from an IBC point of view there is no difference between the space of all Lipschitz operators, equipped with the $W_{\mu, \mathbf{b}}^{1,2}$ -norm, and the space of all $W_{\mu, \mathbf{b}}^{1,2}$ -operators: The adaptive m -widths in both cases coincide. We tightly characterized the dependence of the adaptive m -width on the $\lambda_{\mathbf{b}, i}$ 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 $W_{\mu, \mathbf{b}}^{1,2}$ -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 $W_{\mu, \mathbf{b}}^{1,2}$ -regularity, and Lipschitz regularity in particular, suffices to achieve approximation rates which are arbitrarily close to any algebraic rate, provided that the PCA eigenvalues $\lambda_{\mathbf{b}, i}$ 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 that 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 *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, *Adv. Neural Inf. Process. Syst.*, 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. Begeed-Dov. Lower and upper bounds for the number of lattice points in a simplex. *SIAM J. Appl. Math.*, 22(1):106–108, 1972.
- Y. Benyamini and J. Lindenstrauss. *Geometric Nonlinear Functional Analysis*, volume 48 of *Colloquium Publications*. American Mathematical Society, Providence, RI, 1999.
- 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. *PNAS*, 120(39):e2303904120, 2023.
- 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, *Adv. Neural Inf. Process. Syst.*, volume 35, pages 10945–10958. Curran Associates, Inc., 2022.
- R. DeVore. Nonlinear approximation. *Acta Numer.*, 7:51–150, 1998.
- R. DeVore, R. Howard, and C. Micchelli. Optimal nonlinear approximation. *Manuscr. Math.*, 63(4): 469–478, 1989.
- 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*, volume 63 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics*. Springer International Publishing, Cham, 1 edition, 2016.
- P. Jaillet, D. Lamberton, and B. Lapeyre. Variational inequalities and the pricing of American options. *Acta Appl. Math.*, 21(3):263–289, 1990.
- G. Jovet and E. Bueler. Steady, shallow ice sheets as obstacle problems: Well-posedness and finite element approximation. *SIAM J. Appl. Math.*, 72(4):1292–1314, 2012.
- 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. *Forum of Mathematics, Sigma*, 13:e152, 2025.
- 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. *IMA J. Numer. Anal.*, 46(2):647–712, 2026.
- 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. *Constr. Approx.*, 62(2):261–303, 2025.
- Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *Int. Conf. Learn. Represent.*, 2021. URL <https://openreview.net/forum?id=c8P9NQVtmnO>.
- 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.
- I. López. Sobolev spaces and potential spaces associated to Hermite polynomials expansions. *Extracta Math.*, 33(2):229–249, 2018.
- 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/lehmaterial_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.
- F. Mignot. Contrôle dans les inéquations variationelles elliptiques. *J. Funct. Anal.*, 22(2):130–185, 1976.
- 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.
- A. Pietsch. *Operator Ideals*, volume 20 of *North-Holland Mathematical Library*. Elsevier, 1980.
- A. Pinkus. *N-Widths in Approximation Theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer, Berlin Heidelberg, 1985.
- M. Reiß and M. Wahl. Nonasymptotic upper bounds for the reconstruction error of PCA. *Ann. Statist.*, 48(2):1098–1123, 2020.
- J.-F. Rodrigues. *Obstacle Problems in Mathematical Physics*, volume 134 of *North-Holland Mathematics Studies*. Elsevier, Amsterdam New York, NY, 1987.
- 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. *Anal. Appl.*, 24(01):199–239, 2026.
- 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, Ch. 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 $F : \mathcal{X} \rightarrow \mathcal{Y}$ with finite operator norm $\|F\|_{L(\mathcal{X}; \mathcal{Y})} := \sup_{X \in \mathcal{X}, X \neq 0} \|F(X)\|_{\mathcal{Y}} / \|X\|_{\mathcal{X}}$.

Definition A.1 (Differentiability). Let $\mathcal{M}$ be a collection of non-empty subsets of $\mathcal{X}$, $X \in \mathcal{X}$. A mapping $F : \mathcal{X} \rightarrow \mathcal{Y}$ is said to be differentiable with respect to $\mathcal{M}$ at the point X if there exists a continuous linear mapping, denoted by $D^{\mathcal{M}}F(X) \in L(\mathcal{X}; \mathcal{Y})$, such that for every fixed set $M \in \mathcal{M}$, we have

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

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 . Note that in all three cases, the mapping $D^{\mathcal{M}}F(X)$ is unique and it is called the (*Gâteaux*, *Hadamard*, or *Fréchet*) *derivative of F 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 2.11 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 $\mathcal{E} \rightarrow \mathcal{Y}$, $Z \mapsto F(X + Z)$ is differentiable 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 and Malliavin calculus (see Bogachev, 1998; Da Prato, 2014; Lunardi et al., 2015/2016).

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

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

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 $F : \mathcal{X} \rightarrow \mathcal{Y}$ 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 $\mathcal{X} = \mathbb{R}^n$ for some $n \in \mathbb{N}$ and $Z = \mathbf{e}_i$ is the i th standard unit vector, we use the standard notation $\partial_i F(\mathbf{x}) := \frac{\partial}{\partial \mathbf{e}_i} F(\mathbf{x})$ for $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. If $\mathcal{E}$ is a Hilbert subspace of $\mathcal{X}$ and if $F : \mathcal{X} \rightarrow \mathbb{R}$ 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 $\nabla_{\mathcal{E}}F(X) := Z$.

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, Ch. 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 $A : \text{dom}(A) \rightarrow \mathcal{H}_2$ from a linear subspace $\text{dom}(A) \subset \mathcal{H}_1$ to $\mathcal{H}_2$. The set $\text{dom}(A)$ is

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

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

We can equip $\text{dom}(A)$ with the *graph inner product*,

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

which turns it into a pre-Hilbert space with *graph norm*

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

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

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

$$\text{dom}(\bar{A}) = \left\{ H \in \mathcal{H}_1 : \exists (H_i)_{i=1}^\infty \subset \text{dom}(A) : \lim_{i \rightarrow \infty} H_i = H, (AH_i)_{i=1}^\infty \text{ converges in } \mathcal{H}_2 \right\}.$$

For $H \in \text{dom}(\bar{A})$ and any sequence $(H_i)_{i=1}^\infty \subset \text{dom}(A)$ with $\lim_{i \rightarrow \infty} H_i = H$ in $\mathcal{H}_1$ and $(AH_i)_{i=1}^\infty$ being convergent in $\mathcal{H}_2$, we have $\lim_{i \rightarrow \infty} AH_i = \bar{A}H$ in $\mathcal{H}_2$. In other words, closability of A guarantees that the limit $\lim_{i \rightarrow \infty} AH_i$ is independent of the sequence $(H_i)_{i=1}^\infty$, see [Berezanskij et al. \(1996, Thm. 2.1\)](#). If A is closable, we equip $\text{dom}(\bar{A})$ with the graph inner product, which turns $(\text{dom}(\bar{A}), \langle \cdot, \cdot \rangle_{\Gamma_{\bar{A}}})$ 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 $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ (see [Definition 2.2](#)) and to prove [Theorem 2.11](#) 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 $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ as well as some of its important properties in [Appendix C.2](#). We mainly follow [Lunardi et al. \(2015/2016\)](#) and [Da Prato \(2006\)](#), who consider $\mathcal{Y} = \mathbb{R}$ and the cases $\mathbf{b} = \sqrt{\lambda}$ and $\mathbf{b} = \mathbf{1}$, respectively, and generalize the proofs therein to general Hilbert-valued operators in the case $\mathbf{b} \in \ell^\infty(\mathbb{N})$, $\mathbf{b} > \mathbf{0}$. More information can also be found in [Bogachev \(1998\)](#). Throughout this section, we use notation as introduced in the main text.

C.1. The Cameron-Martin space

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

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

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

$$j : \mathcal{X}^* \rightarrow L_\mu^2(\mathcal{X}), \quad F \mapsto F,$$

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

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

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 $L_\mu^2(\mathcal{X})$, equipped with the inner product from $L_\mu^2(\mathcal{X})$, as well as the mapping

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

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 $\|H\|_{\mathcal{H}} = \|\hat{H}\|_{L_\mu^2(\mathcal{X})}$. Hence, $R_\mu : \mathcal{X}_\mu^* \rightarrow \mathcal{H}$ is an isometric isomorphism that turns $\mathcal{H}$ into a Hilbert space with inner product $\langle H, K \rangle_{\mathcal{H}} := \langle \hat{H}, \hat{K} \rangle_{L_\mu^2(\mathcal{X})}$ 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 $\{\xi_i\}_{i=1}^\infty$ with

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

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

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

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 $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$

Recall the weighted space $\mathcal{X}_{\mathbf{b}}$ with weight sequence $\mathbf{b} = (b_i)_{i=1}^\infty \in \ell^\infty(\mathbb{N})$, $\mathbf{b} > \mathbf{0}$, and orthonormal basis $\{\eta_i\}_{i=1}^\infty$, $\eta_i := b_i \phi_i$, as defined in (2.5) and (2.6), respectively.

C.2.1. Construction

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

Definition C.4 (Cylindrical functionals and operators). A functional $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ is called a *cylindrical functional* if there exist $n \in \mathbb{N}$, $\ell_1, \dots, \ell_n \in \mathcal{X}^*$, and a function $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $\mathcal{FC}_b^1(\mathcal{X})$. Moreover, we define the set of all *cylindrical boundedly Fréchet differentiable $\mathcal{Y}$ -valued operators* by

$$\mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y}) := \text{span} \{ \mathcal{X} \ni X \mapsto \varphi(X)Y \in \mathcal{Y} : \varphi \in \mathcal{FC}_b^1(\mathcal{X}), Y \in \mathcal{Y} \},$$

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

Proposition C.5 (Closability of $D_{\mathcal{X}_b}$). *The (Fréchet) differential operator along $\mathcal{X}_b$, $D_{\mathcal{X}_b} : \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y}) \rightarrow L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$, is closable in $L_\mu^2(\mathcal{X}; \mathcal{Y})$.*

Proof First, a short computation shows that for every $F \in \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$, the derivative $D_{\mathcal{X}_b} F$ lies in $L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$. Hence, the mapping $D_{\mathcal{X}_b} : \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y}) \rightarrow L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$ is well-defined. The proof of closability is a straightforward modification of the proof of Lunardi et al. (2015/2016, Lem. 10.2.4), replacing the derivative along the Cameron-Martin space of μ by the derivative along the space $\mathcal{X}_b$. $\square$

C.2.2. Properties

We first observe that the Hermite polynomials $\{H_{\gamma, \lambda}\}_{\gamma \in \Gamma}$ belong to $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. This is not immediate because $H_{\gamma, \lambda}$ is cylindrical and smooth but not bounded. For this, we require some notation. For $\gamma \in \Gamma$ and $i \in \mathbb{N}$, we define $\gamma^{(i)} = (\gamma_k^{(i)})_{k=1}^\infty \in \Gamma$ as follows: If $\gamma_i = 0$, we set $\gamma^{(i)} := \mathbf{0}$, and if $\gamma_i > 0$, we set

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

Lemma C.6 (Hermite polynomials are Gaussian Sobolev functionals). *For every $\gamma \in \Gamma$, we have $H_{\gamma, \lambda} \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X})$ with weak Gaussian derivatives $\frac{\partial}{\partial \eta_i} H_{\gamma, \lambda} = b_i \sqrt{\frac{\gamma_i}{\lambda_i}} H_{\gamma^{(i)}, \lambda}$ for every $i \in \mathbb{N}$.*

Proof Let $d \in \mathbb{N}$ such that $\text{supp}(\gamma) \subset [d]$. Let $\omega \in C_c^\infty(\mathbb{R}^d)$ be a smooth bump function with $0 \leq \omega \leq 1$, $\omega = 1$ on $[-1, 1]^d$, and $\|\nabla \omega\|_\infty \leq C$ for some constant $C > 0$. We define for $n \in \mathbb{N}$

and $X \in \mathcal{X}$,

$$\Upsilon_n(X) := \omega \left(\frac{\langle X, \phi_1 \rangle_{\mathcal{X}}}{n}, \dots, \frac{\langle X, \phi_d \rangle_{\mathcal{X}}}{n} \right) \quad \text{and} \quad \varphi_n(X) := H_{\gamma, \lambda}(X) \Upsilon_n(X).$$

Notice that $\varphi_n \in \mathcal{FC}_b^1(\mathcal{X})$ with $|\varphi_n| \leq |H_{\gamma, \lambda}|$. Also, since $\lim_{n \rightarrow \infty} \Upsilon_n = 1$ pointwise, we have $\lim_{n \rightarrow \infty} \varphi_n = H_{\gamma, \lambda}$ pointwise. Since $H_{\gamma, \lambda}$ is a polynomial, Fernique's theorem (Theorem C.1) implies that $H_{\gamma, \lambda} \in L_\mu^2(\mathcal{X})$. By the dominated convergence theorem (DCT), we conclude that $\lim_{n \rightarrow \infty} \varphi_n = H_{\gamma, \lambda}$ in $L_\mu^2(\mathcal{X})$.

Now consider the derivatives. By (2.6), we have $\frac{\partial}{\partial \eta_i} = b_i \frac{\partial}{\partial \phi_i}$. Hence, by the product rule,

$$\frac{\partial}{\partial \eta_i} \varphi_n(X) = b_i \frac{\partial}{\partial \phi_i} H_{\gamma, \lambda}(X) \Upsilon_n(X) + b_i H_{\gamma, \lambda}(X) \frac{\partial}{\partial \phi_i} \Upsilon_n(X), \quad \forall X \in \mathcal{X}, \forall i \in \mathbb{N}.$$

Again, since $\frac{\partial}{\partial \phi_i} H_{\gamma, \lambda}$ is a polynomial, it lies in $L_\mu^2(\mathcal{X})$. In fact, by equation (10.16) in Da Prato (2006), we have

$$\frac{\partial}{\partial \phi_i} H_{\gamma, \lambda} = \sqrt{\frac{\gamma_i}{\lambda_i}} H_{\gamma^{(i)}, \lambda}.$$

Since $|\Upsilon_n| \leq 1$, it follows from the DCT that $\lim_{n \rightarrow \infty} b_i \frac{\partial}{\partial \phi_i} H_{\gamma, \lambda} \Upsilon_n = b_i \sqrt{\frac{\gamma_i}{\lambda_i}} H_{\gamma^{(i)}, \lambda}$ in $L_\mu^2(\mathcal{X})$. Further, for $i \in [d]$, notice that $\frac{\partial}{\partial \phi_i} \Upsilon_n(X) = \frac{1}{n} \partial_i \omega \left(\frac{\langle X, \phi_1 \rangle_{\mathcal{X}}}{n}, \dots, \frac{\langle X, \phi_d \rangle_{\mathcal{X}}}{n} \right)$ and hence, $|\frac{\partial}{\partial \phi_i} \Upsilon_n| \leq C/n$. Again by the DCT, we find that $\lim_{n \rightarrow \infty} b_i H_{\gamma, \lambda} \frac{\partial}{\partial \phi_i} \Upsilon_n = 0$ in $L_\mu^2(\mathcal{X})$ for every $i \in [d]$. For $i > d$, we have $\frac{\partial}{\partial \eta_i} \varphi_n = 0$. We conclude that for every $i \in \mathbb{N}$, we have

$$\lim_{n \rightarrow \infty} \frac{\partial}{\partial \eta_i} \varphi_n = b_i \sqrt{\frac{\gamma_i}{\lambda_i}} H_{\gamma^{(i)}, \lambda} \quad \text{in } L_\mu^2(\mathcal{X}). \quad (\text{C.3})$$

Altogether, $\lim_{n \rightarrow \infty} \varphi_n = H_{\gamma, \lambda}$ in $L_\mu^2(\mathcal{X})$ and $(\nabla_{\mathcal{X}_b} \varphi_n)_{n=1}^\infty$ is convergent in $L_\mu^2(\mathcal{X}; \mathcal{X}_b)$. This implies by definition that $H_{\gamma, \lambda} \in W_{\mu, b}^{1,2}(\mathcal{X})$ and it follows from (C.3) that the weak Gaussian derivatives are given by $\frac{\partial}{\partial \eta_i} H_{\gamma, \lambda} = b_i \sqrt{\frac{\gamma_i}{\lambda_i}} H_{\gamma^{(i)}, \lambda}$. $\square$

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

Lemma C.7. *Suppose that $(F_n)_{n=1}^\infty \subset W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})$ with $\lim_{n \rightarrow \infty} F_n = F$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$.*

- (i) *If $\sup_{n \in \mathbb{N}} \|D_{\mathcal{X}_b} F_n\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))} < \infty$, then $F \in W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})$.*
- (ii) *If $\lim_{n \rightarrow \infty} D_{\mathcal{X}_b} F_n = G$ in $L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$, then $F \in W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})$ with $D_{\mathcal{X}_b} F = G$.*

Proof Since $W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})$ is a Hilbert space, it is reflexive. Notice that in both cases the assumptions imply that $(F_n)_{n=1}^\infty$ is bounded in $W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})$. Hence, there exists a subsequence $(F_{n_k})_{k=1}^\infty$ which converges weakly in $W_{\mu, b}^{1,2}(\mathcal{X}; \mathcal{Y})$ to some $\tilde{F}$ as $k \rightarrow \infty$. Since $\lim_{k \rightarrow \infty} F_{n_k} = F$

in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ by assumption, we conclude that $F = \tilde{F}$. This shows the claim in case (i). In case (ii), the assumptions imply additionally that $D_{\mathcal{X}_b} \tilde{F} = G$. $\square$

Next, we study the ℓ^2 -characterization of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$, see Theorem 2.6.

Proposition C.8. *Let $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. Then,*

$$\frac{\partial}{\partial \eta_i} F = \sum_{\gamma \in \Gamma} b_i \sqrt{\frac{\gamma_i}{\lambda_i}} \left(\int_{\mathcal{X}} F H_{\gamma, \lambda} d\mu \right) H_{\gamma^{(i)}, \lambda}, \quad \forall i \in \mathbb{N}, \quad (\text{C.4})$$

and

$$\|F\|_{W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})}^2 = \sum_{\gamma \in \Gamma} \left(1 + \sum_{i=1}^{\infty} b_i^2 \frac{\gamma_i}{\lambda_i} \right) \left\| \int_{\mathcal{X}} F H_{\gamma, \lambda} d\mu \right\|_{\mathcal{Y}}^2. \quad (\text{C.5})$$

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

$$\sum_{\gamma \in \Gamma} \left(1 + \sum_{i=1}^{\infty} b_i^2 \frac{\gamma_i}{\lambda_i} \right) \|Y_\gamma\|_{\mathcal{Y}}^2 < \infty, \quad (\text{C.6})$$

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

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

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

Since $\frac{\partial}{\partial \eta_i} \varphi = b_i \frac{\partial}{\partial \phi_i} \varphi$, it follows from (C.7) together with orthonormality of the Hermite polynomials that

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

By linearity, the identities (C.7) and (C.8) hold in fact for every cylindrical operator $\varphi \in \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$. Now suppose that $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. It suffices to prove (C.4), as (C.5) then follows by Parseval's identity (2.4). For this, it is enough to show that

$$\int_{\mathcal{X}} \left(\frac{\partial}{\partial \eta_i} F \right) H_{\gamma^{(i)}, \lambda} d\mu = b_i \sqrt{\frac{\gamma_i}{\lambda_i}} \int_{\mathcal{X}} F H_{\gamma, \lambda} d\mu, \quad \forall i \in \mathbb{N}. \quad (\text{C.9})$$

By definition of $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$, there exists a sequence $(F_n)_{n=1}^\infty \subset \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$ such that $\lim_{n \rightarrow \infty} F_n = F$ and $\lim_{n \rightarrow \infty} \frac{\partial}{\partial \eta_i} F_n = \frac{\partial}{\partial \eta_i} F$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$. We set $\varphi = F_n$ in (C.8) and take the limit $n \rightarrow \infty$ to obtain (C.9).

Conversely, suppose that (C.6) holds for a family of vectors $(Y_\gamma)_{\gamma \in \Gamma} \subset \mathcal{Y}$. We fix an enumeration $\tau : \mathbb{N} \rightarrow \Gamma$ of Γ and define

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

We deduce from Parseval's identity and (C.6) that $(F_n)_{n=1}^\infty$ is a Cauchy sequence in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ with limit $F := \sum_{i=1}^\infty Y_{\tau(i)} H_{\tau(i), \lambda}$. Notice that the limit is independent of the enumeration τ . We can thus write $F = \sum_{\gamma \in \Gamma} Y_\gamma H_{\gamma, \lambda}$. Lemma C.6 implies that $F_n \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ with

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

Consequently, we can bound the $L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$ -norm of the derivatives $D_{\mathcal{X}_b} F_n$ by

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

The right-hand side is finite by (C.6). Lemma C.7(i) now implies that $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. $\square$

Next, we consider operators with special approximation properties that will become important in the proof of Theorem 2.11 in Appendix D. To this end, recall from (C.1), (C.2) the orthonormal basis $\{\xi_i\}_{i=1}^\infty$ 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 $F \in L_\mu^2(\mathcal{X}; \mathcal{Y})$ and $n \in \mathbb{N}$, we define

$$\mathbb{E}_n F : \mathcal{X} \rightarrow \mathcal{Y}, \quad \mathbb{E}_n F := \mathbb{E}[F \mid \hat{\xi}_1, \dots, \hat{\xi}_n], \quad (\text{C.10})$$

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

$$P_n : \mathcal{X} \rightarrow \text{span}\{\xi_i : i \in [n]\}, \quad P_n X := \sum_{i=1}^n \hat{\xi}_i(X) \xi_i. \quad (\text{C.11})$$

Proposition C.9 (Properties of $\mathbb{E}_n F$ in L_μ^2). *For $F \in L_\mu^2(\mathcal{X}; \mathcal{Y})$, let $\mathbb{E}_n F$ and P_n be given as in (C.10) and (C.11), 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

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

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

(ii) For every $F \in L^2_\mu(\mathcal{X}; \mathcal{Y})$, the sequence $(\mathbb{E}_n F)_{n=1}^\infty$ 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. $\square$

Proposition C.10 (Properties of $\mathbb{E}_n F$ in $W^{1,2}_{\mu, \mathbf{b}}$). *Let $F \in W^{1,2}_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y})$ and $\mathbb{E}_n F$ be defined as in (C.10). Then, $\mathbb{E}_n F \in W^{1,2}_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y})$ for every $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} \mathbb{E}_n F = F$ in $W^{1,2}_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y})$.*

We need the following preliminary result for cylindrical operators:

Proposition C.11. *Let $\varphi \in \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$. Then, for every $n \in \mathbb{N}$, we have $\mathbb{E}_n \varphi \in W^{1,2}_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y})$ and the following properties hold true:*

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

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

(ii) We have $\lim_{n \rightarrow \infty} \mathbb{E}_n \varphi = \varphi$ in $W^{1,2}_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y})$.

Proof In the case $\mathcal{Y} = \mathbb{R}$ and $\mathbf{b} = \sqrt{\lambda}$ (where $\eta_i = \xi_i$), the claim follows from [Lunardi et al. \(2015/2016, Prop. 10.1.2\)](#). The case of general $\mathbf{b}$ can be readily checked by using the identity $\frac{\partial}{\partial \eta_i} = b_i \lambda_i^{-1/2} \frac{\partial}{\partial \xi_i}$, see (2.6) and (C.1). The lifting to general $\mathcal{Y}$ is straightforward. $\square$

Proof of Proposition C.10 Let $F \in W^{1,2}_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y})$. Let $(\varphi_m)_{m=1}^\infty \subset \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$ be a sequence of cylindrical operators which converges to F in $W^{1,2}_{\mu, \mathbf{b}}(\mathcal{X}; \mathcal{Y})$. Using Proposition C.9(i), it can be readily checked that $\mathbb{E}_n \varphi_m \in \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$. By the contraction property of the conditional

expectation, we find

$$\|\mathbb{E}_n \varphi_m - \mathbb{E}_n F\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})} \leq \|\varphi_m - F\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})},$$

and hence, $\lim_{m \rightarrow \infty} \mathbb{E}_n \varphi_m = \mathbb{E}_n F$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$. Moreover, using Parseval's identity, Proposition C.11(i), and again the contraction property, we compute for any $\varphi \in \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$

$$\begin{aligned} \|D_{\mathcal{X}_b} \mathbb{E}_n \varphi\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))}^2 &= \int_{\mathcal{X}} \sum_{i=1}^{\infty} \left\| \frac{\partial}{\partial \eta_i} \mathbb{E}_n \varphi(X) \right\|_{\mathcal{Y}}^2 d\mu(X) = \sum_{i=1}^n \left\| \mathbb{E}_n \left(\frac{\partial}{\partial \eta_i} \varphi \right) \right\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})}^2 \\ &\leq \sum_{i=1}^n \left\| \frac{\partial}{\partial \eta_i} \varphi \right\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})}^2 \leq \int_{\mathcal{X}} \sum_{i=1}^{\infty} \left\| \frac{\partial}{\partial \eta_i} \varphi(X) \right\|_{\mathcal{Y}}^2 d\mu(X) \\ &= \|D_{\mathcal{X}_b} \varphi\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))}^2. \end{aligned}$$

By setting $\varphi = \varphi_m - \varphi_{m'}$ with $m, m' \in \mathbb{N}$, we deduce that $(D_{\mathcal{X}_b} \mathbb{E}_n \varphi_m)_{m=1}^\infty$ is a Cauchy sequence in $L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))$. We conclude that $\mathbb{E}_n F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ with $\lim_{m \rightarrow \infty} \mathbb{E}_n \varphi_m = \mathbb{E}_n F$ in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. We know from Proposition C.9 that $\lim_{n \rightarrow \infty} \mathbb{E}_n F = F$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$. Furthermore,

$$\begin{aligned} \|D_{\mathcal{X}_b} \mathbb{E}_n F - D_{\mathcal{X}_b} F\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))} &\leq \|D_{\mathcal{X}_b} \mathbb{E}_n F - D_{\mathcal{X}_b} \mathbb{E}_n \varphi_m\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))} \\ &\quad + \|D_{\mathcal{X}_b} \mathbb{E}_n \varphi_m - D_{\mathcal{X}_b} \varphi_m\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))} \\ &\quad + \|D_{\mathcal{X}_b} \varphi_m - D_{\mathcal{X}_b} F\|_{L_\mu^2(\mathcal{X}; \text{HS}(\mathcal{X}_b; \mathcal{Y}))}. \end{aligned}$$

The first and third term on the right-hand side converge to zero as $m \rightarrow \infty$ by the arguments above. The second term converges to zero as $n \rightarrow \infty$ by Proposition C.11(ii). By a diagonal argument, we conclude that the left-hand side converges to zero as $n \rightarrow \infty$. In conclusion, we have $\lim_{n \rightarrow \infty} \mathbb{E}_n F = F$ in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. $\square$

The final two results identify families of compositional operators in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ that will be needed in Appendices D and E.2. To this end, we set

$$T_n : \mathcal{X} \rightarrow \mathbb{R}^n, \quad X \mapsto (\hat{\xi}_1(X), \dots, \hat{\xi}_n(X)), \quad (\text{C.12})$$

with the functionals $\hat{\xi}_i = \lambda_i^{-1/2} \langle \cdot, \phi_i \rangle_{\mathcal{X}} \in \mathcal{X}^*$, $i \in \mathbb{N}$, defined in (C.2).

Lemma C.12. *Let $F : \mathcal{X} \rightarrow \mathcal{Y}$ be an operator of the form $F = G \circ T_n$ with $G \in W^{1,2}(\mathbb{R}^n; \mathcal{Y})$ for some $n \in \mathbb{N}$ and T_n given by (C.12). Then, $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ with*

$$\frac{\partial}{\partial \eta_i} F(X) = \begin{cases} b_i \lambda_i^{-1/2} \partial_i G(T_n(X)) & \text{if } 1 \leq i \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

for μ -a.e. $X \in \mathcal{X}$.

Proof First observe that the pushforward measure $T_n\#\mu$ is given by $\bigotimes_{i=1}^n \mathcal{N}(0, 1)$. It is then easy to see that $F \in L_\mu^2(\mathcal{X}; \mathcal{Y})$. Next, we construct a sequence of cylindrical operators that converges to F in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ and which yields the claimed derivative expressions. By Hytönen et al. (2016, Cor. 2.5.6), there exists a sequence $(G_m)_{m=1}^\infty \subset C_c^\infty(\mathbb{R}^n; \mathcal{Y})$ such that $\lim_{m \rightarrow \infty} G_m = G$ in $W^{1,2}(\mathbb{R}^n; \mathcal{Y})$. Let $\{\psi_j\}_{j=1}^\infty$ be an orthonormal basis of $\mathcal{Y}$ and define $G_{m,k} := \sum_{j=1}^k \langle G_m, \psi_j \rangle_{\mathcal{Y}} \psi_j$ as well as $F_{m,k} := G_{m,k} \circ T_n$. By construction, we have $F_{m,k} \in \mathcal{FC}_b^1(\mathcal{X}; \mathcal{Y})$. Using Parseval's identity, we compute

$$\begin{aligned} \|F_{m,k} - F\|_{L_\mu^2(\mathcal{X}; \mathcal{Y})}^2 &= \int_{\mathcal{X}} \sum_{j=1}^k \left| \langle G_m(T_n(X)) - G(T_n(X)), \psi_j \rangle_{\mathcal{Y}} \right|^2 d\mu(X) \\ &\quad + \int_{\mathcal{X}} \sum_{j=k+1}^\infty \left| \langle G(T_n(X)), \psi_j \rangle_{\mathcal{Y}} \right|^2 d\mu(X) \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \sum_{j=1}^k \left| \langle G_m(\mathbf{x}) - G(\mathbf{x}), \psi_j \rangle_{\mathcal{Y}} \right|^2 e^{-\|\mathbf{x}\|_2^2/2} d\mathbf{x} \\ &\quad + \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \sum_{j=k+1}^\infty \left| \langle G(\mathbf{x}), \psi_j \rangle_{\mathcal{Y}} \right|^2 e^{-\|\mathbf{x}\|_2^2/2} d\mathbf{x}. \end{aligned}$$

The first term on the right-hand side converges for fixed k to zero as $m \rightarrow \infty$ by construction of the G_m . The second term converges to zero as $k \rightarrow \infty$ because $G \in L^2(\mathbb{R}^n; \mathcal{Y})$. By a diagonal argument, we can thus choose $m = m(k)$ such that $\lim_{k \rightarrow \infty} F_{m(k),k} = F$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$.

Next, we consider the derivatives. By the chain rule,

$$\frac{\partial}{\partial \eta_i} F_{m,k}(X) = \begin{cases} b_i \lambda_i^{-1/2} \partial_i G_{m,k}(T_n(X)) & \text{if } 1 \leq i \leq n, \\ 0 & \text{otherwise,} \end{cases}$$

for every $X \in \mathcal{X}$. By a similar computation as above, after possibly adapting $m = m(k)$, we find that $\lim_{k \rightarrow \infty} (\partial_i G_{m(k),k}) \circ T_n = (\partial_i G) \circ T_n$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ for $i \in [n]$, as well as $\lim_{k \rightarrow \infty} \frac{\partial}{\partial \eta_i} F_{m,k} = 0$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ for $i > n$. This yields the claim. $\square$

Lemma C.13. *Let $F : \mathcal{X} \rightarrow \mathcal{Y}$ be an operator of the form $F = G \circ T_n$ with $G \in \text{Lip}(\mathbb{R}^n; \mathcal{Y})$ for some $n \in \mathbb{N}$ and T_n given by (C.12). Then, $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ with*

$$\frac{\partial}{\partial \eta_i} F(X) = \begin{cases} b_i \lambda_i^{-1/2} \partial_i G(T_n(X)) & \text{if } 1 \leq i \leq n, \\ 0 & \text{otherwise,} \end{cases}$$

for μ -a.e. $X \in \mathcal{X}$.

Proof Suppose that G is L -Lipschitz for some $L > 0$. It follows from the generalized Rademacher theorem that G is Hadamard differentiable $\mathcal{L}^n$ -almost everywhere in $\mathbb{R}^n$, see Aronszajn (1976, Thm. 1 in Ch. 2 & Rmk. 2 in Ch. 1). Since $\mathbb{R}^n$ is finite-dimensional, G is

in fact Fréchet differentiable $\mathcal{L}^n$ -almost everywhere—recall Appendix A. Moreover, the derivative is bounded by $\|DG(\mathbf{x})\|_{\text{HS}(\mathbb{R}^n; \mathcal{Y})} \leq \sqrt{n}L$ for $\mathcal{L}^n$ -a.e. $\mathbf{x} \in \mathbb{R}^n$. Consequently, $G \in W_{\text{loc}}^{1,\infty}(\mathbb{R}^n; \mathcal{Y})$.

Next, let $\omega \in C_c^\infty(\mathbb{R}^n)$ be a smooth bump function with $0 \leq \omega \leq 1$, $\omega = 1$ on $[-1, 1]^n$, and $\|\nabla \omega\|_\infty \leq C$ for some constant $C > 0$. For $m \in \mathbb{N}$, we set $\omega_m(\mathbf{x}) = \omega(\mathbf{x}/m)$ for $\mathbf{x} \in \mathbb{R}^n$. We further set $G_m := \omega_m G$ and $F_m := G_m \circ T_n$. Notice that $G_m \in W^{1,2}(\mathbb{R}^n; \mathcal{Y})$. Consequently, we find by Lemma C.12 that $F_m \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ with

$$\frac{\partial}{\partial \eta_i} F_m(X) = \begin{cases} b_i \lambda_i^{-1/2} \partial_i G_m(T_n(X)) & \text{if } 1 \leq i \leq n, \\ 0 & \text{otherwise,} \end{cases}$$

for μ -a.e. $X \in \mathcal{X}$. Using similar arguments as in the proof of Lemma C.6, we find that $\lim_{m \rightarrow \infty} F_m = F$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$. In addition, $\lim_{m \rightarrow \infty} \frac{\partial}{\partial \eta_i} F_m = b_i \lambda_i^{-1/2} (\partial_i G) \circ T_n$ for $i \in [n]$ and $\lim_{m \rightarrow \infty} \frac{\partial}{\partial \eta_i} F_m = 0$ for $i > n$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$. The claim now follows by Lemma C.7(ii). $\square$

D. Proof of Theorem 2.11

The argument follows along the lines of the proof of Lunardi et al. (2015/2016, Prop. 10.1.4) with several important modifications. Let $F : \mathcal{X} \rightarrow \mathcal{Y}$ be Lipschitz continuous with $L := [F]_{\text{Lip}(\mathcal{X}; \mathcal{Y})}$. The idea is to use Lemmas C.7 and C.13 to show that $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$.

By Lipschitz continuity, we have $\|F(X)\|_{\mathcal{Y}} \leq \|F(0)\|_{\mathcal{Y}} + L\|X\|_{\mathcal{X}}$ for every $X \in \mathcal{X}$, which implies that $F \in L_\mu^2(\mathcal{X}; \mathcal{Y})$ by Fernique's theorem (Theorem C.1). For $n \in \mathbb{N}$, we set $F_n := \mathbb{E}_n F$, as defined in (C.10). By Proposition C.9, we have $\lim_{n \rightarrow \infty} \mathbb{E}_n F = F$ in $L_\mu^2(\mathcal{X}; \mathcal{Y})$ and we can write $\mathbb{E}_n F(X) = V_n(T_n(X))$ for some $V_n : \mathbb{R}^n \rightarrow \mathcal{Y}$ and $T_n : \mathcal{X} \rightarrow \mathbb{R}^n$, $T_n(X) := (\hat{\xi}_1(X), \dots, \hat{\xi}_n(X))$ with $\xi_i, \hat{\xi}_i$ defined in (C.1), (C.2). Notice that V_n inherits Lipschitz continuity from F . Indeed, using Proposition C.9(i), we find for $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$

$$\begin{aligned} \|V_n(\mathbf{x} + \mathbf{z}) - V_n(\mathbf{x})\|_{\mathcal{Y}} &= \left\| \mathbb{E}_n F \left(\sum_{i=1}^n x_i \xi_i + \sum_{i=1}^n z_i \xi_i \right) - \mathbb{E}_n F \left(\sum_{i=1}^n x_i \xi_i \right) \right\|_{\mathcal{Y}} \\ &\leq \int_{\mathcal{X}} \left\| F \left(\sum_{i=1}^n x_i \xi_i + \sum_{i=1}^n z_i \xi_i + (\text{Id} - P_n)Z \right) - F \left(\sum_{i=1}^n x_i \xi_i + (\text{Id} - P_n)Z \right) \right\|_{\mathcal{Y}} d\mu(Z) \\ &\leq L \left\| \sum_{i=1}^n z_i \xi_i \right\|_{\mathcal{X}} = L \left\| \sum_{i=1}^n \sqrt{\lambda_i} z_i \phi_i \right\|_{\mathcal{X}} = L \left(\sum_{i=1}^n \lambda_i z_i^2 \right)^{1/2}. \end{aligned} \quad (\text{D.1})$$

Next, write $m = \dim(\mathcal{Y})$ and let $\{\psi_j\}_{j=1}^m$ be an orthonormal basis of $\mathcal{Y}$. For $j \in [m]$, we define the function(al)s

$$\begin{aligned} V_n^{(j)} : \mathbb{R}^n &\rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \langle V_n(\mathbf{x}), \psi_j \rangle_{\mathcal{Y}} \\ \mathbb{E}_n^{(j)} F : \mathcal{X} &\rightarrow \mathbb{R}, \quad X \mapsto \langle \mathbb{E}_n F(X), \psi_j \rangle_{\mathcal{Y}} = V_n^{(j)}(T_n(X)). \end{aligned}$$

We equip $\mathbb{R}^n$ with the rescaled Euclidean inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}_\lambda^n} := \sum_{i=1}^n \lambda_i x_i y_i$, which induces the norm $\|\mathbf{x}\|_{\mathbb{R}_\lambda^n} := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbb{R}_\lambda^n}}$. We denote the hereby defined space by $\mathbb{R}_\lambda^n$ and observe that

$\{\lambda_i^{-1/2} \mathbf{e}_i\}_{i=1}^n$ is an orthonormal basis. Recall that $\mathbf{e}_i$ denotes i th standard unit vector in $\mathbb{R}^n$. Now notice that (D.1) implies that $V_n^{(j)}$ is L -Lipschitz as a function $\mathbb{R}_{\mathbf{X}}^n \rightarrow \mathbb{R}$. We can thus apply Lemma C.13 with $\mathcal{Y} = \mathbb{R}$ (notice that it is independent of the inner product structure in $\mathbb{R}^n$). Consequently,

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

for μ -a.e. $X \in \mathcal{X}$.

Proof of (i). Suppose that $m \in \mathbb{N}$. For $\mathcal{L}^n$ -almost every $\mathbf{x} \in \mathbb{R}^n$, we have

$$\begin{aligned} \sum_{i=1}^n \lambda_i^{-1} |\partial_i V_n^{(j)}(\mathbf{x})|^2 &= \sum_{i=1}^n |DV_n^{(j)}(\mathbf{x})(\lambda_i^{-1/2} \mathbf{e}_i)|^2 \\ &= \|DV_n^{(j)}(\mathbf{x})\|_{\text{HS}(\mathbb{R}_{\mathbf{X}}^n; \mathbb{R})}^2 = \|DV_n^{(j)}(\mathbf{x})\|_{L(\mathbb{R}_{\mathbf{X}}^n; \mathbb{R})}^2 \leq L^2. \end{aligned}$$

Consequently, for μ -a.e. $X \in \mathcal{X}$,

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

By Lemma C.7(i), we may conclude that $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$.

Next, we show that $C^{0,1}(\mathcal{X}; \mathcal{Y})$ is continuously embedded in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. By (D.2), we know that $\|D_{\mathcal{X}_{\mathbf{b}}}(\mathbb{E}_n F)(X)\|_{L_{\mu}^2(\mathcal{X}; \text{HS}(\mathcal{X}_{\mathbf{b}}; \mathcal{Y}))} \leq \sqrt{m} \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})} L$. Moreover, it follows from Proposition C.9(i) that $\|\mathbb{E}_n F\|_{L_{\mu}^2(\mathcal{X}; \mathcal{Y})} \leq \sup_{X \in \mathcal{X}} \|F(X)\|_{\mathcal{Y}}$. In total, we have

$$\begin{aligned} \|\mathbb{E}_n F\|_{W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})} &\leq \left(\sup_{X \in \mathcal{X}} \|F(X)\|_{\mathcal{Y}}^2 + m \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})}^2 L^2 \right)^{1/2} \\ &\leq \max \left\{ 1, \sqrt{m} \|\mathbf{b}\|_{\ell^\infty(\mathbb{N})} \right\} \cdot \|F\|_{C^{0,1}(\mathcal{X}; \mathcal{Y})}. \end{aligned}$$

As $\lim_{n \rightarrow \infty} \mathbb{E}_n F = F$ in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ by Proposition C.10, the claim follows.

Proof of (ii). Suppose that $m = \infty$. Notice that in this case we cannot use the same argument as in the proof of (i), as (D.2) does not give a meaningful bound for $m = \infty$. We thus have to argue differently. To this end, assume that $\mathbf{b} \in \ell^2(\mathbb{N})$. From (D.1) it follows that $\|\partial_i V_n(\mathbf{x})\|_{\mathcal{Y}} \leq$

$L\sqrt{\lambda_i}$ for $\mathcal{L}^n$ -a.e. $\mathbf{x} \in \mathbb{R}^n$ and every $1 \leq i \leq n$. Consequently, for μ -a.e. $X \in \mathcal{X}$, we have

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

By Lemma C.7(i), we conclude that $F \in W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$. In order to show that $C^{0,1}(\mathcal{X}; \mathcal{Y})$ is continuously embedded in $W_{\mu, \mathbf{b}}^{1,2}(\mathcal{X}; \mathcal{Y})$ with the claimed bound, we can argue analogously as in the proof of (i). $\square$

E. Further proofs

For $d \in \mathbb{N} \cup \{\infty\}$, we write $\mu_d := \bigotimes_{i=1}^d \mathcal{N}(0, 1)$. In Sections E.2 and E.3 we will need the complementary error function $\text{erfc} : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$, which satisfies

$$\lim_{t \rightarrow \infty} t^m \text{erfc}(t) = 0, \quad \forall m \in \mathbb{N}. \quad (\text{E.1})$$

E.1. Proof of Proposition 3.10

As a preliminary step, we show that

$$\max_{\gamma \in \pi([s])} \|\gamma\|_{\ell^1(\mathbb{N})} \leq s-1 \leq s \quad \text{and} \quad \max_{\gamma \in \pi([s])} |\text{supp}(\gamma)| \leq s-1 \leq s. \quad (\text{E.2})$$

We start with the ℓ^1 -norm. Let $\gamma^* \in \pi([s])$ with $\|\gamma^*\|_{\ell^1(\mathbb{N})} = t \geq s$. By Assumption 2.7, we have $\lambda_{\mathbf{b}, i} \leq \lambda_{\mathbf{b}, 1}$ for every i . Hence,

$$u_{\gamma^*} = \left(1 + \sum_{i=1}^{\infty} \frac{\gamma_i^*}{\lambda_{\mathbf{b}, i}}\right)^{-1/2} \leq \left(1 + \frac{\|\gamma^*\|_{\ell^1(\mathbb{N})}}{\lambda_{\mathbf{b}, 1}}\right)^{-1/2} < \left(1 + \frac{j}{\lambda_{\mathbf{b}, 1}}\right)^{-1/2} = u_{j\mathbf{e}_1}, \quad \forall j = 0, \dots, s-1,$$

where $\mathbf{e}_1 = (\delta_{1,j})_{j=1}^{\infty} \in \Gamma$. Consequently, there exist at least s elements $\gamma \in \Gamma$ whose weights u_{γ} exceed u_{γ^*} . It follows that $\gamma^* \notin \pi([s])$, as required. The bound on the maximum support in (E.2) now follows from $|\text{supp}(\gamma)| \leq \|\gamma\|_{\ell^1(\mathbb{N})}$.

Suppose now that Assumption 3.9 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 $\tilde{F} := r^{-1} \|F\|_{C^{0,1}(\mathcal{X}; \mathcal{Y})}^{-1} F$, where $r := \max\{1, \sqrt{\dim(\mathcal{Y})} \cdot \|\mathbf{b}\|_{\ell^{\infty}(\mathbb{N})}\}$ if $\mathcal{Y}$ is finite-dimensional and $r := \max\{1, \|\mathbf{b}\|_{\ell^2(\mathbb{N})}\}$ if $\mathcal{Y}$ is infinite-dimensional. This implies $\tilde{F} \in B_{\mu, \mathbf{b}}^{\text{Lip}}(\mathcal{X}; \mathcal{Y})$ by Theorem 2.11. Next, let $\bar{s} = \bar{s}(F)$ be the constant in Assumption 3.9 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 $s := \lceil \epsilon^{-2/\beta} \rceil$ so that $s \geq \bar{s}$ and $\epsilon \leq 2^{\beta/2} s^{-\beta/2} \leq 2^{\beta/2} \epsilon$.

Recall from (2.3) that the truncated Wiener-Hermite PC expansion $\tilde{F}_{\pi([s])}$ is defined via the Hermite polynomials $H_{\pi(\mathbf{1}), d_1}, \dots, H_{\pi(\mathbf{s}), d_s}$ with $d_i := \max\{j : j \in \text{supp}(\pi(\mathbf{i}))\}$. Notice that we can interpret each $H_{\pi(\mathbf{i}), d_i}$ as a function $H_{\pi(\mathbf{i}), \infty}$ on $\mathbb{R}^{\mathbb{N}}$ (simply by ignoring all components x_i

with $i \notin \text{supp}(\gamma)$. Observe in addition that the set $\pi([s])$ is downward closed, i.e., if $\gamma \in \pi([s])$, then $\gamma' \in \pi([s])$ for every $\gamma' \leq \gamma$. Together with (E.2) we can thus directly apply Schwab and Zech (2023, Thm. 3.9) to conclude that there exists a ReLU neural network $\psi = (\psi_1, \dots, \psi_s) : \mathbb{R}^d \rightarrow \mathbb{R}^s$ with $d := \max\{d_1, \dots, d_s\}$ and with each ψ_i depending solely on the variables $x_1, \dots, x_d$ such that

$$\max_{i \in [s]} \|H_{\pi(i), \infty} - \psi_i\|_{L_{\mu_\infty}^2(\mathbb{R}^N)} \leq \epsilon^{(\beta+1)/\beta}. \quad (\text{E.3})$$

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

$$\text{size}(\psi) \leq C' s^6 \log(s) \log(\epsilon^{-(\beta+1)/\beta}) \leq 2^7 C' \left(1 + \frac{1}{\beta}\right) \epsilon^{-14/\beta-1} \quad (\text{E.4})$$

with some global constant $C' > 0$. We now define the encoder and decoder mappings

$$\begin{aligned} \tilde{\mathcal{E}}_{\mathcal{X}} : \mathcal{X} &\rightarrow \mathbb{R}^d, \quad \tilde{\mathcal{E}}_{\mathcal{X}}(X) := \left(\frac{\langle X, \phi_1 \rangle_{\mathcal{X}}}{\sqrt{\lambda_1}}, \dots, \frac{\langle X, \phi_d \rangle_{\mathcal{X}}}{\sqrt{\lambda_d}} \right), \\ \tilde{\mathcal{D}}_{\mathcal{Y}} : \mathbb{R}^s &\rightarrow \mathcal{Y}, \quad \tilde{\mathcal{D}}_{\mathcal{Y}}(x) := \sum_{i=1}^s \left(\int_{\mathcal{X}} \tilde{F} H_{\pi(i), \lambda} d\mu \right) x_i. \end{aligned}$$

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

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

By Assumption 3.9, we can bound $T_1(F)$ by $T_1(F) \leq c(F) s^{-\beta} \leq c(F) \epsilon^2 \leq c(F) \epsilon$, where $c(F)$ is the constant in (3.20). For $T_2(F)$, we find by an application of (E.3), the Cauchy-Schwarz inequality, and Parseval's identity that

$$\begin{aligned} T_2(F) &\leq \sum_{i=1}^s \left\| \int_{\mathcal{X}} \tilde{F} H_{\pi(i), \lambda} d\mu \right\|_{\mathcal{Y}} \|H_{\pi(i), \infty} - \psi_i\|_{L_{\mu_\infty}^2(\mathbb{R}^N)} \\ &\leq \epsilon^{(\beta+1)/\beta} s^{1/2} \left(\sum_{i=1}^s \left\| \int_{\mathcal{X}} \tilde{F} H_{\pi(i), \lambda} d\mu \right\|_{\mathcal{Y}}^2 \right)^{1/2} \leq \sqrt{2} \epsilon \|\tilde{F}\|_{L_{\mu}^2(\mathcal{X}; \mathcal{Y})} \leq \sqrt{2} \epsilon. \end{aligned}$$

Altogether, we conclude that $\|F - \Psi\|_{L_{\mu}^2(\mathcal{X}; \mathcal{Y})} \leq r \|F\|_{C^{0,1}(\mathcal{X}; \mathcal{Y})} (c(F) + \sqrt{2}) \epsilon$. Upon rescaling ϵ and $\bar{\epsilon}$ by the factor $r \|F\|_{C^{0,1}(\mathcal{X}; \mathcal{Y})} (c(F) + \sqrt{2})$, the claim follows in light of (E.4) with $\alpha = \beta/(14 + \beta)$. $\square$

E.2. Proof of Lemma 4.11

Let $\gamma \in \Gamma$ with $\text{supp}(\gamma) \subset [d]$ for some $d \in \mathbb{N}$. We first consider convergence in $L_{\mu}^2(\mathcal{X})$. To this end, recall the d -dimensional Hermite polynomials $H_{\gamma, d}$ defined in (2.1). By Fubini's theorem

and a change of variables, one has

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

It thus suffices to show that

$$\lim_{R \rightarrow \infty} \tilde{H}_{\gamma, R, d} = H_{\gamma, d} \quad \text{in } L^2_{\mu_d}(\mathbb{R}^d), \quad \forall d \in \mathbb{N}. \quad (\text{E.5})$$

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

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

where we used the notation $[-R, R]^c := \mathbb{R} \setminus [-R, R]$. By (E.1), we have $\lim_{R \rightarrow \infty} T_1(R) = 0$. Since $H_n \in L^2_{\mu_1}(\mathbb{R})$, the second term $T_2(R)$ converges to zero as $R \rightarrow \infty$ by the dominated convergence theorem. Next, let $d > 1$ and suppose that (E.5) holds for any $1 \leq d' < d$. Without loss of generality we may assume $R \geq 1$. For $\mathbf{x} \in \mathbb{R}^d$, let us write $\mathbf{x}_{[d-1]} = (x_1, \dots, x_{d-1}) \in \mathbb{R}^{d-1}$. Then, by Fubini's theorem, we have

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

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

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

By induction hypothesis for $d' = d - 1$, the term $\tilde{H}_{\gamma, R, d-1}$ converges in $L^2_{\mu_{d-1}}(\mathbb{R}^{d-1})$ as $R \rightarrow \infty$. Hence, the supremum over $R \geq 1$ of the first integral is finite. Applying the induction hypotheses for $d' = 1$, we conclude that the second integral over $\mathbb{R}$ converges to zero as $R \rightarrow \infty$. A similar argument shows that $\lim_{R \rightarrow \infty} t_2(R) = 0$. This completes the proof of (E.5).

Next, we consider convergence of $\nabla_{\mathcal{X}_b} \tilde{H}_{\gamma, R, \lambda}$ in $L^2_{\mu}(\mathcal{X}; \mathcal{X}_b)$. For this, recall the basis $\{\eta_i\}_{i=1}^{\infty}$ of $\mathcal{X}_b$ defined in (2.6). Note that the capped Hermite polynomials $\tilde{H}_{n, R}$, as defined in (4.6),

are Lipschitz continuous. We write $x_i := \lambda_i^{-1/2} \langle X, \phi_i \rangle_{\mathcal{X}}$ for $X \in \mathcal{X}$ and $i \in [d]$. By Lemma C.13 with $\mathcal{Y} = \mathbb{R}$, we have $\tilde{H}_{\gamma,R,\lambda} \in W_{\mu,\mathbf{b}}^{1,2}(\mathcal{X})$ and for μ -a.e. $X \in \mathcal{X}$ and $i \in [d]$ the weak partial derivatives are given by

$$\begin{aligned} \frac{\partial}{\partial \eta_i} \tilde{H}_{\gamma,R,\lambda}(X) &= b_i \lambda_i^{-1/2} \partial_i \tilde{H}_{\gamma,R,d}(x_1, \dots, x_d) \\ &= \begin{cases} b_i \lambda_i^{-1/2} \partial_i H_{\gamma_i}(x_i) \prod_{\substack{j=1 \\ j \neq i}}^d \tilde{H}_{\gamma_j,R}(x_j) & \text{if } x_i \in [-R, R], \\ 0 & \text{if } x_i \in [-R, R]^c. \end{cases} \end{aligned}$$

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

$$\begin{aligned} \left\| \nabla_{\mathcal{X}_{\mathbf{b}}} \tilde{H}_{\gamma,R,\lambda} - \nabla_{\mathcal{X}_{\mathbf{b}}} H_{\gamma,\lambda} \right\|_{L_{\mu}^2(\mathcal{X}; \mathcal{X}_{\mathbf{b}})}^2 &= \int_{\mathcal{X}} \sum_{i=1}^d \left| \frac{\partial}{\partial \eta_i} \tilde{H}_{\gamma,R,\lambda}(X) - \frac{\partial}{\partial \eta_i} H_{\gamma,\lambda}(X) \right|^2 d\mu(X) \\ &= \sum_{i=1}^d \int_{\mathbb{R}^{i-1} \times [-R, R] \times \mathbb{R}^{d-i}} b_i^2 \lambda_i^{-1} |\partial_i H_{\gamma_i}(x_i)|^2 \left| \prod_{\substack{j=1 \\ j \neq i}}^d \tilde{H}_{\gamma_j,R}(x_j) - \prod_{\substack{j=1 \\ j \neq i}}^d H_{\gamma_j}(x_j) \right|^2 d\mu_d(\mathbf{x}) \\ &\quad + \sum_{i=1}^d \int_{\mathbb{R}^{i-1} \times [-R, R]^c \times \mathbb{R}^{d-i}} b_i^2 \lambda_i^{-1} |\partial_i H_{\gamma,d}(\mathbf{x})|^2 d\mu_d(\mathbf{x}). \end{aligned}$$

Since $\partial_i H_{\gamma_i} \in L_{\mu_1}^2(\mathbb{R})$, the first term on the right-hand side converges to zero as $R \rightarrow \infty$ by (E.5) and the dominated convergence theorem. Similarly, the second term converges to zero as $R \rightarrow \infty$ by the dominated convergence theorem. $\square$

E.3. Proof of Lemma 4.12

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

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

Note that $H_n H_m$ is a polynomial of order $n + m$. Hence, by (E.1), we deduce $\lim_{R \rightarrow \infty} T_2(R) = 0$ and $\lim_{R \rightarrow \infty} T_3(R) = 0$. Moreover, the dominated convergence theorem and orthonormality of

the Hermite polynomials imply that

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

where $\delta_{n,m}$ denotes the Kronecker delta. Altogether,

$$\lim_{R \rightarrow \infty} \left\langle \tilde{H}_{n,R}, \tilde{H}_{m,R} \right\rangle_{L^2_{\mu_1}(\mathbb{R})} = \delta_{n,m}. \quad (\text{E.6})$$

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

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

we conclude by (E.6) that

$$\lim_{R \rightarrow \infty} \left\langle \tilde{H}_{\gamma,R,\lambda}, \tilde{H}_{\gamma',R,\lambda} \right\rangle_{L^2_{\mu}(\mathcal{X})} = \delta_{\gamma,\gamma'} := \prod_{i=1}^{\infty} \delta_{\gamma_i,\gamma'_i}.$$

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

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

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

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

Similarly, we have

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

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