Bayesian inference for Gaussian models: Inverse problems and evolution equations

Yan, D.

Author: Yan, D.
Title: Bayesian inference for Gaussian models: Inverse problems and evolution equations
Issue Date: 2020-03-03

Chapter 2

Smoothness Class

Smoothness, also known as *regularity*, characterises how 'well' a function behaves. A set of function spaces, each of which consists of functions with the same smoothness property, is called a *smoothness class*. We introduce a scale of smoothness classes, i.e. smoothness classes that are indexed by an ordered set, that is particularly suitable for studying Gaussian linear models in Hilbert space framework.

While most of the statements in this chapter are given in a general Hilbert space setting, they can always be translated to a concrete L^2 function space $L^2(\mathfrak{D}, \mu)$, square integrable functions on a domain $\mathfrak{D}$ with respect to measure μ. When the underlying domain $\mathfrak{D}$ is one-dimensional, the smoothness is naturally considered to be a scalar. In the higher dimensional case, i.e. the dimension of $\mathfrak{D}$ being $d > 1$, a function $f \in L^2(\mathfrak{D}, \mu)$ does not necessarily behave identically along each coordinate direction. Hence, the scalar smoothness should as well adapt to the multi-dimensional nature. In this chapter, we first present the *isotropic* scale, that is a class of spaces parametrised by scalar-valued indices. Subsequently, the concept is generalised to the *anisotropic* scale, with $\mathbb{R}^d$-valued indices. Studied later in Part II, the inverse problems are placed in isotropic scales. In contrast, for evolution equations examined in Part III, it is more adequate to consider the spatial and temporal regularity of functions separately, and as a consequence, anisotropic scales are used.

In this chapter, we define smoothness scales, introduce important examples that will be used as underlying parameter spaces for the later statistical study, and examine their properties, especially those important for establishing approximation error estimate.

2.1 Smoothness scales

The parameter θ of interest in the Gaussian linear model (I.1) is assumed to be an element of a Hilbert space H. We embed this space as the space $H = H_0$ in a 'scale of smoothness classes', defined as follows.

Definition 2.1 (Smoothness scale)**.** For every $s \in \mathbb{R}$ the space H_s is an infinite-dimensional, separable Hilbert space, with inner product $\langle \cdot, \cdot \rangle_s$ and induced norm $\|\cdot\|_s$. The spaces $(H_s)_{s\in\mathbb{R}}$ satisfy the following conditions:

(i) For $s < t$ the space H_t is a dense subspace of H_s and $\|f\|_s \lesssim \|f\|_t$, for $f \in H_t$.

(ii) For $s \geq 0$ and $f \in H_0$ viewed as element of $H_{-s} \supset H_0$,

$$\|f\|_{-s} = \sup_{\|g\|_s \leq 1} \langle f, g \rangle_0, \qquad f \in H_0. \tag{2.1}$$

The notion of scales of smoothness classes is standard in the literature. In the preceding definition we have stripped it to the bare essentials needed in our general result on posterior contraction. Concrete examples, as well as more involved structures such as Hilbert scales, are introduced in the subsequent section.

Remark 2.2 (Norm duality)**.** The *norm duality* (2.1) is implied if, for $s > 0$, the space H_{-s} can be identified with the dual space H_s^* of H_s and the embedding $\iota : H_0 \to H_{-s}$ is the adjoint of the embedding $\iota : H_s \to H_0$, after the usual identification of H_0 and its dual space H_0^*. (The three nested spaces $H_{-s} \supset H_0 \supset H_s$ then form a 'Gelfand triple'.) Indeed, by definition the image $\iota^* f$ of $f \in H_0 = H_0^*$ under the adjoint $\iota^* : H_0^* \to H_s^*$ is the map $g \mapsto (\iota^* f)(g) = \langle \iota g, f \rangle_0 = \langle g, f \rangle_0$ from $H_s \to \mathbb{R}$. The norm of this map as an element of H_s^* is $\sup_{\|g\|_s \leq 1} (\iota^* f)(g)$. The norm duality follows if $\iota^* f$ is identified with the element $f \in H_0 \subset H_{-s}$.

Since every H_s is a Hilbert space, one can also identify H_s^* with itself in the usual way, but this involves the inner product in H_s, and is different from the identification of H_s^* with the 'bigger space' H_{-s}.

We assume that the smoothness scale allows good finite-dimensional approximations, as in the following condition.

Assumption 2.3 (Approximation property)**.** For every $j \in \mathbb{N}$ and $s \in (0, S)$, for some $S > 0$, there exists a $(j-1)$-dimensional linear subspace $V_j \subset H_0$ and a number $\delta(j, s)$ such that $\delta(j, s) \to 0$ as $j \to \infty$, and such that

$$\inf_{g \in V_j} \|f - g\|_0 \lesssim \delta(j, s) \, \|f\|_s, \tag{2.2}$$

$$\|g\|_s \lesssim \frac{1}{\delta(j, s)} \, \|g\|_0, \qquad \forall g \in V_j. \tag{2.3}$$

This assumption is also common in the literature on numerical analysis, approximation theory, and inverse problems, etc. The two inequalities (2.2) and (2.3) are known as of Jackson and Bernstein type, respectively, see, e.g., [12]. The approximation property (2.2) shows that 'smooth elements' $f \in H_s$ are well approximated in $\|\cdot\|_0$ by their projection onto a finite-dimensional space V_j, with approximation error tending to zero as the dimension of V_j tends to infinity. Naturally one expects the numbers $\delta(j, s)$ that control the approximation to be decreasing in both j and s. In our examples we shall mostly have polynomial dependence $\delta(j, s) = j^{-s/d}$, in the case that H_0 consists of functions on a d-dimensional domain. The stability property (2.3) quantifies the smoothness norm of the projections in terms of the approximation numbers. Both conditions are assumed up to a maximal order of smoothness $S > 0$, and it follows from (2.3) that V_j must be contained in the space H_S.

The following estimates derived from the approximation property are convenient for the later study.

Lemma 2.4. *If V_j is a finite-dimensional space as in Assumption 2.3 such that (2.2) and (2.3) hold, then, for $P_j : H_0 \to V_j$ the orthogonal projection onto V_j, and $0 \leq s, t < S$,*

$$\|f - P_j f\|_{-t} \lesssim \delta(j,t)\delta(j,s)\|f\|_s, \qquad f \in H_0, \tag{2.4}$$

$$\|g\|_s \lesssim \frac{1}{\delta(j,s)\delta(j,t)}\|g\|_{-t}, \qquad g \in V_j. \tag{2.5}$$

Proof. By the dual norm relation in (ii) of Definition 2.1, and the orthogonality of $f - P_j f$ to V_j,

$$\begin{aligned}\|f - P_j f\|_{-t} &= \sup_{\|g\|_t \leq 1} \langle f - P_j f, g\rangle_0 = \sup_{\|g\|_t \leq 1} \langle f - P_j f, g - P_j g\rangle_0 \\ &\leq \|f - P_j f\|_0 \sup_{\|g\|_t \leq 1} \|g - P_j g\|_0,\end{aligned}$$

by the Cauchy-Schwarz inequality. Here $\|f - P_j f\|_0 \lesssim \delta(j,s)\|f\|_s$ and $\|g - P_j g\|_0 \lesssim \delta(j,t)\|g\|_t$, both by (2.2). Inequality (2.4) follows.

For the second inequality we have, for $g \in V_j$,

$$\|g\|_0 = \sup_{f \in V_j : \|f\|_0 \leq 1} \langle g, f\rangle_0 \lesssim \sup_{f \in V_j : \|f\|_0 \leq 1} \|g\|_{-t}\|f\|_t,$$

again by the dual norm relation. Here we can bound $\|f\|_t$ by $\|f\|_0/\delta(j,t)$, with the help of (2.3). We obtain (2.5) by first bounding $\|g\|_s$ with the help of (2.3) and next using the preceding display. □

The approximation property (2.2) can also be stated in terms of the 'approximation numbers' of the canonical embedding $\iota : H_s \to H_0$. The jth *approximation number* of a general bounded linear operator $T : G \to H$ between normed spaces is defined as

$$a_j(T : G \to H) = \inf_{U:\mathrm{Rank}U<j} \sup_{f:\|f\|_G \leq 1} \|(T - U)f\|_H, \tag{2.6}$$

where the infimum is taken over all linear operators $U : G \to H$ of rank less than j. It is immediate from the definitions that the numbers $\delta(j,s)$ in (2.2) can be taken equal to the approximation numbers $a_j(\iota : H_s \to H_0)$. The set of approximation numbers $a_j(\iota : H_{s+t} \to H_t)$ of the canonical embedding describes many characteristics of the smoothness scale $(H_s)_{s\in\mathbb{R}}$. In particular, Assumption 2.3 implies that the canonical embedding $\iota : H_s \to H_0$ is a limit of a sequence of finite rank operators, and hence is compact. Later we give a brief discussion in Section 2.4.

Example 2.5 (Sobolev classes)**.** The most important examples of smoothness classes satisfying Definition 2.1 are fractional Sobolev spaces on a bounded domain $\mathfrak{D} \subset \mathbb{R}^d$. For a natural number $s \in \mathbb{N}$ the Sobolev space of order s can be defined by

$$H_s(\mathfrak{D}) = W^{s,2}(\mathfrak{D}) := \Big\{f \in \mathscr{D}'(\mathfrak{D}) : \|f\|_s := \sum_{|\alpha|\leq s} \|D^\alpha f\|_{L^2(\mathfrak{D})} < \infty\Big\}.$$

Here $\mathscr{D}'(\mathfrak{D})$ is the space of generalized functions on $\mathfrak{D}$ (distributions), i.e. the topological dual space of the space $C_c^\infty(\mathfrak{D})$ of infinitely differentiable functions with compact support in $\mathfrak{D}$; the sum ranges over the multi-indices $\alpha = (\alpha_1, \cdots, \alpha_d) \in (\{0\} \cup \mathbb{N})^d$ with $|\alpha|:= \sum_{i=1}^s \alpha_i \leq s$; and D^α is the differential operator

$$D^\alpha := \frac{\partial^{\alpha_1}\partial^{\alpha_2}\cdots\partial^{\alpha_d}}{\partial x_1^{\alpha_1} x_2^{\alpha_2}\cdots \partial x_d^{\alpha_d}}.$$

The definition can be extended to $s \in \mathbb{R}\backslash\mathbb{N}$ in several ways. All constructions are equivalent to the Besov space $B^s_{2,2}(\mathfrak{D})$, see [92, 93].

It is well known that the approximation numbers of the scale of Sobolev spaces satisfy Assumption 2.3 with $\delta(j,t) = j^{-t/d}$, see [45].

Example 2.6 (Sequence spaces)**.** Suppose $(\phi_i)_{i\in\mathbb{N}}$ is a given orthonormal sequence in a given Hilbert space H, and $1 \leq b_i \uparrow \infty$ is a given sequence of numbers. For $s \geq 0$, define H_s as the set of all elements $f = \sum_{i\in\mathbb{N}} f_i\phi_i \in H$ with $\sum_{i\in\mathbb{N}} b_i^{2s} f_i^2 < \infty$, equipped with the norm

$$\|f\|_s = \Big(\sum_{i\in\mathbb{N}} b_i^{2s} f_i^2\Big)^{1/2}.$$

Then $H_0 = H$ is embedded in H_s, for every $s > 0$, and the norms $\|f\|_s$ are increasing in s. Every space H_s is a Hilbert space; in fact H_s is isometric to H_0 under the map $(f_i) \to (f_i b_i^s)$, where we have identified the series with their coefficients for simplicity of notation.

For $s < 0$, we equip the elements $f = \sum_{i\in\mathbb{N}} f_i\phi_i$ of H, where $(f_i) \in \ell_2$, with the norm as in the display, which is now automatically finite, and next define H_s as the metric completion of H under this norm. The space H_s is isometric to the set of all sequences $(f_i)_{i\in\mathbb{N}}$ with $\sum_{i\in\mathbb{N}} f_i^2 b_i^{2s} < \infty$ equipped with the norm given on the right hand side of the preceding display, but the series $\sum_{i\in\mathbb{N}} f_i\phi_i$ may not possess a concrete meaning, for instance as a function if H is a function space.

By Parseval's identity the inner product on $H = H_0$ is given by $\langle f, g\rangle_0 = \sum_{i\in\mathbb{N}} f_i g_i$, and the norm duality (2.1) follows with the help of the Cauchy-Schwarz inequality.

The natural approximation spaces for use in Assumption 2.3 are $V_j = \mathsf{Span}(\phi_i : i < j)$. Inequalities are satisfied with the approximation numbers taken equal to $\delta(j,t) = b_j^{-t}$.

2.2 Hilbert Scale

As we are going to show in this section, two Hilbert spaces such that one is dense in the other, naturally define a smoothness scale with additional structure, which is called Hilbert scale. In many applications, the smoothness scales turn out to be Hilbert scales.

A *Hilbert scale* is generated by a *densely defined unbounded self-adjoint strictly positive* operator $\Lambda : \mathrm{Dom}(\Lambda) \subset H_0 \to H_0$, with domain $\mathrm{Dom}(\Lambda)$ such that

(a) $\mathrm{Dom}(\Lambda)$ is dense in H_0, (i.e. 'Λ is densely defined'),

(b) $\mathrm{Dom}(\Lambda) = \mathrm{Dom}(\Lambda^*)$,

(c) $\langle \Lambda x, y\rangle = \langle x, \Lambda y\rangle$ for all $x, y \in \mathrm{Dom}(\Lambda)$, (i.e. '$\Lambda$ is symmetric'),

(d) $\langle \Lambda x, x\rangle \geq \kappa \|x\|^2$, for all $x \in \mathrm{Dom}(\Lambda)$, and some $\kappa > 0$.

The set $\mathrm{Dom}(\Lambda^*)$ in (b) is the domain of the adjoint Λ^* of Λ, which is defined as the set of all $y \in H$ such that the map $x \mapsto \langle \Lambda x, y\rangle$ from $\mathrm{Dom}(\Lambda)$ to $\mathbb{R}$ is continuous (see Appendix A). Note that this depends on the domain $\mathrm{Dom}(L)$, which is considered part of the definition of Λ and is restricted by (a) only. Together, requirements (b) and (c) are equivalent to the requirement that Λ be *self-adjoint*.

The domain of the k-th power of the operator Λ is defined, by induction for $k = 2, 3, \ldots$, as (with $\Lambda^1 = \Lambda$)

$$\mathrm{Dom}(\Lambda^k) = \{f \in \mathrm{Dom}(\Lambda^{k-1}) : \Lambda f \in \mathrm{Dom}(\Lambda)\}, \qquad k > 1.$$

All powers Λ^k, for $k \in \mathbb{N}$, are defined on

$$H_\infty := \bigcap_{k\in\mathbb{N}} \mathrm{Dom}(\Lambda^k). \tag{2.7}$$

It can be shown that H_∞ is dense in H_0 (Lemma 8.17 in [29]). Next, using spectral theory, fractional powers Λ^s can be defined as well on the domain H_∞, for every $s \in \mathbb{R}$, through integration with respect to the spectral family (E_λ) of Λ, i.e.

$$\Lambda^s := \int_{\mathbb{R}} \lambda^s \, dE_\lambda = \int_\kappa^\infty \lambda^s \, dE_\lambda.$$

This allows to define an inner product on H_∞ by, for $h, g \in H_\infty$ and $s \in \mathbb{R}$,

$$\langle h, g\rangle_s := \langle \Lambda^s h, \Lambda^s g\rangle. \tag{2.8}$$

Definition 2.7 (Hilbert scale)**.** The Hilbert space H_s is the completion of H_∞ with respect to the norm induced by the inner product $\langle\cdot,\cdot\rangle_s$ defined in (2.8). The family $(H_s)_{s\in\mathbb{R}}$ is called the *Hilbert scale generated by* Λ.

The following proposition, adapted from Proposition 8.19 in [29], lists basic properties of Hilbert scales.

Proposition 2.8. *Let Λ be a densely defined unbounded operator satisfying (a)–(d). Then the Hilbert scale $(H_s)_{s\in\mathbb{R}}$ is a smoothness scale in the sense of Definition 2.1, with $\|f\|_s \leq \kappa^{s-t}\|f\|_t$, for $f \in H_t$, and $s < t$.*

In addition, a Hilbert scale possesses the following properties.

(i) If $s \geq 0$, then $H_s = \mathrm{Dom}(\Lambda^s)$, and H_{-s} is the dual space of H_s, i.e.

$$H_{-s} = (H_s)^*.$$

(ii) The following interpolation inequality *holds. If $f \in H_t$,*

$$\|f\|_s \leq \|f\|_r^\lambda \|f\|_t^{1-\lambda}, \quad \textit{with } \lambda = (t-s)/(t-r), \tag{2.9}$$

for $-\infty < r < s < t < \infty$.

Furthermore, for any $s, t \in \mathbb{R}$ the operator Λ^{t-s} has a unique extension from H_∞ to a bounded, self-adjoint operator $\Lambda^{t-s} : H_t \to H_s$, satisfying

(iii) $\|\Lambda^{t-s} f\|_s \simeq \|f\|_t$, *for* $f \in H_t$.

(iv) $\Lambda^{t-s} = \Lambda^t \Lambda^{-s}$.

(v) $(\Lambda^s)^{-1} = \Lambda^{-s}$.

Somewhat abusing notation, we have denoted the extension of Λ^{t-s} in the proposition using the same symbol Λ^{s-t}. Taking $s = 0$ or $t = 0$, we see that $\Lambda^s : H_s \to H_0$ and $\Lambda^s : H_0 \to H_{-s}$ are norm isomorphisms, for every $s \in \mathbb{R}$. In particular, the unbounded densely defined operator $\Lambda : D(\Lambda) \subset H_0 \to H_0$ that generates the scale can be extended to a bounded operator $\Lambda : H_1 \to H_0$, by strengthening the norm on its domain, and also to a bounded operator $\Lambda : H_0 \to H_{-1}$, by extending its range space and weakening the norm of its range space. Moreover, the inverse map is a norm isomorphism $\Lambda^{-1} : H_0 \to H_1$, and hence is certainly bounded as an operator $\Lambda^{-1} : H_0 \to H_0$.

The eigenvalues of Λ^{-1} are closely connected to the approximation property in Assumption 2.3.

Proposition 2.9. *If $\Lambda^{-1} : H_0 \to H_0$ is compact with eigenvalues $\lambda_j \downarrow 0$, then Assumption 2.3 is satisfied in the Hilbert scale $(H_s)_{s\in\mathbb{R}}$ generated by Λ, with $\delta(j, t) \simeq \lambda_j^t$ and $S = \infty$. In fact, there exist linear spaces V_j of dimension $j - 1$ such that, for $s \geq 0$ and $t \in \mathbb{R}$,*

$$\inf_{g\in V_j} \|f - g\|_t \lesssim \delta(j, s)\, \|f\|_{s+t}, \tag{2.10}$$

$$\|g\|_{s+t} \lesssim \frac{1}{\delta(j, s)}\, \|g\|_t, \qquad \forall g \in V_j. \tag{2.11}$$

Proof. Because $\Lambda^{-1} : H_0 \to H_0$ is compact, there exists an orthonormal basis $(\phi_i)_{i\in\mathbb{N}}$ of eigenfunctions in H_0. It may be checked that $f = \sum_{i\in\mathbb{N}} f_i\phi_i$ has $L^s f = \sum_{i\in\mathbb{N}} f_i\lambda_i^{-s}\phi_i$, and square norm $\|f\|_s^2 = \sum_{i\in\mathbb{N}} f_i^2\lambda_i^{-2s}$, provided the latter series converges. Take V_j equal to the linear span of the first $j-1$ eigenfunctions. Then $f - P_j f = \sum_{i\geq j} f_i\phi_i$ and hence $\|f - P_j f\|_t^2 = \sum_{i\geq j} f_i^2\lambda_i^{-2t} \leq \lambda_j^{2s} \sum_{i\geq j} f_i^2\lambda_i^{-2t-2s} \leq \lambda_j^{2s}\|f\|_{s+t}^2$, for $s, t \geq 0$, and for $f \in V_j$ we have $\|f\|_{s+t}^2 = \sum_{i<j} f_i^2\lambda_i^{-2s-2t} \leq \lambda_j^{-2s} \sum_{i\leq j} f_i^2\lambda_i^{-2t} = \lambda_j^{-2s}\|f\|_t^2$. □

The sequence spaces of Example 2.6 are one class of examples of Hilbert scales, generated by the operator $L : (f_i) \mapsto (f_i b_i)$. More intricate Hilbert scales arise from (elliptic) differential operators. These are useful in that they can incorporate boundary conditions, which are then automatically inherited by a Gaussian prior attached to such a scale. The following one-dimensional example is simplistic, but illustrative.

Example 2.10 (Sobolev scales)**.** Consider the one-dimensional negative Laplacian

$$-\Delta = -\frac{d^2}{dx^2}$$

as an operator on the space $C_c^\infty(0,1)$ of infinitely often differentiable functions with compact support in $(0,1)$, viewed as subset of $L^2(0,1)$, with range space $L^2(0,1)$. On this domain this operator is not self-adjoint, but it has a self-adjoint extension (with differentiation interpreted in the sense of distributions) to the space of all functions $f \in W^{2,2}(0,1)$ satisfying the *Dirichlet boundary condition*

$$f(0) = 0 = f(1). \tag{2.12}$$

(See Theorem 4.23 in [41].) The eigenfunctions of the Laplacian under the Dirichlet boundary condition are the functions $x \mapsto \sin(j\pi x)$, for $j \in \mathbb{N}$, with eigenvalues of the order $b_j \asymp j^{-1}$. The corresponding Hilbert scale can also be described as the sequence space generated by this orthogonal basis.

Because the Laplacian is a second derivative it is natural to half the scale parameter, or equivalently use the root negative Laplacian $\Lambda := \sqrt{-\Delta}$ as the generator of the scale (where the root is defined through the spectral decomposition).

The boundary conditions play an important role in defining the scale. Technically they are needed to create a domain on which the operator is self-adjoint. An alternative choice to the Dirichlet is the *Cauchy boundary condition*

$$f'(0) = 0 = f(1).$$

This leads to the sequence scale generated by the eigenfunctions $x \mapsto \cos((j - 1/2)\pi x)$, for $j \in \mathbb{N}$, and is different from the Dirichlet scale. Again the eigenvalues of Λ^{-1} are of the order j^{-1}.

Incidentally, it is shown in [73] that the full Sobolev scale ($s \in \mathbb{R}$) of Example 2.5 is not a Hilbert scale for any generating operator Λ. Also in that sense the boundary conditions are essential.

Until now, the Hilbert scales have been constructed from H_0 with the help of a generating operator $(\Lambda, \mathrm{Dom}(\Lambda))$. Alternatively, given a dense subset G of H, the existence of a generating operator Λ with $\mathrm{Dom}\,\Lambda = G$ is guaranteed, as shown in the following theorem.

Theorem 2.11. *Let G and H be two Hilbert spaces such that G is densely and continuously embedded into H such that $\|g\|_H \leq \|g\|_G$ for all $g \in G$. Then there exists a unique operator Λ, which is positive-definite and self-adjoint, and generates a Hilbert scale $(H_s)_{s\in\mathbb{R}}$ such that $H_1 = G$ and $H_0 = H$.*

Proof. The Λ is constructed using Lemma A.8. The rest is to check that Λ satisfies the properties of a generating operator. ☐

In Proposition 2.8, a natural duality structure is possessed by Hilbert scales, i.e. $H_{-s} = (H_s)^*$, given $s \geq 0$. Now we are going to have a closer look at the norm duality mentioned in Remark 2.2 and its connection to Hilbert scales. First let us recall a noted duality structure.

Definition 2.12. Let G and H be two Hilbert spaces such that G is a dense subset of H and the canonical embedding $G \hookrightarrow H$ is continuous. The triplet

$$G \subset H = H^* \subset G^*$$

is a *Gelfand triple*, where the subsets are dense in all bigger spaces.

Remark 2.13. The Gelfand triple is a well-defined object, see Lemma A.7. Its general version is that the *pivot space* H is a Hilbert space and the dense space G is a reflexive Banach space, but for this thesis it is sufficient to only work with Hilbert spaces.

We further elaborate on how to identify dual spaces with the spaces with negative indices in Hilbert scales.

Lemma 2.14. *Let $\{H_s\}_{s\in\mathbb{R}}$ be a Hilbert scale. For any fixed $t > 0$ and $f \in H_{-t}$, the mapping,*

$$\mathcal{J} : H_{-t} \to (H_t)^*,$$
$$(\mathcal{J}f)(g) := \langle \Lambda^{-2t} f, g\rangle_{H_0}, \tag{2.13}$$

with the generating operator Λ of $\{H_s\}_{s\in\mathbb{R}}$, is an isometric isomorphism between H_{-t} and $(H_t)^*$.

Proof. It is convenient to recall that $\Lambda^t : H_{s+t} \to H_s$ is an isometric isomorphism.

For $f \in H_{-t}$, the map $g \mapsto \langle f, g\rangle_{H_{-t}}$ is a bounded linear functional on H_t, and therefore by the Riesz representation theorem, there is a unique $g_f \in H_t$ such that for all $g \in H_t$,

$$\langle f, g\rangle_{H_{-t}} = \langle \Lambda^{-t} f, \Lambda^{-t} g\rangle_{H_0} = \langle \Lambda^{-2t} f, g\rangle_{H_0} = \langle \Lambda^{t}\Lambda^{-4t} f, \Lambda^{t} g\rangle_{H_0} = \langle g_f, g\rangle_{H_t},$$

where $g_f = \Lambda^{-4t} f$. Consequently, for any $f \in H_{-t}$ and $g \in H_t$,

$$(\mathcal{J}f)(g) = \langle \Lambda^{-4t} f, g\rangle_{H_t} = \langle \Lambda^{-2t} f, g\rangle_{H_0}.$$

Conversely, if $\ell \in H_t^*$ is a bounded linear functional on H_t, then again by the Riesz representation theorem, there is a unique $h_\ell \in H_t$ such that

$$\ell(h) = \langle h_\ell, h\rangle_{H_t} = \langle \Lambda^{2t} f_\ell, h\rangle_{H_0},$$

where $f_\ell = \Lambda^{2t} h_\ell \in H_{-t}$. Furthermore, since $\|f_\ell\|_{H_{-t}} = \|h_\ell\|_{H_t}$,

$$|\mathcal{J}(f_\ell)(h)| = |\ell(h)| \leq \|h_\ell\|_{H_t}\|h\|_{H_t} = \|f_\ell\|_{H_{-t}}\|h\|_{H_t},$$

which implies $\|\ell\|_{H_t^*} \leq \|f_\ell\|_{H_{-t}}$. On the other hand,

$$\ell(\Lambda^{-2t} f_\ell) = \langle h_\ell, \Lambda^{-2t} f_\ell\rangle_{H_t} = \langle \Lambda^{-t} f_\ell, \Lambda^{-t} f_\ell\rangle_{H_0} = \|f_\ell\|^2_{H_{-t}}.$$

Because $\|f_l\|_{H_{-t}} = \|h_l\|_{H_t}$, the preceding equation implies $\|\ell\| \geq \|f_\ell\|_{H_{-t}}$.

Combining the results above, we conclude that $\mathcal{J} : H_{-t} \to (H_t)^*$ is an isometric isomorphism. □

The connection between Gelfand triples and Hilbert scales is stated in the following result.

Theorem 2.15. *Let $(H_s)_{s\in\mathbb{R}}$ be a Hilbert scale. Then, (H_{s+t}, H_s, H_{s-t}) is a Gelfand triple. Conversely, for any Gelfand triple (G, H, G^*), there exists a unique Hilbert scale $(H_s)_{s\in\mathbb{R}}$ such that $H_1 = G, H_0 = H$, and $H_{-1} = G^*$.*

Proof. Identify the dual space $(H_s)^*$ of H_s with itself. the first statement follows the same argument from Lemma 2.14 with

$$\mathcal{J} : H_{s-t} \to (H_{s+t})^*,$$
$$(\mathcal{J} f)(g) := \langle \Lambda^{-2t} f, g \rangle_{H_s}.$$

The second part is a corollary of Theorem 2.11. □

2.2.1 Relation to Boundary Conditions

Boundary conditions play an important role in the formulation of multi-dimensional problems, and Hilbert scales naturally cope with this issue. While for functions with domain an interval of the real line a boundary condition just concerns the values at the two endpoints of the interval, on multi-dimensional domains boundary conditions are a subtle issue. In the latter case the boundary is itself a continuous, possibly multi-dimensional, domain, and the boundary condition will involve a space of functions defined on the boundary, an infinite-dimensional space. Generally speaking, Hilbert scales are useful in the sense that the functions in the Hilbert scale automatically satisfy the boundary condition if L is chosen properly. It is tightly connected to the Hilbert (L^2) theory of elliptic equations.

To see this in more detail, consider the case of a partial differential equation

$$\mathcal{L}u = f, \tag{2.14}$$

where $\mathcal{L}$ is a second order elliptic differential operator, and $u, f : \mathfrak{D} \to \mathbb{R}$ are functions on a bounded domain $\mathfrak{D} \subset \mathbb{R}^d$ with sufficiently regular boundary (so that trace operators are well-defined). Given $f \in L^2(\mathfrak{D})$, different choices of the domain $\mathrm{Dom}(\mathcal{L})$ of the operator L lead to different realizations of solution spaces. The closure of $\mathcal{L}|_{C_c^\infty(\mathfrak{D})}$ as an operator in $L^2(\mathfrak{D})$ is known as the *minimal realization* associated with $\mathcal{L}$, denoted L_{min} (c.f., Definition 4.2 in [41]). On the other hand, the *maximal realization*, denoted L_{max}, has domain of definition $\{u \in L^2(\mathfrak{D}) : \exists f \in L^2(\mathfrak{D}) \text{ such that } \mathcal{L}u = f \text{ weakly}\}$ (see Definition 4.1 in [41]). In the one-dimensional case when $\mathfrak{D}$ is a bounded interval in $\mathbb{R}$, the domains $\mathrm{Dom}(\mathcal{L}_{min})$ and $\mathrm{Dom}(\mathcal{L}_{max})$ of the two operators differ only by a two-dimensional space, but when $d \geq 2$ the difference is an infinite-dimensional space. Moreover, the domain $\mathrm{Dom}(\mathcal{L}_{max})$ of the maximal realization can be larger than the canonical Sobolev space $W^{2,2}(\mathfrak{D})$, see Example 2.5 for the definition. In this section we also use

$$W_0^{k,2}(\mathfrak{D}) = \{f \in W^{k,2}(\mathfrak{D}) : D^\alpha f|_{\partial\mathfrak{D}} = 0, |\alpha| \leq k-1\}, \qquad k = 1, 2, \cdots,$$

which is the closure of $C_c^\infty(\mathfrak{D})$ under $W^{k,2}$-norm.

In the definition of a Hilbert scale it is assumed that $\mathcal{L}$ is self-adjoint, which requires both the structural property $\langle \mathcal{L}x, y \rangle = \langle x, \mathcal{L}y \rangle$ and that the domains of $\mathcal{L}$ and its adjoint $\mathcal{L}^*$ be identical. The domain of $\mathcal{L}^*$ is determined by the domain of $\mathcal{L}$ (see Appendix A) and hence the latter must be chosen carefully.

As an example, consider the operator $\mathcal{L} = -\Delta$, where Δ is the d-dimensional Laplacian. Given $\mathrm{Dom}(\mathcal{L}_{min}) = W_0^{2,2}(\mathfrak{D})$ (see Theorem 10.19 in [84]), $\mathcal{L}_{min}$ is too small to be self-adjoint. Indeed, by the Green's identity (only real functions for simplicity)

$$\int_{\mathfrak{D}} (-\Delta u) v dx = \int_{\mathfrak{D}} u((-\Delta)^{\dagger} v) dx + \int_{\partial \mathfrak{D}} u \partial_{\nu} v - (\partial_{\nu} u) v d\sigma,$$

where $(-\Delta)^{\dagger}$ is the formal adjoint and ∂_{ν} is the directional derivative in the direction of outward pointing normal ν to the surface element $d\sigma$. This implies that $D(-\Delta^{\dagger}) \supseteq W^{2,2}(\mathfrak{D}) \cap W_0^{1,2}(\mathfrak{D}) \supsetneq W_0^{2,2}(\mathfrak{D})$. On the other hand, $\mathrm{Dom}(\mathcal{L}_{max}) \supset W^{2,2}(\mathfrak{D})$ is too big (see Exercise 11.10 in [84] for example). There are several self-adjoint extensions $\mathcal{L}$ of the minimal operator $\mathcal{L}_{min}$ that represent boundary conditions. For example, $\mathcal{L} = -\Delta$ with $\mathrm{Dom}(\mathcal{L}_{min}) = W_0^{2,2}(\mathfrak{D})$ has a self-adjoint extension $-\Delta_D$ with $\mathrm{Dom}(-\Delta_{\mathfrak{D}}) = W^{2,2}(\mathfrak{D}) \cap W_0^{1,2}(\mathfrak{D})$, corresponding to the Dirichlet condition $f|_{\partial\mathfrak{D}} = 0$ (Theorem 10.19, [84]). For the Neumann condition $\frac{\partial f}{\partial \nu}|_{\partial\mathfrak{D}} = 0$, one may use the variational form of $-\Delta u = f$ (see Section 10.6.2, [84]), to show that there exists a self-adjoint extension $-\Delta_N$ with domain $\mathrm{Dom}(-\Delta_N) = W^{1,2}(\mathfrak{D})$ (Theorem 10.20, [84]).

In the situations given above, $\sqrt{-\Delta}$ can be defined using the spectral measure of the self-adjoint extension of $-\Delta$. Consequently, the Hilbert scale $(H_s)_{s\in\mathbb{R}}$ generated by $\sqrt{-\Delta}$ (see Definition 2.7) is the scale of Sobolev spaces $\widetilde{W}^{s,2}(\mathfrak{D})$ of (generalized) functions $u \in W^{s,2}(\mathfrak{D})$ that satisfy the corresponding boundary condition. More advanced techniques, such as the pseudo-differential method, are necessary for more sophisticated boundary conditions, see [41, 55].

It is reasonable to consider Sobolev scales of functions that satisfy boundary conditions, since the existence and uniqueness of the solution of the forward problem (2.14) is proved by establishing the fact that $\mathcal{L}$ is isomorphism between Sobolev spaces satisfying boundary conditions, see [55]. As an immediate consequence, the isomorphism of the forward operator $\mathcal{A} = \mathcal{L}^{-1}$ is clear in the context of the corresponding inverse problem in form (I.1).

2.3 Smoothness in Higher Dimensions

In this section, we briefly review how to construct Hilbert spaces of functions on higher dimensional domains. We adopt the convention of notations for multidimensional vectors and multi-indices from Section 1.5.

First we recall some facts about tensor products. Let H_1, H_2 be two real separable Hilbert spaces. For each $f_1 \in H_1$, $f_2 \in H_2$, define a bilinear form $f_1 \otimes f_2$ acting on $H_1 \times H_2$ such as

$$f_1 \otimes f_2(h_1, h_2) := \langle f_1, h_1 \rangle_{H_1} \langle f_2, h_2 \rangle_{H_2}.$$

On the set $\mathscr{T}$ of finite linear combinations of the bilinear forms defined above, one can define an inner product

$$\langle f_1 \otimes f_2, g_1 \otimes g_2 \rangle_{\mathscr{T}} = \langle f_1, g_1 \rangle_{H_1} \langle f_2, g_2 \rangle_{H_2}. \tag{2.15}$$

The *tensor product* of H_1 and H_2 is the completion of $\mathscr{T}$ under the inner product defined in (2.15), denoted by $H_1 \otimes H_2$. A tensor product space inherits a base from the original spaces. That is, if $\{\varphi_i\}$ and $\{\psi_j\}$ are orthonormal bases for H_1 and H_2, then $\{\varphi_i \otimes \psi_j\}_{i,j}$ is an orthonormal basis for $H_1 \otimes H_2$.

We are only interested in the case that H_1 and H_2 are L^2 spaces on Euclidean domains. The results needed are summarised in the following theorem.

Theorem 2.16 (Theorem II.10, [81])**.** *For $i = 1, 2$, let $(\mathfrak{D}_i, \mu_i)$ be measure spaces such that $L^2(\mathfrak{D}_i, \mu_i)$ are separable. Then, the following statements hold.*

(i) There is a unique isomorphism from $L^2(\mathfrak{D}_1, \mu_1) \otimes L^2(\mathfrak{D}_2, \mu_2)$ to $L^2(\mathfrak{D}_1 \otimes \mathfrak{D}_2, \mu_1 \otimes \mu_2)$ such that

$$f \otimes g \mapsto fg.$$

(ii) Let $\widetilde{H}$ be a separable Hilbert space. Then there is a unique isomorphism from $L^2(\mathfrak{D}_1, \mu_1) \otimes \widetilde{H}$ to $L^2(\mathfrak{D}_1, \mu_1; \widetilde{H})$ such that

$$f(x) \otimes h \mapsto f(x)h.$$

(iii) In particular, there is a unique isomorphism from $L^2(\mathfrak{D}_1 \otimes \mathfrak{D}_2, \mu_1 \otimes \mu_2)$ to $L^2(\mathfrak{D}_1; L^2(\mathfrak{D}_2, \mu_2))$ such that

$$f(x, y) \mapsto (x \mapsto f(x, \cdot)),$$

and

$$\int_{\mathfrak{D}_1 \times \mathfrak{D}_2} |f(x, y)|^2 \, dxdy = \int_{\mathfrak{D}_1} \|f(x, \cdot)\|^2_{L^2(\mathfrak{D}_2, \mu_2)} \, dx.$$

2.3.1 Multi-dimensional Smoothness

Sobolev spaces on multi-dimensional domains are defined using the general statement from the previous subsection. Throughout this subsection, we assume that $\mathfrak{D}$ is a bounded domain in $\mathbb{R}^d$ with sufficiently regular boundary, e.g C^k with $k \in \mathbb{N}$ larger than the order of $\mathcal{L}$ as in Section 2.2.1.

Let H_1 be $L^2([0, T]; \mathbb{R})$ and H_2 be $L^2(\mathfrak{D}; \mathbb{R})$ We will omit the codomain if it is the real space $\mathbb{R}$. Following from Theorem 2.16,

$$L^2([0, T]; L^2(\mathfrak{D})) \cong L^2([0, T] \times \mathfrak{D}) \cong L^2(\mathfrak{D}) \otimes L^2([0, T]) \cong L^2([0, T]) \otimes L^2(\mathfrak{D}),$$

Moreover, by Theorem 2.16, if there exist orthonormal bases $\{\varphi_i\}_{i \in \mathbb{N}^d}$ for $L^2(\mathfrak{D})$ and $\{\psi_i\}_{i \in \mathbb{N}}$ for $L^2([0, T])$, the tensor orthonormal basis $\{\varphi_i \otimes \psi_j\}_{(i,j) \in \mathbb{N}^{d+1}}$ is an orthonormal basis for $L^2(\mathfrak{D}) \otimes L^2([0, T])$. In particular, there exists a unique isomorphism from $H \otimes L^2([0, T])$ to $L^2([0, T]; L^2(\mathfrak{D}))$ so that $\varphi \otimes f(t) \mapsto \varphi f(t)$. As a consequence, for any element $f \in L^2([0, T]; L^2(\mathfrak{D}))$, it admits the following representation

$$f(x, t) = \sum_{(i,j) \in \mathbb{N}^{d+1}} f_{(i,j)} \varphi_i \otimes \psi_j(x, t) = \sum_{(i,j) \in \mathbb{N}^{d+1}} f_{(i,j)} \varphi_i(x) \psi_j(t),$$

with

$$f_{(i,j)} = \langle f(x, t), \varphi_i \otimes \psi_j \rangle_{L^2(dx \times dt)} = \int_{[0,T]} \int_{\mathfrak{D}} f(x, t) \varphi_i(x) \, dx \; \psi_j(t) \, dt.$$

Introduce $\mathfrak{D}_T := \mathfrak{D} \times [0, T]$. The tensor space is isomorphic to the ordinary space $L^2(\mathfrak{D}_T)$, by the isomorphism statement above. On the other hand, the series representation above sheds light on how to obtain concrete smoothness scales centred at $L^2(\mathfrak{D}_T)$.

Consider an arbitrary L^2 space $L^2(\widetilde{\mathfrak{D}})$ with a bounded domain $\widetilde{\mathfrak{D}} \subset \mathbb{R}^m, m \in \mathbb{N}$. Since it is a separable Hilbert space, there exists an orthonormal basis $\{\varphi_k\}_{k\in\mathbb{N}^m}$. By abstract Parseval's identity, any function f in $L^2(\widetilde{\mathfrak{D}})$ admits a series expansion such that $f = \sum_k f_k \varphi_k$ in L^2 sense. With a multi-index $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m) \in \mathbb{R}^m_+$, an anisotropic smoothness class is defined to be the completion of

$$\left\{ f = \sum_{k\in\mathbb{N}^m} f_k \varphi_k \,\middle|\, \|f\|_{H_{\boldsymbol{\beta}}} := \left(\sum_{k\in\mathbb{N}^m} \left|\boldsymbol{\lambda}_k^{\boldsymbol{\beta}}\right|^2 f_k^2 \right)^{1/2} < \infty \right\}, \tag{2.16}$$

where $\boldsymbol{\lambda}_k = (\lambda_{k_1}, \cdots, \lambda_{k_m})$ for $k \in \mathbb{N}^m$, and

$$\left|\boldsymbol{\lambda}_k^{\boldsymbol{\beta}}\right| = \left|\boldsymbol{\lambda}_k^{\boldsymbol{\beta}}\right|_2 = \left(\sum_{i\leq m} \lambda_{k_i}^{2\beta_i} \right)^{1/2}.$$

Similarly, the type of smoothness classes above can be introduced to sequence spaces. Let $\ell^2(\mathbb{N}^m)$ with $m \in \mathbb{N}$ be the m-dimensional square integrable sequence space, i.e. for any $f = \{f_k\}_{k\in\mathbb{N}^m} \in \ell^2(\mathbb{N}^m)$,

$$\|f\|_{\ell^2} = \left(\sum_{k\in\mathbb{N}^m} f_k^2 \right)^{1/2} < \infty.$$

The anisotropic ellipsoid $h_{\boldsymbol{\beta}}$ with $\boldsymbol{\beta} \in \mathbb{R}^m_+$ is the completion of the set $f \in \ell^2(\mathbb{N}^m)$ such that, with same $\{\boldsymbol{\lambda}_k\}_k$,

$$\|f\|_{h_{\boldsymbol{\beta}}} = \left(\sum_{k\in\mathbb{N}^m} |\boldsymbol{\lambda}_k^{\boldsymbol{\beta}}|^2 f_k^2 \right)^{1/2} < \infty \tag{2.17}$$

under the norm $\|\cdot\|_{h_{\boldsymbol{\beta}}}$.

The Sequence space is convenient as we often deal with the coefficients. Once the basis $\{\varphi_k\}_{k\in\mathbb{N}^m}$ is fixed, given $f = \sum_k f_k \varphi_k$ and $\widetilde{f} = \{f_k\}_{k\in\mathbb{N}^m}$, we have the isometry,

$$\|f\|_{L^2} = \|\widetilde{f}\|_{\ell^2}, \quad \|f\|_{H_{\boldsymbol{\beta}}} = \|\widetilde{f}\|_{h_{\boldsymbol{\beta}}}.$$

Remark 2.17. It is worth noting that so far we only assume that $\{\varphi_k\}$ is an orthonormal basis of $L^2(\mathfrak{D})$ and the smoothness is characterised by weighted ℓ^2 norms of the coefficients. In order to establish the connection to canonical Sobolev spaces, additional requirements are necessary, see the upcoming subsections.

Different types of smoothness might be more suitable for various problems. As we will see in the study of the inference for evolution equations in Part III, for the recovery of an initial condition, a type of spatial smoothness classes is expected, while for the recovery of a drift term, we need to introduce a proper smoothness class to describe the space-time regularity. To distinguish the different types of smoothness, we call a smoothness class *isotropic* when the smoothness index $\boldsymbol{\beta}$ satisfies $\beta_i = \beta_j$ for all $i, j \leq m$, or otherwise *anisotropic*.

It is also convenient to introduce the harmonic mean, which will be used to describe the 'balanced' smoothness of anisotropic Sobolev spaces (see Section 10.4). For a multi-index $\beta \in \mathbb{R}_+^m$, the harmonic mean is defined as

$$\mathcal{H}(\beta) := \frac{m}{\sum_{i=1}^m (1/\beta_i)}. \tag{2.18}$$

Below we collect some elementary but useful lemmas related to the harmonic mean, which will be used mainly in Part III. We restrict to the case that $\lambda_{k_i} \simeq k_i, i = 1, \cdots, m$, in (2.16).

Lemma 2.18. *Given a multi-index $\alpha = (\alpha_1, \cdots, \alpha_m) \in \mathbb{R}_+^m$, if $\mathcal{H}(\alpha) > m/2$, then $\sum_{k\in\mathbb{N}^m} |k^\alpha|^{-2}$ is finite.*

Proof. Since $|k^\alpha|^2 = \sum_{i\leq m} k_i^{2\alpha_i} \geq \frac{1}{m} \prod_{i\leq m} k_i^{2\alpha_i/m}$,

$$\sum_{k\in\mathbb{N}^m} |k^\alpha|^{-2} \leq_m \sum_{k\in\mathbb{N}^m} \prod_{i\leq m} k_m^{-2\alpha_i/m}.$$

Introduce the hypercubes

$$C_n = \{k \in \mathbb{N}^m : k_i \lesssim n^{\mathcal{H}(\alpha)/\alpha_i},\ i = 1, \cdots, m\},$$

where the constants are independent from m, i and n. The number of points in $\mathbb{N}^m$ covered by C_n is $\#C_n \simeq n^m$, and consequently $\#[C_n \backslash C_{n-1}] \simeq n^{m-1}$. Hence, the summation can be estimated as,

$$\sum_{n\in\mathbb{N}} \sum_{k\in[C_n\backslash C_{n-1}]} \prod_{i\leq m} k_i^{-2\alpha_i/m} \simeq \sum_{n\in\mathbb{N}} n^{m-1} n^{-2\mathcal{H}(\alpha)}.$$

The result immediately follows. □

The following elementary lemma entitles us the freedom to choose between equivalent norms.

Lemma 2.19. *Let $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m) \in \mathbb{R}_+^m$ and $k = (k_1, \cdots, k_m)$ be a multi-index. Then, with a constant only dependent on m,*

$$|\boldsymbol{k}^{\boldsymbol{\beta}}|^2 = \sum_{i=1}^m k_i^{2\beta_i} \simeq_m \left(\sum_{i=1}^m k_i^{\beta_i}\right)^2.$$

In particular, if $\beta_1 = \cdots = \beta_m = \beta > 0$, for any $q \in [0, \infty]$,

$$|k^{\boldsymbol{\beta}}|_2 \simeq_m |k|_q^\beta,$$

where the constant is between 1 and m.

Proof. The first equivalence is elementary. The second one is argued as follows. Recall $|\cdot|_p$ is the p-norm on $\mathbb{R}^d$. From power mean inequality one can derive $|v|_p \leq |v|_r \leq d^{1/r-1/p}|v|_p$ for $0 < r < p$. The proof is concluded by applying the inequality to $k^{\boldsymbol{\beta}}$ with $\boldsymbol{\beta} = (\beta, \cdots, \beta) \in \mathbb{R}_+^d$. □

The following lemma provides an error estimate on the truncation of a sequence in an anisotropic Sobolev ellipsoid, which can be easily translated to the projection in a Sobolev space.

Lemma 2.20. *Let P_N be the projection of a sequence to the coordinates*

$$\{k < N : k_i < N_i,\ i \leq m\}$$

with $N = (N_1, \cdots, N_m)$. For $f \in h_{\boldsymbol{\beta}}$ with $\lambda_{k_i} \simeq k_i, i = 1, \cdots, m$,

$$\|P_N f - f\|_{\ell^2}^2 \leq \left(\sum_{i \leq m} N_i^{-2\beta_i} \right) \|f\|_{h_\beta}^2.$$

In particular, when $N_i \simeq n^{\mathcal{H}(\boldsymbol{\beta})/(\beta_i m)}$, we have $\prod_{i \leq m} N_i \simeq n$ and

$$\|P_N f - f\|_{\ell^2} \lesssim_m n^{-\mathcal{H}(\boldsymbol{\beta})/m} \|f\|_{h_\beta}.$$

Proof.

$$\|P_N f - f\|_{\ell^2}^2 \leq \sum_{i \leq m} \left[\sum_{k_i \geq N_i} \sum_{\substack{k_j \in \mathbb{N}: \\ j \leq m \\ j \neq i}} f_k^2 \right] = \sum_{i \leq m} S_i.$$

The items S_i can be bounded from above by,

$$\sum_{k_i \geq N_i} \sum_{\substack{k_j \in \mathbb{N}: \\ j \leq m \\ j \neq i}} \frac{|\boldsymbol{k}^{\boldsymbol{\beta}}|^2}{N_i^{2\beta_i}} f_k^2 \leq N_i^{-2\beta_i} \|f\|_{h_{\boldsymbol{\beta}}}^2.$$

□

Using the lemmas above, one can show that the anisotropic smoothness class is a multi-dimensional version of a smoothness class.

Corollary 2.21. *Consider an anisotropic smoothness class defined in (2.16) with $\lambda_{k_i} \simeq k_i, i = 1, \cdots, m$, then it is a (multi-dimensional) smoothness scale in the following sense.*

(i) For $\boldsymbol{s} < \boldsymbol{t}$, i.e. $s_i \leq t_i$ for all $i \leq m$ and $s_i < t_i$ for at least one $i \leq m$, the space $H_{\boldsymbol{t}}$ is a dense subspace of $H_{\boldsymbol{s}}$ and $\|f\|_{\boldsymbol{s}} \lesssim \|f\|_{\boldsymbol{t}}$, for $f \in H_{\boldsymbol{t}}$.

(ii) For $\boldsymbol{s} \geq 0$, i.e. $s_i \geq 0$ for all $i \leq m$, $f \in H_0$ can be viewed as element of $H_{-\boldsymbol{s}} \supset H_0$,

$$\|f\|_{-\boldsymbol{s}} = \sup_{\|g\|_{\boldsymbol{s}} \leq 1} \langle f, g \rangle_0, \qquad f \in H_0. \tag{2.19}$$

Furthermore, Assumption 2.3 is satisfied with $\delta(j, \boldsymbol{s}) \simeq j^{-\mathcal{H}(\boldsymbol{s})/m}$.

2.4 Approximation Number and Metric Entropy

The jth *approximation number* of a bounded linear operator $T : G \to H$ between normed spaces is defined as

$$a_j(T : G \to H) = \inf_{U:\mathrm{Rank}U<j} \|T - U\|_{G\to H}, \tag{2.20}$$

where the infimum is taken over all linear operators $U : G \to H$ of rank (i.e., dimension of the range space) strictly less than j, and the norm on the right is the operator norm $\|T-U\|_{G\to H}= \sup_{f:\|f\|_G\le 1}\|(T-U)f\|_H$. The approximation numbers measure the possibility of approximating an operator by simpler operators of finite-dimensional rank. There is a rich literature on approximation numbers. The main purpose of the present section is to note their relationship to singular values and to metric entropy. Metric entropy plays an important role in the characterization of contraction rates of Bayesian posterior distributions.

If $G \subset H$, we can take T equal to the embedding $\iota : G \to H$, and then by linearity we see that there exists an operator U of rank smaller than j such that

$$\|f - Uf\|_H \lesssim a_j(\iota : G \to H)\,\|f\|_G, \qquad \forall f \in G.$$

If H is a Hilbert space, then the minimizing finite-rank operator U is of course the orthogonal projection P_j on V_j. However, the approximation numbers also 'search' an optimal projection space. If we take $G = H_s$ and $H = H_0$, then the range space V_j of U satisfies the approximation property (2.2), with the numbers $\delta(j, s)$ taken equal to the approximation numbers $a_j(\iota : H_s \to H_0)$.

The approximation number is an example of an *s-number*, as introduced in [76]. In general s-numbers are defined as maps $T \mapsto (s_j(T))_{j\in\mathbb{N}}$, attaching to every operator T a sequence of nonnegative numbers $s_j(T)$, satisfying certain axiomatic properties. In general, approximation numbers attached to operators $T : H \to H$ are the 'largest' possible s-numbers, but on Hilbert spaces there is only one s-number: all s-numbers are the same (see 2.11.9 in [77]). Because the singular values are also s-numbers, the latter unicity yields the important relation that the approximation numbers of operators on Hilbert spaces are equal to their singular values. Recall here that the singular values of a compact operator $T : G \to H$ are the roots of the eigenvalues of the self-adjoint operator $T^*T : G \to G$.

The finite-rank approximations U that (nearly) achieve the infimum in the definition of the approximation numbers for different j are not a-priori ordered. However, in many cases there exists a basis $(\phi_i)_{i\in\mathbb{N}}$ such that the projections on the linear span of the first $j-1$ basis elements achieve the infimum. For Sobolev spaces e.g. spline bases, the Fourier basis, or wavelet bases are all 'optimal' in this sense (see [18, 79]).

Approximation numbers are strongly connected to metric entropy. In the literature the connection is usually made through the notion of 'entropy numbers', which are defined as follows. The j-th *entropy number* $e_j(T)$ of an operator $T : G \to H$ is defined as the infimum of the numbers $\varepsilon > 0$ so that the image $T(U_G) \subset H$ of the unit ball U_G in G can be covered by 2^{j-1} balls of radius ε in

H; or more formally, with U_H the unit ball in H,

$$e_j(T) = \inf\Big\{\varepsilon > 0 : T(U_G) \subset \bigcup_{i=1}^{2^j-1} (h_i + \varepsilon U_H), \text{for some } h_1, \ldots, h_{2^j-1} \in H\Big\}.$$

The function $j \mapsto e_j(T)$ is roughly the inverse function of the metric entropy of $T(U_G)$ relative to the metric induced by $\|\cdot\|_H$. Recall that the *metric entropy* of a metric space (U, d) is the logarithm of the covering number $N(\varepsilon, U, d)$, which is the minimal number of d-balls of radius $\varepsilon > 0$ needed to cover the space U. Presently we consider the metric entropy $H(\varepsilon, T) = \log N(\varepsilon, T(U_G), \|\cdot\|_H)$ of $T(U_G)$ under the metric of H. Roughly we have that

$$N(\varepsilon, T(U_G), \|\cdot\|_H) \simeq 2^{j-1}, \qquad \text{if} \qquad e_j(T) \simeq \varepsilon.$$

If we use the logarithm at base 2, then the map $\varepsilon \mapsto H(\varepsilon, T)$ is approximately inverse to the map $j \mapsto e_j(T)$.

Now it is proved in [26] that for any operator $T : G \to H$ between Hilbert spaces with infinite-dimensional ranges:

$$e_{j+1}(T) \leq 2a_{J+1}(T) \leq 2\sqrt{2}e_{J+2}(T),$$

for any natural numbers j, J satisfying:

$$j \log 2 \geq 2 \sum_{i=1}^{J} \log \frac{3a_i(T)}{a_{J+1}(T)}.$$

As shown in [26] this relationship between entropy numbers and approximation numbers may be solved to derive the entropy number from the approximation numbers in many cases.

The following lemma gives one example, important to the present thesis.

Lemma 2.22 (Metric entropy)**.** *For a smoothness scale $(H_s)_{s\in\mathbb{R}}$ satisfying (2.2) with $\delta(j, s) = j^{-s/d}$, and $s > 0$ and $t \geq 0$,*

$$\log N(\varepsilon, \{f \in H_s : \|f\|_s \leq 1\}, \|\cdot\|_{-t}) \sim \varepsilon^{-d/(s+t)}. \tag{2.21}$$

Proof. By (2.5) the approximation number $a_j(\iota : H_s \to H_{-t})$ is of the order $\delta(j, s)\delta(j, t) = j^{-(s+t)/d}$. It is shown in [26] that the entropy numbers $e_j(\iota : H_s \to H_{-t})$ are of the order $j^{-(s+t)/d}$. By the preceding reasoning this can be inverted to obtain the order of the metric entropy of the image of the unit ball in H_{-t}. □

Similarly, the results attained above can be extended to the multi-dimensional scale as follows.

Lemma 2.23. *Let $h_\beta(\mathbb{N}^m)$ be a Sobolev ellipsoid given in (2.17). When $\boldsymbol{\beta} > 0$, i.e. $\beta_i > 0$, for all $1 \leq i \leq m$, then both the approximation number and entropy number of the canonical embedding $\iota : h_{\boldsymbol{\beta}}(\mathbb{N}^m) \to \ell^2(\mathbb{N}^m)$ are of the order*

$$a_j(\iota : h_{\boldsymbol{\beta}} \to \ell^2) \simeq e_j(\iota : h_{\boldsymbol{\beta}} \to \ell^2) \simeq j^{-\mathcal{H}(\boldsymbol{\beta})/m}.$$

Proof. Let P_N be the truncation of a sequence to $\{k < N : k_i < N_i,\ i \leq m\}$ with $N = (N_1, \cdots, N_m)$. From Lemma 2.20, for the truncation at n coefficients, i.e. $\prod_{i\leq m} N_i \simeq n$, the minimal projection error

$$\|P_N f_0 - f_0\|_{\ell^2} \lesssim_d n^{-\mathcal{H}(\boldsymbol{\beta})/m} \|f\|_{h_{\boldsymbol{\beta}}}$$

is achieved when N_i are balanced, i.e. $N_i \simeq n^{\mathcal{H}(\boldsymbol{\beta})/(\beta_i m)}$ for all $i \leq m$. This estimate leads to the upper bound of the approximation number

$$a_n(\iota : h_{\boldsymbol{\beta}} \to \ell^2) \lesssim_d n^{-\mathcal{H}(\boldsymbol{\beta})/m}.$$

Taking a sequence whose only nonzero entry is 1 and its multi-index satisfies $k_i \simeq n^{\mathcal{H}(\boldsymbol{\beta})/(\beta_i m)}$ for $1 \leq i \leq d$, it is also straightforward to show that

$$a_n(\iota : h_{\boldsymbol{\beta}} \to \ell^2) \gtrsim n^{-\mathcal{H}(\boldsymbol{\beta})/m}.$$

Then by the uniqueness of s-number on Hilbert spaces, we conclude that the entropy number is of the same order as the approximation number. □

Corollary 2.24 (Metric entropy)**.** *Under the same assumption in Lemma 2.23, the metric entropy is given by, as* $\varepsilon \downarrow 0$,

$$H(\varepsilon, \iota) := \log N\Big(\varepsilon, \{f \in h_{\boldsymbol{\beta}} : \|f\|_{h_{\boldsymbol{\beta}}} \leq 1\}, \|\cdot\|_{\ell^2}\Big) \sim \varepsilon^{-m/\mathcal{H}(\boldsymbol{\beta})}.$$

Proof. The proof is same as Lemma 2.22 and hence is omitted. □

In a similar way it is possible to invert approximation numbers that are not of the polynomial form $j^{-s/d}$. There are many examples of this type, for instance, involving additional logarithmic terms, or exponentially decreasing rates. We defer the related discussion until Chapter 10 in Part III.

2.5 Notes

Hilbert scales

Hilbert scales and the relate concepts such as Gelfand triples have been studied be many authors from various fields with different motivations. The Gelfand triples were introduced in [32] to study the theory of generalised functions[1]. The idea of scales of function spaces can be traced back to Krein's work, [62, 63], whose main focus was on the interpolation theory of linear operators. In [28], rigged Hilbert space was studied under another name 'Sobolev towers' in the context of operator semigroups. In the field of stochastic analysis, Hilbert scales can be found in Hida's analysis of white noise functionals (as known as generalised stochastic processes) [47], and it is also used to construct the solution spaces in some studies of stochastic partial differential equations [48, 85]. Its application in regularization theory can be found e.g. in Chapter 8 of [29]. For the application in physics, we refer to the survey paper [31].

[1] We would like to point out that [32] contains a flaw as addressed by the translator. On the bottom of page 122, the translator raised a concern on the proof of spectral theorem of normal operators in rigged Hilbert spaces. The correct proof is given in [40].

Compactness

Compared to separable spaces mentioned in the Section 1.6, the elements in a compact space can be estimated by a finite dimensional approximation but yet retaining reasonable accuracy. This property is important for any estimation procedure. [50] systematically describes the connections between compactness and statistical estimation. Applications can be found in [51, 52], which are also related to Part III of this thesis.