Ruprecht-Karls-Universität Heidelberg
Institut für Angewandte Mathematik
Prof. Dr. Jan JOHANNES

Outline of the lecture course

STATISTICS OF INVERSE PROBLEMS

Summer semester 2017

Preliminary version: May 17, 2017

If you find **errors in the outline**, please send a short note
by email to johannes@math.uni-heidelberg.de

MΛTHEMΛTIKON, Im Neuenheimer Feld 205, 69120 Heidelberg
phone: +49 6221 54.14.190 – fax: +49 6221 54.53.31
email: johannes@math.uni-heidelberg.de
webpage: www.razbaer.eu/ag-johannes/vl/SIP-SS17/

Table of contents

Chapter 1

Introduction

Short summary

Statistical ill-posed inverse problems are becoming increasingly important in a diverse range of disciplines, including geophysics, astronomy, medicine and economics. Roughly speaking, in all of these applications the observable signal $g = Tf$ is a transformation of the functional parameter of interest f under a linear operator T. Statistical inference on f based on an estimation of g which usually requires an inversion of T is thus called an *inverse problem*. The lecture course focuses on statistical ill-posed inverse problems with noise in the operator where neither the signal g nor the linear operator T are known in advance, although they can be estimated from the data. Our objective in this context is the construction of minimax-optimal fully data-driven estimation procedures of the unknown function f. Special attention is given to four models and their extensions, namely Gaussian inverse regression, density deconvolution, functional linear regression and non-parametric instrumental regression, which lead naturally to statistical ill-posed inverse problems with noise in the operator.

Applications

Density deconvolution with unknown error distribution. The biologist who is interested in the density f of a gene-expression intensity X, can record in a cDNA microarray the expressed gene intensity X only corrupted by the intensity of a background noise ε, that is $Y = X + \varepsilon$. If the additive measurement error ε is independent of X then the density $g = f \star \phi$ of Y equals the convolution of f and the error density ϕ. Consequently, recovering f from the estimated density $g = C_\phi f$ of Y is an inverse problem where C_ϕ is the convolution operator defined by the error density ϕ. In this situation, the density f of the random variable X has to be estimated non-parametrically based on an iid. sample from a noisy observation Y of X which is called a density deconvolution problem. There is a vast literature on deconvolution with known error density which leads to a statistical ill-posed inverse problem with known operator. On the other hand, if the error density ϕ is estimated from an additional calibration sample of the error ε then the deconvolution problem corresponds to a statistical ill-posed inverse problem with noise in the operator.

Functional linear regression. In climatology, prediction of level of ozone pollution based on continuous measurements of pollutant indicators is often modelled by a functional linear model. In this context a scalar response Y (i.e. the ozone concentration) is modelled in dependence of a random function X (i.e. the daily concentration curve of a pollutant indicator). Typically the dependence is assumed to be linear which finds its expression in a linear normal equation $g = \Gamma f$ where g is the cross-correlation between Y and X, and Γ is the covariance operator associated to the indicator X. Note that both the cross-correlation function g and the covariance operator Γ need to be estimated in practice. Consequently, the non-parametric estimation of the functional slope parameter f based on an iid. sample from (Y, X) leads to a statistical ill-posed inverse problem with noise in the operator.

Non-parametric instrumental regression. An econometrician who wants to analyse an economic relation between a response Y and an endogenous vector X of explanatory variables, might incorporate a vector of exogenous instruments Z. This situation is usually treated by considering a conditional moment equation $g = K f$ where $g = \mathbb{E}_{Y|Z}$ is the conditional expectation function of Y given Z and K is the conditional expectation operator of X given Z. As these are unknown in practice, inference on f based on an iid. sample from (Y, X, Z) is a statistical ill-posed inverse problem with noise in the operator.

STATISTICAL ILL-POSED INVERSE PROBLEMS

We study non-parametric estimation of the functional parameter of interest f in an inverse problem, that is, its reconstruction based on an estimation of a linear transformation $g = Tf$. It is important to note that in all the applications discussed above both the signal g and the inherent transformation T are unknown in practice, although they can be estimated from the data. The estimated signal $\widehat{g}$ and operator $\widehat{T}$ respectively given by

$$\widehat{g} = Tf + \sqrt{n}\,\dot{W} \quad \text{and} \quad \widehat{T} = T + \sqrt{m}\,\dot{B}. \tag{1.1}$$

are noisy versions of g and T contaminated by additive random errors $\dot{W}$ and $\dot{B}$ with respective noise levels n and m. Consequently, a statistical inference on the functional parameter of interest f has to take into account that a random noise is present in both the estimated signal $\dot{W}$ and the estimated operator $\dot{B}$.

Gaussian inverse regression with noise in the operator. A particularly interesting situation is given by model (1.1) where the random error $\dot{W}$ and $\dot{B}$ are independent Gaussian white noises. This model is particularly useful to characterise the influence of an *a priori* knowledge of the operator T. To this end we will compare three cases: First, the operator T is *fully known* in advance, i.e., the noise level m is equal to zero. Second, it is *partially known*, that is, the eigenfunctions of T are known in advance but the "observed" eigenvalues of T are contaminated with an additive Gaussian error. Third, the operator T is *unknown*.

MINIMAX-OPTIMAL ESTIMATION

Typical questions in this context are the non-parametric estimation of the functional parameter f on an interval or in a given point, referred to as global or local estimation, respectively. However, these are special cases in a general framework where the accuracy of an estimator $\widehat{f}$ of f given the estimations (1.1) is measured by a distance $\mathfrak{d}_{\text{ist}}(\widehat{f}, f)$. A suitable choice of the distance covers than the global as well as the local estimation problem. Moreover, denoting by $\mathbb{P}^{n,m}_{f,T}|\mathfrak{d}_{\text{ist}}(\widehat{f}, f)|^2$ (or $\mathbb{E}^{n,m}_{f,T}|\mathfrak{d}_{\text{ist}}(\widehat{f}, f)|^2$) its expectation w.r.t. the probability measure $\mathbb{P}^{n,m}_{f,T}$ associated with the observable quantities (1.1) we call the quantity $\mathbb{P}^{n,m}_{f,T}|\mathfrak{d}_{\text{ist}}(\widehat{f}, f)|^2$ risk of the estimator $\widehat{f}$ of f. It is well-known that in terms of its risk the attainable accuracy of an estimation procedure is essentially determined by the conditions imposed on f and the operator T. Typically, these conditions are expressed in the form $f \in \mathcal{F}$ and $T \in \mathcal{T}$ for suitable chosen classes $\mathcal{F}$ and $\mathcal{T}$. The class $\mathcal{F}$ reflects prior information on the solution f, e.g., its level of smoothness, and the class $\mathcal{T}$ imposes among others conditions on the decay of the eigenvalues of the operator T. Consequently, let us introduce the associated family of probability measures $\mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}$. The accuracy of $\widehat{f}$ is hence measured by its maximal risk over the classes $\mathcal{F}$ and $\mathcal{T}$, that is,

$$\mathfrak{R}_{\mathfrak{d}}\big[\widehat{f}\,|\,\mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}\big] := \sup\big\{\mathbb{P}^{n,m}_{f,T}|\mathfrak{d}_{\text{ist}}(\widehat{f},f)|^2, \mathbb{P}^{n,m}_{f,T} \in \mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}\big\}.$$

Moreover, $\widehat{f}$ is called minimax-optimal up to a finite positive constant C if $\mathfrak{R}_{\mathfrak{d}}\big[\widehat{f}\,|\,\mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}\big] \leqslant C\inf_{\widetilde{f}}\mathfrak{R}_{\mathfrak{d}}\big[\widetilde{f}\,|\,\mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}\big]$ where the infimum is taken over all possible estimators of f. Consequently, minimax-optimality of an estimator $\widehat{f}$ based on observations (1.1) is usually shown by establishing both an upper and a lower bound. More precisely, we search a finite positive quantity $\mathcal{R}^{n,n}_{\mathfrak{d}}$ depending only on the noise levels and the classes such that

$$\mathfrak{R}_{\mathfrak{d}}\big[\widehat{f}\,|\,\mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}\big] \leqslant C_1\mathcal{R}^{n,m}_{\mathfrak{d}} \quad \text{and} \quad \mathcal{R}^{n,m}_{\mathfrak{d}} \leqslant C_2\inf_{\widetilde{f}}\mathfrak{R}_{\mathfrak{d}}\big[\widetilde{f}\,|\,\mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}\big]$$

where C_1, C_2 are finite positive constants independent of the noise levels. Moreover, the quantity $\mathcal{R}^{n,m}_{\mathfrak{d}}$ is called the minimax-optimal rate of convergence over the family $\mathbb{P}_{\mathcal{F},\mathcal{T}} := \{\mathbb{P}^{n,m}_{\mathcal{F},\mathcal{T}}, n, m \in (0,1)\}$ if it tends to zero as n and m tend to zero.

Adaptive estimation

In many cases the proposed estimation procedures rely on the choice of at least one tuning parameter, which in turn, crucially influences the attainable accuracy of the constructed estimator. In other words, these estimation procedures can attain the minimax rate $\mathcal{R}^{n,n}_{\mathfrak{d}}$ over the family $\mathbb{P}_{\mathcal{F},\mathcal{T}}$ only if the inherent tuning parameters are chosen optimally. This optimal choice, however, follows often from a classical squared-bias-variance compromise and requires a *a priori* knowledge about the classes $\mathcal{F}$ and $\mathcal{T}$, which is usually inaccessible in practice. This motivates its data-driven choice in the context of non-parametric statistics since its very beginning in the fifties of the last century. A demanding challenge is then a fully data driven method to select the tuning parameters in such a way that the resulting data-driven estimator of f still attains the minimax-rate up to a constant over a variety of classes $\mathcal{F}$ and $\mathcal{T}$. The fully data driven estimation procedure is then called *adaptive*.

Chapter 2

Theoretical basics and terminologies

2.1 Hilbert space

For a detailed and extensive survey on functional analysis we refer the reader, for example, to Werner [2011] or the series of textbooks by Dunford and Schwartz [1988a,b,c].

§2.1.1 **Definition**. A normed vector space $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$ over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ that is complete (in a Cauchy-sense) is called a (real or complex) *Hilbert space* if there exists an inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ on $\mathbb{H} \times \mathbb{H}$ with $|\langle h, h \rangle_{\mathbb{H}}|^{1/2} = \|h\|_{\mathbb{H}}$ for all $h \in \mathbb{H}$. □

§2.1.2 **Property**. *Let $(\mathbb{H}, \|\cdot\|_1)$ and $(\mathbb{H}, \|\cdot\|_2)$ be complete normed vector spaces. If there exists a constant $K > 0$ such that $\|h\|_1 \leqslant K \|h\|_2$ for any $h \in \mathbb{H}$ then, $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent.* □

§2.1.3 **Property**.
(Cauchy-Schwarz inequality) $|\langle h_1, h_2 \rangle_{\mathbb{H}}| \leqslant \|h_1\|_{\mathbb{H}} \cdot \|h_2\|_{\mathbb{H}}$ *for all* $h_1, h_2 \in \mathbb{H}$. □

§2.1.4 **Examples**.

(i) For $k \in \mathbb{N}$ the *Euclidean space* $\mathbb{K}^k$ endowed with the Euclidean inner product $\langle x, y \rangle := \overline{y}^t x$ and the induced Euclidean norm $\|x\| = (\overline{x}^t x)^{1/2}$ for all $x, y \in \mathbb{K}^k$ is a Hilbert space. More generally, given a strictly positive definite $(k \times k)$-matrix W, $\mathbb{K}^k$ endowed with the weighted inner product $\langle x, y \rangle_W := \overline{y}^t W x$ for all $x, y \in \mathbb{K}^k$ is also a Hilbert space.

(ii) Given $\mathcal{J} \subseteq \mathbb{Z}$, denote by $\mathbb{K}^{\mathcal{J}}$ the vector space of all $\mathbb{K}$-valued sequences over $\mathcal{J}$ where we refer to any sequence $(x_j)_{j \in \mathcal{J}} \in \mathbb{K}^{\mathcal{J}}$ as a whole by omitting its index as for example in «the sequence x» and arithmetic operations on sequences are defined element-wise, i.e., $xy := (x_j y_j)_{j \in \mathcal{J}}$. In the sequel, let $\|x\|_{\ell^p} := (\sum_{j \in \mathcal{J}} |x_j|^p)^{1/p}$, for $p \in [1, \infty)$, and $\|x\|_{\ell^\infty} := \sup_{j \in \mathcal{J}} |x_j|$. Thereby, for $p \in [1, \infty]$, consider $\ell^p(\mathcal{J}) := \left\{ (x_j)_{j \in \mathcal{J}} \in \mathbb{K}^{\mathcal{J}}, \|x\|_{\ell^p} < \infty \right\}$, or ℓ^p for short, endowed with the norm $\|\cdot\|_{\ell^p}$. In particular, $\ell^2(\mathcal{J})$ is the usual *Hilbert space of square summable sequences over $\mathcal{J}$* endowed with the inner product $\langle x, y \rangle_{\ell^2} := \sum_{j \in \mathcal{J}} x_j \overline{y}_j$ for all $x, y \in \ell^2(\mathcal{J})$.

(iii) For a strictly positive sequence $\mathfrak{v}$ consider the *weighted norm* $\|x\|^2_{\ell^2_{\mathfrak{v}}} := \sum_{j \in \mathcal{J}} \mathfrak{v}_j^2 |x_j|^2$. We define $\ell^2_{\mathfrak{v}}(\mathcal{J})$, or $\ell^2_{\mathfrak{v}}$ for short, as the completion of $\ell^2(\mathcal{J})$ w.r.t. $\|\cdot\|_{\mathfrak{v}}$ which is a Hilbert space endowed with the inner product $\langle x, y \rangle_{\ell^2_{\mathfrak{v}}} := \langle \mathfrak{v}x, \mathfrak{v}y \rangle_{\ell^2} = \sum_{j \in \mathcal{J}} \mathfrak{v}_j^2 x_j \overline{y}_j$ for all $x, y \in \ell_{\mathfrak{v}}$.

(iv) Let $\mathscr{B}$ be the Borel-σ-algebra on $\mathbb{K}$. Given a measure space $(\Omega, \mathscr{A}, \mu)$ denote by $\mathbb{K}^\Omega$ the vector space of all $\mathbb{K}$-valued functions $f : \Omega \to \mathbb{K}$. Recall that $\|f\|_{L^p_\mu} = (\mu|f|^p)^{1/p} = \left(\int_\Omega |f(\omega)|^p \mu(d\omega) \right)^{1/p}$, for $p \in [1, \infty)$, and $\|f\|_{L^\infty_\mu} := \inf\{c : \mu(|f| > c) = 0\}$, where for $p \in [1, \infty]$, we write $L^p(\Omega, \mathscr{A}, \mu) := \{f \in \mathbb{K}^\Omega, \mathscr{A}\text{-}\mathscr{B}\text{-measurable}, \|f\|_{L^p} < \infty\}$, $L^p_\mu(\Omega)$ or L^p_μ for short, which is endowed with the norm $\|\cdot\|_{L^p_\mu}$ for short. In case μ is the Lebesgue measure, then we may write $L^p(\Omega, \mathscr{A})$, $L^p(\Omega)$, L^p and $\|\cdot\|_{L^p}$ for short. Moreover,

$L^2(\Omega, \mathscr{A}, \mu)$, $L^2_\mu(\Omega)$ or L^2_μ for short, is the usual *Hilbert space of square μ-integrable, $\mathscr{A}$-$\mathscr{B}$-measurable functions on Ω* endowed with the inner product $\langle f, g\rangle_{L^2_\mu} := \mu(f\overline{g})$ for all $f, g \in L^2_\mu$.

(v) Let X be a random variable (r.v.) on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ taking its values in a measurable space $(\mathcal{X}, \mathscr{B})$. We denote by $\mathbb{P}^X := \mathbb{P} \circ X^{-1}$ the image probability measure of $\mathbb{P}$ under X on $(\mathcal{X}, \mathscr{B})$. For $p \in [1, \infty]$ we set $L^p_X := L^p(\mathcal{X}, \mathscr{B}, \mathbb{P}^X)$ where L^2_X is a Hilbert space endowed with $\langle f, g\rangle_{L^2_X} = \mathbb{P}^X(f\overline{g})$ for all $f, g \in L^2_X$. □

§2.1.5 **Definition.** A subset $\mathcal{U}$ of a Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ is called *orthogonal* if

$$\forall u_1, u_2 \in \mathcal{U},\ u_1 \neq u_2 :\ \langle u_1, u_2\rangle_{\mathbb{H}} = 0$$

and *orthonormal system (ONS)* if in addition $\|u\|_{\mathbb{H}} = 1, \forall u \in \mathcal{U}$. We say $\mathcal{U}$ is an *orthonormal basis (ONB)* if $\mathcal{U} \subset \mathcal{U}'$ and $\mathcal{U}'$ is ONS, then $\mathcal{U} = \mathcal{U}'$, i.e., if it is a *complete* ONS.

§2.1.6 **Examples.**

(i) Consider the real Hilbert space $L^2([0,1])$ w.r.t. the Lebesgue measure. The *trigonometric basis* $\{\psi_j, j \in \mathbb{N}\}$ given for $t \in [0,1]$ by

$$\psi_1(t) := 1,\ \psi_{2k}(t) := \sqrt{2}\cos(2\pi k t),\ \psi_{2k+1}(t) := \sqrt{2}\sin(2\pi k t),\ k = 1, 2, \dots,$$

is orthonormal and complete, i.e. an ONB.

(ii) Consider the complex Hilbert space $L^2([0,1))$, then the *exponential basis* $\{e_j, j \in \mathbb{Z}\}$ with

$$e_j(t) := \exp(-\iota 2\pi j t) \text{ for } t \in [0,1) \text{ and } j \in \mathbb{Z},$$

is orthonormal and complete, i.e. an ONB. □

§2.1.7 **Properties.**

(Pythagorean formula) *If $h_1, \dots, h_n \in \mathbb{H}$ are orthogonal, then $\|\sum_{j=1}^n h_j\|^2_{\mathbb{H}} = \sum_{j=1}^n \|h_j\|^2_{\mathbb{H}}$.*

(Bessel's inequality) *If $\mathcal{U} \subset \mathbb{H}$ is an ONS, then $\|h\|^2_{\mathbb{H}} \geqslant \sum_{u\in\mathcal{U}} |\langle h, u\rangle_{\mathbb{H}}|^2$ for all $h \in \mathbb{H}$.*

(Parseval's formula) *An ONS $\mathcal{U} \subset \mathbb{H}$ is complete if and only if $\|h\|^2_{\mathbb{H}} = \sum_{u\in\mathcal{U}} |\langle h, u\rangle_{\mathbb{H}}|^2$ for all $h \in \mathbb{H}$.* □

§2.1.8 **Definition.** Let $\mathcal{U}$ be a subset of a Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$. Denote by $\mathbb{U} := \overline{lin}(\mathcal{U})$ the closure of the linear subspace spanned by the elements of $\mathcal{U}$ and its orthogonal complement in $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ by $\mathbb{U}^\perp := \big\{h \in \mathbb{H} : \langle h, u\rangle_{\mathbb{H}} = 0, \forall u \in \overline{lin}(\mathcal{U})\big\}$ where $\mathbb{H} = \mathbb{U} \oplus \mathbb{U}^\perp$. □

§2.1.9 **Remark.** If $\mathcal{U} \subset \mathbb{H}$ is an ONS, then there exists an ONS $\mathcal{V} \subset \mathbb{H}$ such that $\mathbb{H} = \overline{lin}(\mathcal{U}) \oplus \overline{lin}(\mathcal{V})$ and for all $h \in \mathbb{H}$ it holds $h = \sum_{u\in\mathcal{U}} \langle h, u\rangle_{\mathbb{H}} u + \sum_{v\in\mathcal{V}} \langle h, v\rangle_{\mathbb{H}} v$ (in a $\mathbb{H}$-sense). In particular, if $\mathcal{U}$ is an ONB then $h = \sum_{u\in\mathcal{U}} \langle h, u\rangle_{\mathbb{H}} u$ for all $h \in \mathbb{H}$. □

§2.1.10 **Definition.** Given $\mathcal{J} \subset \mathbb{Z}$, a sequence $(u_j)_{j\in\mathcal{J}}$ in $\mathbb{H}$ is said to be *orthonormal and complete* (i.e. orthonormal basis) if the subset $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ is a complete ONS (i.e. ONB). The Hilbert space $\mathbb{H}$ is called *separable*, if there exists a complete orthonormal sequence. □

§2.1.11 **Examples.** The Hilbert space $(\mathbb{R}^k, \langle\cdot,\cdot\rangle_M)$, $(\ell^2_{\mathfrak{v}}, \langle\cdot,\cdot\rangle_{\ell^2_{\mathfrak{v}}})$ and $(L^2_\mu(\Omega), \langle\cdot,\cdot\rangle_{L^2_\mu})$ with σ-finite measure μ are separable. On the contrary, given $\lambda \in \mathbb{R}$ define the function $f_\lambda : \mathbb{R} \to \mathbb{C}$

with $f_\lambda(x) := e^{\iota\lambda x}$ and set $\mathcal{H} = \overline{lin}\{f_\lambda, \lambda \in \mathbb{R}\}$. Observe that $\langle f, g\rangle = \lim_{t\to\infty} \frac{1}{2t}\int_{-t}^{t} f(s)\overline{g(s)}ds$ defines an inner product on $\mathcal{H}$. The completion of $\mathcal{H}$ w.r.t. the induced norm $\|f\| = |\langle f, f\rangle|^{1/2}$ is a Hilbert space which is not separable, since $\|f_\lambda - f_{\lambda'}\| = \sqrt{2}$ for all $\lambda \neq \lambda'$. □

§2.1.12 **Definition**. Given $\mathcal{J} \subseteq \mathbb{Z}$ we call a (possibly finite) sequence $(\mathcal{J}_m)_{m\in\mathcal{M}}$, $\mathcal{M} \subseteq \mathbb{N}$, a *nested sieve in $\mathcal{J}$*, if (i) $\mathcal{J}_k \subset \mathcal{J}_m$, for any $k \in [\![1, m]\!] \cap \mathcal{M}$ and $m \in \mathcal{M}$, (ii) $|\mathcal{J}_m| < \infty$, $m \in \mathcal{M}$, and (iii) $\cup_{m\in\mathcal{M}}\mathcal{J}_m = \mathcal{J}$. We write $\mathcal{J}_m^c := \mathcal{J}\backslash\mathcal{J}_m$, $m \in \mathcal{M}$. Denoting $[\![a, b]\!] := [a, b] \cap \mathbb{Z}$ (analogously, $]\!]a, b]\!] :=]a, b] \cap \mathbb{Z}$, $[\![a, b[\![:= [a, b[\cap\mathbb{Z}$, etc.) we use typically the nested sieve $([\![1, m]\!])_{m\in\mathbb{N}}$ and $([\![-m, m]\!])_{m\in\mathbb{N}}$ in $\mathcal{J} = \mathbb{N}$ and $\mathcal{J} = \mathbb{Z}$, respectively. Analogously, given an ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ and setting $\mathbb{U}_m := \overline{lin}\{u_j, j \in \mathcal{J}_m\}$, $m \in \mathcal{M}$, for a nested sieve $(\mathcal{J}_m)_{m\in\mathcal{M}}$ in $\mathcal{J}$ we call the (possibly finite) sequence $(\mathbb{U}_m)_{m\in\mathcal{M}}$ *a nested sieve in $\mathbb{U} := \overline{lin}\{u_j, j \in \mathcal{J}\}$*. We write $\mathbb{U}_m^\perp := \overline{lin}\{u_j, j \in \mathcal{J}_m^c\}$ where $\mathbb{U} = \mathbb{U}_m \oplus \mathbb{U}_m^\perp$. For convenient notations we set further $\mathbb{1}_{\mathcal{J}_m} := (\mathbb{1}_{\mathcal{J}_m}(j))_{j\in\mathcal{J}}$ with $\mathbb{1}_{\mathcal{J}_m}(j) = 1$ if $j \in \mathcal{J}_m$ and $\mathbb{1}_{\mathcal{J}_m}(j) = 0$ otherwise, and analogously $\mathbb{1}_{\mathcal{J}_m^c} := (\mathbb{1}_{\mathcal{J}_m^c}(j))_{j\in\mathcal{J}}$. □

§2.1.13 **Definition**. We call an ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in L^2_μ (respectively, in ℓ^2)

(i) *regular w.r.t. a nested sieve $(\mathcal{J}_m)_{m\in\mathcal{M}}$ in $\mathcal{J}$ and a weight sequence* $\mathfrak{v}$ if there is a finite constant $\tau_{u\mathfrak{v}} \geqslant 1$ satisfying $\|\sum_{j\in\mathcal{J}_m} \mathfrak{v}_j^2|u_j|^2\|_{L^\infty_\mu} \leqslant \tau_{u\mathfrak{v}}^2 \sum_{j\in\mathcal{J}_m} \mathfrak{v}_j^2$ for all $m \in \mathcal{M}$;

(ii) *regular w.r.t. a weight sequence* $\mathfrak{a}$ if there exists a finite constant $\tau_{u\mathfrak{a}} \geqslant 1$ such that $\|\sum_{j\in\mathcal{J}} \mathfrak{a}_j^2|u_j|^2\|_{L^\infty_\mu} \leqslant \tau_{u\mathfrak{a}}^2$. □

§2.1.14 **Remark**. According to Lemma 6 of Birgé and Massart [1997] assuming in L^2 a regular ONS $\{u_j, j \in \mathbb{N}\}$ w.r.t. the nested sieve $([\![1, m]\!])_{m\in\mathbb{N}}$ and $\mathfrak{v} \equiv 1$ is exactly equivalent to following property: there exists a finite constant $\tau_u \geqslant 1$ such that for any h belonging to the subspace $\mathbb{U}_m$, spanned by the first m functions $\{u_j\}_{j=1}^m$, holds $\|h\|_{L^\infty} \leqslant \tau_u\sqrt{m}\,\|h\|_{L^2}$. Typical example are bounded basis, such as the trigonometric basis, or basis satisfying the assertion, that there exists a positive constant C_∞ such that for any $(c_1, \ldots, c_m) \in \mathbb{R}^m$, $\|\sum_{j=1}^m c_ju_j\|_{L^\infty} \leqslant C_\infty\sqrt{m}|c|_\infty$ where $|c|_\infty = \max_{1\leqslant j\leqslant m} c_j$. Birgé and Massart [1997] have shown that the last property is satisfied for piece-wise polynomials, splines and wavelets. □

§2.1.15 **Example** *(§2.1.6 (i) continued)*. Consider the *trigonometric basis* $\{\psi_j, j \in \mathbb{N}\}$ in the real Hilbert space $L^2([0, 1])$. Since $\sup_{j\in\mathbb{N}} \|\psi_j\|_{L^\infty} \leqslant \sqrt{2}$ setting $\tau_{\psi\mathfrak{v}}^2 := 2$ the trigonometric basis is regular w.r.t. any nested Sieve $(\mathcal{J}_m)_{m\in\mathcal{M}}$ and sequence $\mathfrak{v}$, i.e., §2.1.13 (i) holds with $\|\sum_{j\in\mathcal{J}_m} \mathfrak{v}_j^2|\psi_j|^2\|_{L^\infty} \leqslant \tau_{\psi\mathfrak{v}}^2 \sum_{j\in\mathcal{J}_m} \mathfrak{v}_j^2$. In the particular case of the nested sieve $([\![1, 1 + 2m]\!])_{m\in\mathbb{N}}$ and $\mathfrak{v} \equiv 1$, we have $\sum_{j=1}^{1+2m} |\psi_j|^2 = \mathbb{1}_{[0,1]} + \sum_{j=1}^m\{2\sin^2(2\pi j\bullet) + 2\cos^2(2\pi j\bullet)\} = 1 + 2m$ and thus, the trigonometric basis is regular with $\tau_\psi^2 := 1$. Moreover, the trigonometric basis is regular w.r.t. any square-summable weight sequence $\mathfrak{a}$, i.e., $\|\mathfrak{a}\|_{\ell^2} < \infty$. Indeed, in this situation we have $\|\sum_{j\in\mathbb{N}} \mathfrak{a}_j^2|\psi_j|^2\|_{\ell_\infty} \leqslant 2\,\|\mathfrak{a}\|_{\ell^2}^2$ and hence §2.1.13 holds with $\tau_{\psi\mathfrak{a}}^2 = 2\,\|\mathfrak{a}\|_{\ell^2}^2$. □

2.1.1 Abstract smoothness condition

§2.1.16 **Notations**. Let $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ be an ONS with $\mathbb{U} = \overline{lin}\{u_j, j \in \mathcal{J}\} \subseteq \mathbb{H}$. For any $h \in \mathbb{H}$ consider its associated sequence of generalised Fourier coefficients $[h] := ([h]_j)_{j\in\mathcal{J}}$ with generic elements $[h]_j = \langle h, u_j\rangle_{\mathbb{H}}$, $j \in \mathcal{J}$. Given a strictly positive sequence of weights $\mathfrak{v} = (\mathfrak{v}_j)_{j\in\mathcal{J}}$ for $h, g \in \mathbb{H}$ we define $\langle h, g\rangle_{\mathfrak{v}}^2 := \langle\mathfrak{v}[h], \mathfrak{v}[g]\rangle_{\ell^2} = \sum_{j\in\mathcal{J}} \mathfrak{v}_j^2[h]_j\overline{[g]_j}$ and $\|h\|_{\mathfrak{v}}^2 := \|[h]\|_{\ell^2_{\mathfrak{v}}}^2 = \sum_{j\in\mathcal{J}} \mathfrak{v}_j^2|[h]_j|^2$. Obviously, $\langle\cdot, \cdot\rangle_{\mathfrak{v}}$ and $\|\cdot\|_{\mathfrak{v}}$ restricted on $\mathbb{U}$ defines on $\mathbb{U}$ a (weighted) *inner product* and its induced (weighted) *norm*, respectively. We denote by $\mathbb{U}_{\mathfrak{v}}$ the completion of $\mathbb{U}$ w.r.t. $\|\cdot\|_{\mathfrak{v}}$. If $(u_j)_{j\in\mathcal{J}}$ is complete in $\mathbb{H}$ then let $\mathbb{H}_{\mathfrak{v}}$ be the completion of $\mathbb{H}$ w.r.t. $\|\cdot\|_{\mathfrak{v}}$. □

§2.1.17 **Example** *(§2.1.15 continued)*. Consider the real Hilbert space $L^2([0,1])$ and the *trigonometric basis* $\{\psi_j, j\in\mathbb{N}\}$. Define further a weighted norm $\|\cdot\|_{\mathfrak{v}}$ w.r.t. the trigonometric basis, that is, $\|h\|_{\mathfrak{v}} := \sum_{j\in\mathbb{N}} \mathfrak{v}_j^2 |\langle h, \psi_j\rangle_{L^2}|^2$. Denote by $L^2_{\mathfrak{v}}([0,1])$ or $L^2_{\mathfrak{v}}$ for short, the completion of $L^2([0,1])$ w.r.t. $\|\cdot\|_{\mathfrak{v}}$.

(P) If we set $\mathfrak{v}_1 = 1$, $\mathfrak{v}_{2k} = \mathfrak{v}_{2k+1} = j^p$, $p\in\mathbb{N}$, $k\in\mathbb{N}$, then $L^2_{\mathfrak{v}}([0,1])$ is a subset of the *Sobolev space* of p-times differentiable periodic functions. Moreover, up to a constant, for any function $h\in L^2_{\mathfrak{v}}([0,1])$, the weighted norm $\|h\|^2_{\mathfrak{v}}$ equals the L^2-norm of its p-th weak derivative $h^{(p)}$ (Tsybakov [2009]).

(E) If, on the contrary, $\mathfrak{v}_j = \exp(-1 + j^{2p})$, $p > 1/2$, $j\in\mathbb{N}$, then $L^2_{\mathfrak{v}}([0,1])$ is a *class of analytic functions* (Kawata [1972]).

Note that, the trigonometric basis is regular w.r.t. the weight sequence $1/\mathfrak{v} = \mathfrak{v}^{-1} = (\mathfrak{v}_j^{-1})$ as in §2.1.13 (ii), i.e., $\|1/\mathfrak{v}\|_{\ell^2} < \infty$, in case **(P)** whenever $p > 1/2$ and in case **(E)** if $p > 0$. □

§2.1.18 **Definition** *(Abstract smoothness condition)*. Given a strictly positive sequence of weights $\mathfrak{a} = (\mathfrak{a}_j)_{j\in\mathcal{J}}$ and an ONS $\mathcal{U} = \{u_j, j\in\mathcal{J}\}$ in $\mathbb{H}$ consider the associated weighted norm $\|\cdot\|_{1/\mathfrak{a}}$ and the completion $\mathbb{U}_{1/\mathfrak{a}}$ of $\mathbb{U}$. Let $r > 0$ be a constant. We assume in the following that the function of interest f belongs to the ellipsoid $\mathbb{F}^r_{u\mathfrak{a}} := \{h \in \mathbb{U}_{1/\mathfrak{a}} : \|h\|^2_{1/\mathfrak{a}} \leqslant r^2\}$ and hence, $\Pi_{\mathbb{U}^\perp} f = 0$. □

§2.1.19 **Lemma**. *Let $\mathbb{F}^r_{u\mathfrak{a}}$ be a class of functions w.r.t. an ONS $\mathcal{U} = \{u_j, j\in\mathcal{J}\}$ in L^2_μ (or analogously in ℓ^2) as given in §2.1.18. If the ONS is regular w.r.t. the weight sequence $\mathfrak{a}$ as in §2.1.13 (ii) for some finite constant $\tau_{u\mathfrak{a}} \geqslant 1$, then for each $f\in\mathbb{F}^r_{u\mathfrak{a}}$ holds $\|f\|_{L^\infty_\mu} \leqslant \tau_{u\mathfrak{a}} \|f\|_{1/\mathfrak{a}} \leqslant r\tau_{u\mathfrak{a}}$.*

Proof of Lemma §2.1.19 is given in the lecture. □

§2.1.20 **Example** *(§2.1.17 continued)*. Consider in $L^2_{\mathfrak{v}}([0,1])$ the *trigonometric basis* $\{\psi_j, j\in\mathbb{N}\}$ and a weight sequence $\mathfrak{v}$ satisfying either §2.1.17 **(P)** with $p > 1/2$ or §2.1.17 **(E)** with $p > 0$. In both cases setting $\tau^2_{\psi\mathfrak{v}} = 2\,\|1/\mathfrak{v}\|^2_{\ell^2} < \infty$ the trigonometric basis is regular w.r.t. the weight sequence $1/\mathfrak{v}$. Consequently, setting $\mathfrak{a} = 1/\mathfrak{v}$ and $\mathbb{F}^r_{\psi\mathfrak{a}} = \{h\in L^2_{\mathfrak{v}}([0,1]) : \|h\|^2_{\mathfrak{v}} \leqslant r^2\}$, from Lemma §2.1.19 follows $\|f\|^2_{L^\infty} \leqslant 2\,\|f\|^2_{\mathfrak{v}}\,\|1/\mathfrak{v}\|^2_{\ell^2}$ for all $f\in\mathbb{F}^r_{\psi\mathfrak{a}}$. □

2.2 Linear operator between Hilbert spaces

§2.2.1 **Definition**. A map $T:\mathbb{H}\to\mathbb{G}$ between Hilbert spaces $\mathbb{H}$ and $\mathbb{G}$ is called *linear operator* if $T(ah_1 + bh_2) = aTh_1 + bTh_2$ for all $h_1, h_2\in\mathbb{H}$, $a, b\in\mathbb{K}$. Its *domain* will be denoted by $\mathcal{D}(T)$, its *range* by $\mathcal{R}(T)$ and its *null space* by $\mathcal{N}(T)$. □

§2.2.2 **Property**. *Let $T:\mathbb{H}\to\mathbb{G}$ be a linear operator, then the following assertions are equivalent:* (i) *T is continuous in zero.* (ii) *T is bounded, i.e., there is $M > 0$ such that $\|Th\|_{\mathbb{G}} \leqslant M\,\|h\|_{\mathbb{H}}$ for all $h\in\mathbb{H}$.* (iii) *T is uniformly continuous.* □

§2.2.3 **Definition**. The *class of all bounded linear operators* $T:\mathbb{H}\to\mathbb{G}$ is denoted by $\mathscr{L}(\mathbb{H},\mathbb{G})$, or $\mathscr{L}$ and in case of $\mathbb{H} = \mathbb{G}$, $\mathscr{L}(\mathbb{H})$ for short. For $T\in\mathscr{L}(\mathbb{H},\mathbb{G})$ define its *(uniform)* *norm* as $\|T\|_{\mathscr{L}} := \|T\|_{\mathscr{L}(\mathbb{H},\mathbb{G})} := \sup\{\|Th\|_{\mathbb{G}} : \|h\|_{\mathbb{H}} \leqslant 1, h \in \mathbb{H}\}$. □

§2.2.4 **Examples.**

(i) Let M be a $(m \times k)$ matrix, then $M \in \mathscr{L}(\mathbb{R}^k, \mathbb{R}^m)$. We write $\|M\|_s := \|M\|_{\mathscr{L}(\mathbb{R}^k,\mathbb{R}^m)}$ for short. *(spectral norm)*

(ii) For finite (i.e., $|\mathcal{J}| < \infty$) sequences $(h)_{j\in\mathcal{J}}$ in $\mathbb{H}$ and $(g)_{j\in\mathcal{J}}$ in $\mathbb{G}$ the linear operator $\sum_{j\in\mathcal{J}} h_j \otimes g_j$ defined by $f \mapsto \big[\sum_{j\in\mathcal{J}} h_j \otimes g_j\big] f := \sum_{j\in\mathcal{J}} \langle f, h_j\rangle_{\mathbb{H}}\, g_j$ belongs to $\mathscr{L}(\mathbb{H},\mathbb{G})$ with $\|\sum_{j\in\mathcal{J}} h_j \otimes g_j\|_{\mathscr{L}} \leqslant \sum_{j\in\mathcal{J}} \|h_j\|_{\mathbb{H}} \|g_j\|_{\mathbb{G}}$. Moreover, it has a finite range contained in $\overline{lin}(\{g_j, j \in \mathcal{J}\})$.

(iii) Let $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ be an ONS in $\mathbb{H}$ and for any $f \in \mathbb{H}$ consider its *sequence of generalised Fourier coefficients* $[f] := ([f]_j)_{j\in\mathcal{J}}$ given by $[f]_j := \langle f, u_j\rangle_{\mathbb{H}}$, $j \in \mathcal{J}$. The associated *(generalised) Fourier series transform* U defined by $f \mapsto Uf := [f]$ belongs to $\mathscr{L}(\mathbb{H}, \ell^2(\mathcal{J}))$ with $\|U\|_{\mathscr{L}} = 1$.

(iv) For a sequence $\lambda = (\lambda_j)_{j\in\mathcal{J}}$ consider the *multiplication operator* $\mathrm{M}_\lambda : \mathbb{K}^{\mathcal{J}} \to \mathbb{K}^{\mathcal{J}}$ given by $x \mapsto \mathrm{M}_\lambda x := (\lambda_j x_j)_{j\in\mathcal{J}}$. For any bounded sequence λ, i.e, $\|\lambda\|_{\ell^\infty} < \infty$, we have $\|\mathrm{M}_\lambda\|_{\mathscr{L}(\ell^p)} \leqslant \|\lambda\|_{\ell^\infty}$ and hence, $\mathrm{M}_\lambda \in \mathscr{L}(\ell^p)$ for any $p \in [1,\infty]$. Analogously, given a function $\lambda : \Omega \to \mathbb{K}$ the *multiplication operator* $\mathrm{M}_\lambda : \mathbb{K}^\Omega \to \mathbb{K}^\Omega$ is defined as $f \mapsto \mathrm{M}_\lambda f := f\lambda$ where for any bounded (measurable) function λ, i.e, $\|\lambda\|_{L^\infty_\mu} < \infty$, holds $\|\mathrm{M}_\lambda\|_{\mathscr{L}(L^p_\mu)} \leqslant \|\lambda\|_{L^\infty_\mu} < \infty$ and, hence $\mathrm{M}_\lambda \in \mathscr{L}(L^p_\mu)$. On the other hand side, if λ is real-valued (measurable), μ-a.s. finite and non zero, then the subset $\mathcal{D}(\mathrm{M}_\lambda) := \big\{f \in L^2_\mu : \lambda f \in L^2_\mu\big\}$ is dense in L^2_μ. In this situation the *multiplication operator* $\mathrm{M}_\lambda : L^2_\mu \supset \mathcal{D}(\mathrm{M}_\lambda) \to L^2_\mu$ is densely defined (and self-adjoint).

(v) Given a (generalised) Fourier series transform $U \in \mathscr{L}(\mathbb{H}, \ell^2)$ as in (iii) and a multiplication operator $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2)$ for some bounded sequence $\lambda = (\lambda_j)_{j\in\mathcal{J}}$ as in (iv) the linear operator $\nabla_\lambda : \mathbb{H} \to \mathbb{H}$ given by $\mathcal{N}(U) = \mathcal{N}(\nabla_\lambda)$ and $U\nabla_\lambda = \mathrm{M}_\lambda U$, i.e. $U\nabla_\lambda h = \mathrm{M}_\lambda U h = (\lambda_j [h]_j)_{j\in\mathcal{J}}$ belongs to $\mathscr{L}(\mathbb{H})$ with $\|\nabla_\lambda\|_{\mathscr{L}} \leqslant \|\lambda\|_{\ell^\infty} < \infty$. We call ∇_λ *diagonal* w.r.t. U (or $\mathcal{U}$).

(vi) The *integral operator* $T_k : L^2_{\mu_1}(\Omega_1) \to L^2_{\mu_2}(\Omega_2)$ with kernel $k : \Omega_1 \times \Omega_2 \to \mathbb{K}$ defined by

$$[T_k f](\omega_2) := \int_{\Omega_1} h(\omega_1) k(\omega_1, \omega_2) \mu(d\omega_1),\ \omega_2 \in \Omega_2,\ h \in L^2_{\mu_1}(\Omega_1),$$

belongs to $\mathscr{L}(L^2_{\mu_1}(\Omega_1), L^2_{\mu_2}(\Omega_2))$ if $\|k\|^2_{L^2} = \int_{\Omega_1}\int_{\Omega_2} |k|^2 d\mu_1 d\mu_2 < \infty$.

(vii) Let $X \in L^1(\Omega, \mathscr{A}, \mathbb{P})$ and $\mathscr{F} \subset \mathscr{A}$ be a sub-σ-algebra. There exists $Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ such that $\mathbb{E}(X\mathbb{1}_F) = \mathbb{E}(Y\mathbb{1}_F)$ for all $F \in \mathscr{F}$, moreover, Y is unique up to equality $\mathbb{P}$-a.s.. Each version Y is called conditional expectation of X given $\mathscr{F}$, symbolically, $\mathbb{E}[X|\mathscr{F}] := Y$. For each $p \in [1,\infty]$ the linear map $\mathbb{E}[\bullet|\mathscr{F}] : L^p(\Omega, \mathscr{A}, \mathbb{P}) \to L^p(\Omega, \mathscr{F}, \mathbb{P}) \subseteq L^p(\Omega, \mathscr{A}, \mathbb{P})$ given by $X \mapsto \mathbb{E}[X|\mathscr{F}]$ is a contraction, that is $\|\mathbb{E}[X|\mathscr{F}]\|_{L^p} \leqslant \|X\|_{L^p}$ and thus $\mathbb{E}[\bullet|\mathscr{F}]$ belongs to $\mathscr{L}(L^p(\Omega, \mathscr{A}, \mathbb{P}))$ with $\|\mathbb{E}[\bullet|\mathscr{F}]\|_{\mathscr{L}} = 1$ (keep in mind that $\mathbb{E}[1|\mathscr{F}] = 1$). Given a r.v. Z on $(\Omega, \mathscr{A}, \mathbb{P})$ and the σ-algebra $\sigma(Z)$ generated by Z we set $\mathbb{E}[X|Z] := \mathbb{E}[X|\sigma(Z)]$. The *conditional expectation operator* of X given Z defined by $Kh := \mathbb{E}[h(X)|Z]$ for $h \in L^1_X$ is then an element of $\mathscr{L}(L^p_X, L^p_Z)$ with $\|K\|_{\mathscr{L}} = 1$.

(viii) Let $\phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then the *convolution operator* $C_\phi : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ defined by

$$[C_\phi h](t) := [h * \phi](t) := \int_{\mathbb{R}} h(s)\phi(t-s)ds,\ t \in \mathbb{R},\ h \in L^2(\mathbb{R}),$$

belongs to $\mathscr{L}(L^2(\mathbb{R}))$ with $\|C_\phi\|_{\mathscr{L}} \leqslant \|\phi\|_{L^1} := \int_{\mathbb{R}} |\phi(t)|dt$.

(ix) Let $\phi \in L^2([0,1))$, hence, $\phi \in L^1([0,1))$, and let $\lfloor\cdot\rfloor$ be the floor function, then the *circular convolution operator* $C_\phi : L^2([0,1)) \to L^2([0,1))$ defined by

$$[C_\phi h](t) := [h \circledast \phi](t) := \int_{[0,1)} h(s)\phi(t-s-\lfloor t-s\rfloor)ds,\ t \in [0,1),\ h \in L([0,1)),$$

belongs to $\mathscr{L}(L^2([0,1)))$ with $\|C_\phi\|_{\mathscr{L}} \leqslant \|\phi\|_{L^1} := \int_0^1 |\phi(t)|dt$. □

§2.2.5 **Definition.** A (linear) map $\Phi : \mathbb{H} \supset \mathcal{D}(\Phi) \to \mathbb{K}$ is called *(linear) functional* and given an ONS $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ which belongs to $\mathcal{D}(\Phi)$ we set $[\Phi] = ([\Phi]_j)_{j\in\mathcal{J}}$ with the slight abuse of notations $[\Phi]_j := \Phi(u_j)$. In particular, if $\Phi \in \mathscr{L}(\mathbb{H}, \mathbb{K})$ then $\mathcal{D}(\Phi) = \mathbb{H}$. □

§2.2.6 **Property.** *Let $\Phi \in \mathscr{L}(\mathbb{H}, \mathbb{K})$.*

(Fréchet-Riesz representation) *There exists a function $\phi \in \mathbb{H}$ such that $\Phi(h) = \langle \phi, h\rangle_{\mathbb{H}}$ for all $h \in \mathbb{H}$, and hence, given an ONS $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ we have $[\Phi]_j = [\phi]_j$ for all $j \in \mathcal{J}$.* □

§2.2.7 **Example.** Consider an ONB $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $L^2(\Omega)$ (or analogously in $\ell^2(\mathcal{J})$). By *evaluation at a point* $t_o \in \Omega$ we mean the linear functional Φ_{t_o} mapping $h \in L^2(\Omega)$ to $h(t_o) := \Phi_{t_o}(h) = \sum_{j\in\mathcal{J}}[h]_j u_j(t_o)$. Obviously, a point evaluation of h at t_o is well-defined, if $\sum_{j\in\mathcal{J}} |[h]_j u_j(t_o)| < \infty$. Observe that the point evaluation at t_o is generally not bounded on the subset $\{h \in L^2(\Omega) : \sum_{j\in\mathcal{J}} |[h]_j u_j(t_o)| < \infty\}$. □

§2.2.8 **Definition** *(Regular linear functionals).* Consider an ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ which belongs to the domain $\mathcal{D}(\Phi)$ of a linear functional Φ. In order to guarantee that $\mathbb{U}_{1/\mathfrak{a}}$ and hence the class $\mathbb{F}^r_{u\mathfrak{a}}$ of functions of interest as in §2.1.18 are contained in $\mathcal{D}(\Phi)$ and that $\Phi(f) = \sum_{j\in\mathcal{J}}[\Phi]_j[f]_j$ holds for all $f \in \mathbb{F}^r_{u\mathfrak{a}}$, it is sufficient that $\|[\Phi]\|^2_{\ell^2_\mathfrak{a}} = \sum_{j\in\mathcal{J}} |[\Phi]_j|^2\mathfrak{a}_j^2 < \infty$. Indeed, $|\Phi(f)|^2 \leqslant \|f\|^2_{1/\mathfrak{a}}\|[\Phi]\|^2_{\ell^2_\mathfrak{a}}$ for any $f \in \mathbb{U}_{1/\mathfrak{a}}$ and hence $\Phi \in \mathscr{L}(\mathbb{U}_{1/\mathfrak{a}}, \mathbb{K})$ with $\|\Phi\|_{\mathscr{L}} \leqslant \|[\Phi]\|_{\ell^2_\mathfrak{a}}$. We denote by $\mathcal{L}_\mathfrak{a}$ the set of all linear functionals with $\|[\Phi]\|^2_{\ell^2_\mathfrak{a}} < \infty$. □

§2.2.9 **Remark.** We may emphasise that we neither impose that the sequence $[\Phi] = ([\Phi]_j)_{j\in\mathcal{J}}$ tends to zero nor that it is square summable. The assumption $\Phi \in \mathcal{L}_\mathfrak{a}$, however, enables us in specific cases to deal with more demanding functionals, such as in Example §2.2.7 above the evaluation at a given point. □

§2.2.10 **Example** *(§2.2.7 continued).* Consider an ONB $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $L^2(\Omega)$ and the *evaluation at a point* $t_o \in \Omega$ given by $\Phi_{t_o}(h) = \sum_{j\in\mathcal{J}}[h]_j u_j(t_o)$. Let $L^2_{1/\mathfrak{a}}(\Omega)$ be the completion of $L^2(\Omega)$ w.r.t. a weighted norm $\|\cdot\|_{1/\mathfrak{a}}$ derived from $\mathcal{U}$ and a strictly positive sequence $\mathfrak{a}$. Since $|\Phi_{t_o}(h)|^2 \leqslant \|h\|^2_{1/\mathfrak{a}} \sum_{j\in\mathcal{J}} \mathfrak{a}_j^2|u_j(t_o)|^2$ the point evaluation in t_o is bounded on $L^2_{1/\mathfrak{a}}(\Omega)$ and, thus, belongs to $\mathscr{L}(L^2_{1/\mathfrak{a}}(\Omega), \mathbb{K})$, if $\sum_{j\in\mathcal{J}} \mathfrak{a}_j^2|u_j(t_o)|^2 < \infty$. Consequently, if the ONS $\mathcal{U}$ is regular w.r.t. the weight sequence $\mathfrak{a}$, i.e., §2.1.13 (ii) holds for some finite constant $\tau_{u\mathfrak{a}} \geqslant 1$, then $\|\Phi_{t_o}\|_{\mathscr{L}(L^2_{1/\mathfrak{a}}(\Omega),\mathbb{K})} \leqslant \tau_{u\mathfrak{a}}$ uniformly for any $t_o \in \Omega$. Revisiting the particular situation of Example §2.1.17 and its continuation in §2.1.20, that is, $L^2_\mathfrak{v}([0,1])$ w.r.t. the *trigonometric basis* $\{\psi_j, j \in \mathbb{N}\}$ and weight sequence $\mathfrak{v}$ satisfying either §2.1.17 **(P)** with $p > 1/2$ or §2.1.17 **(E)** with $p > 0$, recall that the trigonometric basis is regular w.r.t. $\mathfrak{a} = 1/\mathfrak{v}$ and hence, the point evaluation Φ_{t_o} belongs to $\mathscr{L}(L^2_\mathfrak{v}([0,1]), \mathbb{R})$, i.e., $\|\Phi_{t_o}\|_{\mathscr{L}} \leqslant \sqrt{2}\,\|1/\mathfrak{v}\|_{\ell^2}$ for each $t_o \in [0,1]$. □

§2.2.11 **Definition**. If $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$, then there exists a uniquely determined *adjoint operator* $T^\star \in \mathscr{L}(\mathbb{G}, \mathbb{H})$ satisfying $\langle Th, g\rangle_{\mathbb{G}} = \langle h, T^\star g\rangle_{\mathbb{H}}$ for all $h \in \mathbb{H}, g \in \mathbb{G}$. □

§2.2.12 **Properties**. *Let $S, T \in \mathscr{L}(\mathbb{H}_1, \mathbb{H}_2)$ and $R \in \mathscr{L}(\mathbb{H}_2, \mathbb{H}_3)$. Then we have*

(i) $(S + T)^\star = S^\star + T^\star$, $(RS)^\star = S^\star R^\star$.

(ii) $\|S^\star\|_{\mathscr{L}} = \|S\|_{\mathscr{L}}$, $\|SS^\star\|_{\mathscr{L}} = \|S^\star S\|_{\mathscr{L}} = \|S\|^2_{\mathscr{L}}$.

(iii) $\mathcal{N}(S) = \mathcal{R}(S^\star)^\perp$, $\mathcal{N}(S^\star) = \mathcal{R}(S)^\perp$, $\mathbb{H}_1 = \mathcal{N}(S) \oplus \overline{\mathcal{R}}(S^\star)$ *and* $\mathbb{H}_2 = \mathcal{N}(S^\star) \oplus \overline{\mathcal{R}}(S)$ *where $\overline{\mathcal{R}}(S)$ (respectively, $\overline{\mathcal{R}}(S^\star)$) denotes the closure of the range of S. In particular, S is injective if and only if $\mathcal{R}(S^\star)$ is dense in $\mathbb{H}$.*

(iv) $\mathcal{N}(S^\star S) = \mathcal{N}(S)$ *and* $\mathcal{N}(SS^\star) = \mathcal{N}(S^\star)$. □

§2.2.13 **Examples** *(§2.2.4 continued).*

(i) The adjoint of a $(k \times m)$ matrix M is its $(m \times k)$ transpose matrix M^t.

(ii) The adjoint $U^\star \in \mathscr{L}(\ell^2(\mathcal{J}), \mathbb{H})$ of the *(generalised) Fourier series transform* $U \in \mathscr{L}(\mathbb{H}, \ell^2(\mathcal{J}))$ satisfies $x \mapsto U^\star x := \sum_{j\in\mathcal{J}} x_j u_j$ for $x \in \ell^2(\mathcal{J})$.

(iii) For finite $\mathcal{J}$ the adjoint operator in $\mathscr{L}(\mathbb{G}, \mathbb{H})$ of $\sum_{j\in\mathcal{J}} h_j \otimes g_j \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ satisfies $[\sum_{j\in\mathcal{J}} h_j \otimes g_j]^\star g = \sum_{j\in\mathcal{J}} \langle g, g_j\rangle_{\mathbb{G}} h_j = [\sum_{j\in\mathcal{J}} g_j \otimes h_j]g$.

(iv) Let $\mathrm{M}_\lambda \in \mathscr{L}(L^2_\mu(\Omega))$ (or analogously $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2)$) be a *multiplication operator*, then its adjoint operator $\mathrm{M}^\star_\lambda = \mathrm{M}_{\lambda^\star}$ is a multiplication operator with $\lambda^\star(t) = \overline{\lambda(t)}$, $t \in \Omega$.

(v) Let $T_k \in \mathscr{L}(L^2_{\mu_1}(\Omega_1), L^2_{\mu_2}(\Omega_2))$ be an *integral operator* with kernel k, then its adjoint $T^\star_k = T_{k^\star} \in \mathscr{L}(L^2_{\mu_2}(\Omega_2), L^2_{\mu_1}(\Omega_1))$ is again an integral operator satisfying

$$[T_{k^\star} g](\omega_1) := \int_{\Omega_2} g(\omega_2) k^\star(\omega_2, \omega_1) \mu_2(d\omega_2),\ \omega_1 \in \Omega_1,\ g \in L^2_{\mu_2}(\Omega_2),$$

with kernel $k^\star(\omega_2, \omega_1) := \overline{k(\omega_1, \omega_2)}$, $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$.

(vi) Let $K \in \mathscr{L}(L^2_X, L^2_Z)$ be the *conditional expectation* of X given Z, then its adjoint operator $K^\star = K \in \mathscr{L}(L^2_Z, L^2_X)$ is the conditional expectation of Z given X satisfying $Kg = \mathbb{E}[g(Z)|X]$ for all $g \in L^2_Z$.

(vii) Let $C_g \in \mathscr{L}(L^2(\mathbb{R}))$ be a *convolution operator*, then its adjoint operator $C^\star_g = C_{g^\star}$ is a convolution operator, i.e, $C_{g^\star} h = g^\star * h$, with $g^\star(t) = \overline{g(-t)}$, $t \in \mathbb{R}$. □

§2.2.14 **Definition**.

(i) The *identity* in $\mathscr{L}(\mathbb{H})$ is denoted by $\mathrm{Id}_{\mathbb{H}}$.

(ii) Let $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$. Obviously, $T : \mathcal{N}(T)^\perp \to \mathcal{R}(T)$ is bijective and continuous whereas its *inverse* $T^{-1} : \mathcal{R}(T) \to \mathcal{N}(T)^\perp$ is continuous (i.e. bounded) if and only if $\mathcal{R}(T)$ is closed. In particular, if $T : \mathbb{H} \to \mathbb{G}$ is bijective (invertible) then its inverse $T^{-1} \in \mathscr{L}(\mathbb{G}, \mathbb{H})$ satisfies $\mathrm{Id}_{\mathbb{G}} = TT^{-1}$ and $\mathrm{Id}_{\mathbb{H}} = T^{-1}T$.

(iii) $U \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ is called *unitary*, if U is invertible with $UU^\star = \mathrm{Id}_{\mathbb{G}}$ and $U^\star U = \mathrm{Id}_{\mathbb{H}}$.

(iv) $V \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ is called *partial isometry*, if $V : \mathcal{N}(V)^\perp \to \mathcal{R}(V)$ is unitary.

(v) $T \in \mathscr{L}(\mathbb{H})$ is called *self-adjoint*, if $T = T^\star$, i.e., $\langle Th, g\rangle_{\mathbb{H}} = \langle h, T^\star g\rangle_{\mathbb{H}}$ for all $h, g \in \mathbb{H}$.

(vi) $T \in \mathscr{L}(\mathbb{H})$ is called *normal*, if $TT^\star = T^\star T$, i.e., $\langle Th, Tg\rangle_{\mathbb{H}} = \langle T^\star h, T^\star g\rangle_{\mathbb{H}}$ for all $h, g \in \mathbb{H}$.

(vii) A self-adjoint $T \in \mathscr{L}(\mathbb{H})$ is called *positive semi-definite (non-negative definite)* or $T \geqslant 0$ for short, if $\langle Th, h\rangle_{\mathbb{H}} \geqslant 0$ for all $h \in \mathbb{H}$ and *strictly positive definite* or $T > 0$ for short, if $\langle Th, h\rangle_{\mathbb{H}} > 0$ for all $h \in \mathbb{H}\backslash\{0\}$.

(viii) $\Pi \in \mathscr{L}(\mathbb{H})$ is called *projection* if $\Pi^2 = \Pi$. For $\Pi \neq 0$ are equivalent: (a) Π is an orthogonal projection ($\mathbb{H} = \mathcal{R}(\Pi) \oplus \mathcal{N}(\Pi)$); (b) $\|\Pi\|_{\mathscr{L}} = 1$; (c) Π is non-negative. □

§2.2.15 **Property**. *Let $T \in \mathscr{L}(\mathbb{H})$. If T is invertible, then it is $T^\star$, where $(T^{-1})^\star = (T^\star)^{-1}$. Moreover, if T is normal, then $\|T\|_{\mathscr{L}} = \sup\{|\langle Th, h\rangle_{\mathbb{H}}| : \|h\|_{\mathbb{H}} \leqslant 1, h \in \mathbb{H}\}$.*

(Neumann series) *If $\|T\|_{\mathscr{L}} < 1$, then $\|(\mathrm{Id}_{\mathbb{H}} - T)^{-1}\|_{\mathscr{L}} \leqslant (1 - \|T\|_{\mathscr{L}})^{-1}$.* □

§2.2.16 **Examples** *(§2.2.4 continued).*

(i) The *(generalised) Fourier series transform* U is a partial isometry with adjoint operator $U^\star x = \sum_{j\in\mathcal{J}} x_j u_j$ for $x \in \ell^2(\mathcal{J})$. Moreover, the orthogonal projection $\Pi_{\mathbb{U}}$ onto $\mathbb{U}$ satisfies $\Pi_{\mathbb{U}} f = U^\star U f = \sum_{j\in\mathcal{J}} [f]_j u_j$ for all $f \in \mathbb{H}$. If $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ is complete (i.e. ONB), then U is invertible with $UU^\star = \mathrm{Id}_{\ell^2}$ and $U^\star U = \mathrm{Id}_{\mathbb{H}}$ due to Parseval's formula, and hence U is unitary.

(ii) Let $\mathcal{F} \in \mathscr{L}(L^2(\mathbb{R}))$ denote the *Fourier-Plancherel transform* satisfying

$$[\mathcal{F}h](t) = \int_{\mathbb{R}} h(x) e^{-\iota 2\pi x t} dx, \quad \forall h \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}).$$

Then $\mathcal{F}$ is unitary with $[\mathcal{F}^\star h](t) = \int h(x) e^{\iota 2\pi x t} dx$ for all $h \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. We note further for all $h \in L^1$ that $\|\mathcal{F}h\|_{L^\infty} \leqslant \|h\|_{L^1}$, and that $\mathcal{F}h$ is continuous and tends to zero in infinity. Keeping in mind the convolution defined in Examples §2.2.4 (viii) the convolution theorem states $\mathscr{F}(f * g) = \mathscr{F}f \cdot \mathscr{F}g$ for any $f, g \in L^1(\mathbb{R})$.

(iii) A *multiplication operator* $\mathrm{M}_\lambda \in \mathscr{L}(L^2_\mu)$ is normal. If λ is in addition real, it is self-adjoint and if λ is non-negative, then it is non-negative.

(iv) A *diagonal operator* $\nabla_\lambda \in \mathscr{L}(\mathbb{H})$ w.r.t. a partial isometry $U \in \mathscr{L}(\mathbb{H}, \ell^2)$ satisfies $\nabla_\lambda = U^\star \mathrm{M}_\lambda U$ and it shares the properties of the *multiplication operator* $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2)$.

(v) A *conditional expectation operator* $K \in \mathscr{L}(L^2_X, L^2_Z)$ is an orthogonal projection.

(vi) A *convolution operator* $C_g \in \mathscr{L}(L^2(\mathbb{R}))$ is normal and if g is in addition a real and even ($g(-t) = g(t)$) function, then it is self-adjoint.

(vii) A *circular convolution operator* $C_g \in \mathscr{L}(L^2([0, 1)))$ is normal and if g is in addition a real and even ($g(t) = g(1 - t)$) function, then it is self-adjoint. □

2.2.1 Compact, nuclear and Hilbert-Schmidt operator

§2.2.17 **Definition**. An operator $K \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ is called *compact*, if $\{Kh : \|h\|_{\mathbb{H}} \leqslant 1, h \in \mathbb{H}\}$ is relatively compact in $\mathbb{G}$. We denote by $\mathscr{K}(\mathbb{H}, \mathbb{G})$ the *subset of all compact operator* in $\mathscr{L}(\mathbb{H}, \mathbb{G})$, and we write $\mathscr{K}(\mathbb{H}) = \mathscr{K}(\mathbb{H}, \mathbb{H})$ for short. □

§2.2.18 **Properties**. *Let $K \in \mathscr{L}(\mathbb{H}, \mathbb{G})$.*

(Schauder's theorem) *K is compact, if and only if its adjoint $K^\star \in \mathscr{L}(\mathbb{G}, \mathbb{H})$ is compact.*

If there are $K_j \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ with finite dimensional range for each $j \in \mathbb{N}$ such that $\lim_{j\to\infty} \|K_j - K\|_{\mathscr{L}} = 0$, then K is compact. If in addition $\mathbb{G}$ is separable, then the converse holds also true. □

§2.2.19 **Examples** *(§2.2.4 continued)*.

(i) For finite $\mathcal{J}$ the operator $\sum_{j\in\mathcal{J}} h_j \otimes g_j \in \mathscr{L}(\mathbb{H},\mathbb{G})$ is compact.

(ii) A *multiplication operator* $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2)$ is compact, if λ has either only a finite number of entries not equal to zero or zero is the only accumulation point.

(iii) A *diagonal operator* $\nabla_\lambda = U^\star \mathrm{M}_\lambda U \in \mathscr{L}(\mathbb{H})$ w.r.t. a partial isometry $U \in \mathscr{L}(\mathbb{H},\ell^2)$ is compact if the multiplication operator $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2)$ is compact.

(iv) A *convolution operator* $C_g \in \mathscr{L}(L^2(\mathbb{R}))$ is not compact.

(v) A *circular convolution operator* $C_g \in \mathscr{L}(L^2([0,1)))$ is compact. □

§2.2.20 **Remark**. Every finite linear combination of compact operators is compact, and hence $\mathscr{K}(\mathbb{H},\mathbb{G})$ is a vector space. □

§2.2.21 **Definition**. An operator $T \in \mathscr{L}(\mathbb{H},\mathbb{G})$ is called *nuclear*, if there are sequences $(h_j)_{j\in\mathbb{N}}$ in $\mathbb{H}$ and $(g_j)_{j\in\mathbb{N}}$ in $\mathbb{G}$ with $\sum_{j\in\mathbb{N}} \|h_j\|_{\mathbb{H}} \|g_j\|_{\mathbb{G}} < \infty$ such that $\lim_{n\to\infty} \|\sum_{j=1}^n h_j \otimes g_j - T\|_{\mathscr{L}} = 0$, or $T = \sum_{j\in\mathbb{N}} h_j \otimes g_j$ for short. We denote by $\mathscr{N}(\mathbb{H},\mathbb{G})$ the *subset of all nuclear operator* in $\mathscr{L}(\mathbb{H},\mathbb{G})$, and we write $\mathscr{N}(\mathbb{H}) := \mathscr{N}(\mathbb{H},\mathbb{H})$. Furthermore, let $(f_j)_{j\in\mathbb{N}}$ be any ONB in $\mathbb{H}$ and $T \in \mathscr{N}(\mathbb{H})$, then $\mathrm{tr}(T) := \sum_{j\in\mathbb{N}} \langle T f_j, f_j\rangle_{\mathbb{H}}$ denotes the *trace* of T. □

§2.2.22 **Remark**. We have $\mathscr{N}(\mathbb{H},\mathbb{G}) \subset \mathscr{K}(\mathbb{H},\mathbb{G}) \subset \mathscr{L}(\mathbb{H},\mathbb{G})$. The trace does not depend on the choice of the ONB and is a continuous linear functional on $\mathscr{N}(\mathbb{H})$ with $\|\mathrm{tr}\|_{\mathscr{L}} = 1$. □

§2.2.23 **Properties**. *Let $T \in \mathscr{L}(\mathbb{H},\mathbb{G})$ and $S \in \mathscr{L}(\mathbb{G},\mathbb{H})$.*

(i) *T is nuclear, if and only if its adjoint $T^\star \in \mathscr{L}(\mathbb{G},\mathbb{H})$ is nuclear.*

(ii) *If T is nuclear, then $TS \in \mathscr{N}(\mathbb{H})$, $ST \in \mathscr{N}(\mathbb{G})$ and $\mathrm{tr}(TS) = \mathrm{tr}(ST)$.* □

§2.2.24 **Example**. A *multiplication operator* $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2)$ and, hence an associated *diagonal operator* $U^\star \mathrm{M}_\lambda U \in \mathscr{L}(\mathbb{H})$, is nuclear, if λ is absolute summable, i.e., $\|\lambda\|_{\ell^1} < \infty$, and $\mathrm{tr}(\mathrm{M}_\lambda) = \mathrm{tr}(\nabla_\lambda) = \sum_{j\in\mathcal{J}} \lambda_j$. □

§2.2.25 **Definition**. An operator $T \in \mathscr{L}(\mathbb{H},\mathbb{G})$ is called *Hilbert-Schmidt*, if there exists an ONB $(h_j)_{j\in\mathbb{N}}$ in $\mathbb{H}$ such that $\|T\|_{\mathscr{H}}^2 := \sum_{j\in\mathbb{N}} \|T h_j\|_{\mathbb{G}}^2 < \infty$. The number $\|T\|_{\mathscr{H}}$ is called Hilbert-Schmidt norm of T and satisfies $\|T\|_{\mathscr{L}} \leqslant \|T\|_{\mathscr{H}}$. We denote by $\mathscr{H}(\mathbb{H},\mathbb{G})$ the *subset of all Hilbert-Schmidt operator* in $\mathscr{L}(\mathbb{H},\mathbb{G})$, and we write $\mathscr{H}(\mathbb{H}) := \mathscr{H}(\mathbb{H},\mathbb{H})$. □

§2.2.26 **Remark**. We have $\mathscr{N}(\mathbb{H},\mathbb{G}) \subset \mathscr{H}(\mathbb{H},\mathbb{G}) \subset \mathscr{K}(\mathbb{H},\mathbb{G})$. The number $\|T\|_{\mathscr{H}}$ does not depend on the choice of the ONB. The product TS of two Hilbert-Schmidt operator T and S is nuclear. The space $\mathscr{H}(\mathbb{H},\mathbb{G})$ endowed with the inner product $\langle T, S\rangle_{\mathscr{H}} := \mathrm{tr}(S^\star T)$, $S, T \in \mathscr{H}(\mathbb{H},\mathbb{G})$ is a Hilbert space and $\|\cdot\|_{\mathscr{H}}$ the induced norm. □

§2.2.27 **Property**. *If $T \in \mathcal{H}(\mathbb{H},\mathbb{G})$ and $S \in \mathscr{L}(\mathbb{G})$ then $\mathrm{tr}(TST^\star) \leqslant \mathrm{tr}(TT^\star)\,\|S\|_{\mathscr{L}}$.* □

§2.2.28 **Examples**.

(i) Let $T \in \mathscr{L}(L^2_{\mu_1}(\Omega_1), L^2_{\mu_2}(\Omega_2))$. The operator T is Hilbert-Schmidt if and only if it is an *integral operator* $T = T_k$ with square integrable kernel k and it holds $\|T\|_{\mathscr{H}}^2 = \int_{\Omega_1}\int_{\Omega_2} |k(\omega_1,\omega_2)|^2 \mu_1(d\omega_1)\mu_2(d\omega_2)$.

(ii) A *multiplication operator* $\mathrm{M}_\lambda \in \mathscr{L}(\ell(\mathcal{J}))$ and, hence an associated *diagonal operator* $U^\star \mathrm{M}_\lambda U \in \mathscr{L}(\mathbb{H})$, is Hilbert-Schmidt, if $\lambda = (\lambda_j)_{j\in\mathcal{J}}$ is square summable and $\|\mathrm{M}_\lambda\|_{\mathscr{H}} = \|\nabla_\lambda\|_{\mathscr{H}} = \|\lambda\|_{\ell^2} < \infty$.

(iii) Consider the *conditional expectation operator* $K \in \mathscr{L}(L^2_X, L^2_Z)$ of X given Z. Let in addition $p_{X,Z}$, p_X and p_Z be, respectively, the joint and marginal densities of (X, Z), X and Z w.r.t. a σ-finite measure. In this situation, the operator K is Hilbert Schmidt if and only if $\mathbb{E}\left[\frac{|p_{XZ}(X,Z)|^2}{|p_X(X)p_Z(Z)|^2}\right] < \infty$. □

2.2.2 Spectral theory and functional calculus

§2.2.29 **Definition**. Consider $T \in \mathscr{L}(\mathbb{H})$. The set $\rho(T) = \{\lambda \in \mathbb{K} : (\lambda \operatorname{Id}_{\mathbb{H}} - T)^{-1} \in \mathscr{L}(\mathbb{H})\}$ and its complement $\sigma(T) = \mathbb{K}\backslash\rho(T)$ is called *resolvent set* and *spectrum* of T, respectively. The subset $\sigma_p(T) = \{\lambda \in \mathbb{K} : \lambda \operatorname{Id}_{\mathbb{H}} - T \text{ is not injective}\}$ of $\sigma(T)$ is called *point spectrum* of T. An element λ of $\sigma_p(T)$ and $h \in \mathbb{H}\backslash\{0\}$ with $Th = \lambda h$ is called *eigenvalue* and *eigenfunction* (eigenvector), respectively. □

§2.2.30 **Properties**. *Consider $T \in \mathscr{K}(\mathbb{H})$.*

(i) *If T is self-adjoint, then $\sigma(T) \subset \mathbb{R}$.*

(ii) *If $\mathbb{H}$ is infinite dimensional, then $0 \in \sigma(T)$.*

(iii) *The (possibly empty) set $\sigma(T)\backslash\{0\}$ is at most countable.*

(iv) *Any $\lambda \in \sigma(T)\backslash\{0\}$ is an eigenvalue of T and its multiplicity is the (finite) dimension of the associated eigenspace $\mathcal{N}(\lambda \operatorname{Id}_{\mathbb{H}} - T)$.*

(v) *In $\sigma(T)$ the only possible accumulation point is zero.* □

§2.2.31 **Example**. The spectrum of a *multiplication operator* $\mathrm{M}_\lambda \in \mathscr{K}(\ell^2)$ and its associated *diagonal operator* $\nabla_\lambda = U^\star \mathrm{M}_\lambda U \in \mathscr{K}(\mathbb{H})$ is given by $\sigma(\mathrm{M}_\lambda) = \sigma(\nabla_\lambda) = \{\lambda_j, j \in \mathcal{J}\} \subset \mathbb{K}$. □

§2.2.32 **Definition**. Let $T \in \mathscr{K}(\mathbb{H})$ be normal ($\mathbb{K} = \mathbb{C}$) or self-adjoint ($\mathbb{K} = \mathbb{R}$). There exist

(i) a sequence $\lambda = (\lambda_j)_{j\in\mathcal{J}}$ in $\mathbb{K}\backslash\{0\}$ with $\|T\|_{\mathscr{L}} = \sup_{j\in\mathcal{J}} |\lambda_j|$ which has either a finite number of entries or zero as accumulation point, and determines a multiplication operator $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2(\mathcal{J}))$,

(ii) an ONS $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ with $\mathbb{U} := \overline{\operatorname{lin}}\{u_j, j \in \mathcal{J}\}$ and associated generalised Fourier series transform $\mathcal{U} \in \mathscr{L}(\mathbb{H}, \ell^2(\mathcal{J}))$ as defined in §2.2.4,

such that $\mathbb{H} = \mathcal{N}(T) \oplus \mathbb{U}$ and $T = \sum_{j\in\mathcal{J}} \lambda_j u_j \otimes u_j = \mathcal{U}^\star \mathrm{M}_\lambda \mathcal{U} = \nabla_\lambda$ (see §2.2.4 (ii), (iv) and (v)). For $j \in \mathcal{J}$, λ_j and u_j are, respectively, a non-zero *eigenvalue* and *associated eigenvector* of T respectively. $\{(\lambda_j, u_j), j \in \mathcal{J}\}$ is called an *eigensystem* of T. □

§2.2.33 **Properties**. *Let $T \in \mathscr{K}(\mathbb{H})$ be self-adjoint with eigensystem $\{(\lambda_j, u_j), j \in \mathcal{J}\}$, i.e., $\sigma(T)\backslash\{0\} = \{\lambda_j, j \in \mathcal{J}\} \subset \mathbb{R}$ denotes the (possibly empty) countable point spectrum of T. The sequence $\lambda = (\lambda_j)_{j\in\mathcal{J}}$ contains each eigenvalue of T repeated according to its multiplicity.*

(i) *If T is nuclear, then λ is absolute summable, i.e. $\|\lambda\|_{\ell_1} < \infty$, and $\operatorname{tr}(T) = \sum_{j\in\mathcal{J}} \lambda_j$.*

(ii) *If T is Hilbert-Schmidt, then λ is square summable and $\|T\|_{\mathscr{H}} = \|\lambda\|_{\ell^2} < \infty$.* □

§2.2.34 **Definition** *(Class of operators with given eigenfunctions)*. Given an ONS $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ let $\mathcal{E}_u(\mathbb{H})$ or $\mathcal{E}_u$ for short be the subset of $\mathscr{K}(\mathbb{H})$ containing all compact, normal (self-adjoint), linear operators having for some $\mathcal{J}' \subseteq \mathcal{J}$, $\{u_j, j \in \mathcal{J}'\}$ as eigenfunctions, i.e., for each $T \in \mathcal{E}_u(\mathbb{H})$ there exist $\mathcal{J}' \subseteq \mathcal{J}$ and a sequence $(\lambda_j)_{j\in\mathcal{J}'}$ in $\mathbb{K}\backslash\{0\}$ such that T admits $\{(\lambda_j, u_j), j \in \mathcal{J}'\}$ as eigensystem, i.e., $\mathcal{E}_u(\mathbb{H}) \subset \{\nabla_\lambda, \lambda \in \mathbb{K}^{\mathcal{J}}\}$. □

§2.2.35 **Example**. Let $C_g \in \mathscr{K}(L^2([0,1)))$ be a *circular convolution operator*. Consider as in §2.1.6 (ii) the *exponential basis* $\{e_j\}_{j\in\mathbb{Z}}$ in $L^2([0,1))$ and for $f \in L^2([0,1))$ the associated Fourier coefficients $[f]_j = \langle f, e_j\rangle_{L^2}$, $j \in \mathbb{Z}$. Keep in mind that C_g is normal and for all $f \in L^2([0,1))$ the convolution theorem states $[g \circledast f]_j = [g]_j[f]_j$ for all $j \in \mathbb{Z}$. Thereby, $\{([g]_j, e_j),\ j \in \mathbb{Z}\}$ is an eigensystem of the circular convolution operator C_g. In other words, for each $g \in L([0,1))$ we have $C_g \in \mathcal{E}_e(L^2([0,1)))$. □

§2.2.36 **Property**. *Let $T \in \mathscr{K}(\mathbb{H})$ be strictly positive definite and let $(\lambda_j)_{j\in\mathbb{N}}$ be a strictly positive, monotonically non-increasing sequence containing each eigenvalue of T repeated according to its multiplicity. For $m \in \mathbb{N}$ let $\mathcal{H}_m$ be the set of all m-dimensional subspaces $\mathbb{U}_m$ in $\mathbb{H}$, and denote by $\mathbb{U}_m^\perp$ the orthogonal complement of $\mathbb{U}_m$ in $\mathbb{H}$. Furthermore, let $\mathbb{B}_{\mathbb{U}_m} := \{h \in \mathbb{U}_m : \|h\|_{\mathbb{H}} = 1\}$ and $\mathbb{B}_{\mathbb{U}_m^\perp}$ be the unit ball in $\mathbb{U}_m$ and $\mathbb{U}_m^\perp$, respectively.*

(Courant's max-min-principle) $\lambda_m = \max_{\mathbb{U}_m \in \mathcal{H}_m} \min_{h\in\mathbb{B}_{\mathbb{U}_m}} \langle Th, h\rangle_{\mathbb{H}}$,

(Courant's min-max-principle) $\lambda_m = \min_{\mathbb{U}_{m-1} \in \mathcal{H}_{m-1}} \max_{h\in\mathbb{B}_{\mathbb{U}_{m-1}^\perp}} \langle Th, h\rangle_{\mathbb{H}}$. □

§2.2.37 **Definition**. Let $T \in \mathscr{K}(\mathbb{H}, \mathbb{G})$. There exist

(i) a sequence $\mathfrak{s} := (\mathfrak{s}_j)_{j\in\mathcal{J}}$ in $\mathbb{K}\backslash\{0\}$ with $\|T\|_{\mathscr{L}} = \sup_{j\in\mathcal{J}} |\mathfrak{s}_j|$ which has either a finite number of entries or zero as only accumulation point, and determines a multiplication operator $\mathrm{M}_\mathfrak{s} \in \mathscr{L}(\ell^2(\mathcal{J}))$,

(ii) an (possibly finite) ONS $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ with $\mathbb{U} := \overline{lin}\{u_j, j \in \mathcal{J}\}$ and associated generalised Fourier series transform $\mathcal{U} \in \mathscr{L}(\mathbb{H}, \ell^2(\mathcal{J}))$ (a partial isometry),

(iii) an (possibly finite) ONS $\{v_j, j \in \mathcal{J}\}$ in $\mathbb{G}$ with $\mathbb{V} := \overline{lin}\{v_j, j \in \mathcal{J}\}$ and associated generalised Fourier series transform $\mathcal{V} \in \mathscr{L}(\mathbb{G}, \ell^2(\mathcal{J}))$ (a partial isometry),

such that $\mathbb{H} = \mathcal{N}(T) \oplus \mathbb{U}$, $\mathbb{G} = \mathcal{N}(T^\star) \oplus \mathbb{V}$ and $T = \mathcal{V}^\star \mathrm{M}_\mathfrak{s} \mathcal{U} = \sum_{j\in\mathcal{J}} \mathfrak{s}_j\, u_j \otimes v_j$. In particular, $\{(|\mathfrak{s}_j|^2, u_j), j \in \mathcal{J}\}$ and $\{(|\mathfrak{s}_j|^2, v_j), j \in \mathcal{J}\}$ are an eigensystem of $T^\star T$ and $TT^\star$ respectively. The numbers $\{\mathfrak{s}_j, j \in \mathcal{J}\}$ and triplets $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathcal{J}\}$ are, respectively, called *singular values* and *singular system* of T. □

§2.2.38 **Properties**. *Let $T \in \mathscr{K}(\mathbb{H}, \mathbb{G})$ with singular system $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathcal{J}\}$ where the (possibly empty) countable point spectrum of $T^\star T$ (respectively, $TT^\star$) is given by $\sigma(T^\star T)\backslash\{0\} = \{|\mathfrak{s}_j|^2, j \in \mathcal{J}\} \subset \mathbb{R}$. The sequence $(|\mathfrak{s}_j|^2)_{j\in\mathcal{J}}$ contains each eigenvalue of $T^\star T$ repeated according to its multiplicity.*

(i) *If T is nuclear, then $\mathfrak{s}$ is absolute summable, i.e. $\|\mathfrak{s}\|_{\ell^1} < \infty$.*

(ii) *If T is Hilbert-Schmidt, then $\mathfrak{s}$ is square summable and $\|T\|_{\mathscr{H}} = \|\mathfrak{s}\|_{\ell^2} < \infty$.* □

§2.2.39 **Definition** *(Class of operators with known eigenfunctions)*. Given an ONS $\{u_j, j \in \mathcal{J}\}$ and $\{v_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and $\mathbb{G}$, respectively, let $\mathcal{S}_{uv}(\mathbb{H}, \mathbb{G})$ or $\mathcal{S}_{uv}$ for short, be the subset of $\mathscr{K}(\mathbb{H}, \mathbb{G})$ containing all compact, linear operators having for some $\mathcal{J}' \subseteq \mathcal{J}$, $\{u_j, j \in \mathcal{J}'\}$

and $\{u_j, j \in \mathcal{J}'\}$ as eigenfunctions, i.e., for each $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G})$ there exist $\mathcal{J}' \subseteq \mathcal{J}$ and a sequence $(\mathfrak{s}_j)_{j\in\mathcal{J}'}$ in $\mathbb{K}\backslash\{0\}$ such that T admits $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathcal{J}'\}$ as singular system. □

§2.2.40 **Property** *(Spectral theorem). If $T \in \mathscr{L}(\mathbb{H})$ is self-adjoint, then T is isometrically equivalent to a multiplication operator, i.e., there exist*

(i) *a measurable space (Ω, μ) (σ-finite, if $\mathbb{H}$ is separable),*

(ii) *a bounded (measurable) and μ-a.s. non zero function $\lambda : \Omega \to \mathbb{R}$ with associated multiplication operator $M_\lambda \in \mathscr{L}(L^2_\mu(\Omega))$, and*

(iii) *a partial isometry $\mathcal{U} \in \mathscr{L}(\mathbb{H}, L^2_\mu(\Omega))$,*

such that $T = \mathcal{U}^\star M_\lambda \mathcal{U}$. □

§2.2.41 **Example.** Let $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ be a real and even function. Consider the associated self-adjoint *convolution operator* $C_g \in \mathscr{L}(L^2(\mathbb{R}))$. Recall that the convolution theorem states $\mathcal{F}(g * f) = \mathcal{F}g \cdot \mathcal{F}f$ for all $f \in L^2(\mathbb{R})$ where $\mathcal{F}$ denotes the *Fourier-Plancherel transform*. Consequently, the operator C_g is unitarily equivalent to the multiplication operator $\mathrm{M}_\lambda \in \mathscr{L}(L^2(\mathbb{R}))$ with $\lambda = [\mathcal{F}g]$, that is $C_g = \mathcal{F}^{-1}\mathrm{M}_\lambda\mathcal{F}$. □

§2.2.42 **Property** *(Spectral theorem Halmos [1963]). Let $T : \mathbb{H} \supset \mathcal{D}(T) \to \mathbb{H}$ be a densely-defined self-adjoint operator. There exist*

(i) *a measurable space (Ω, μ) (σ-finite, if $\mathbb{H}$ is separable),*

(ii) *an unitary operator $\mathcal{U} \in \mathscr{L}(\mathbb{H}, L^2_\mu(\Omega))$,*

(iii) *a (measurable) function $\lambda : \Omega \to \mathbb{R}$ (μ-a.s. finite and non zero) and an associated multiplication operator $\mathrm{M}_\lambda : L^2_\mu(\Omega) \supset \mathcal{D}(\mathrm{M}_\lambda) \to L^2_\mu(\Omega)$ with $\mathcal{D}(\mathrm{M}_\lambda) = \big\{f \in L^2_\mu(\Omega) : \lambda f \in L^2_\mu(\Omega)\big\}$*

such that $\mathcal{D}(T) = \{h \in \mathbb{H} : \mathcal{U}h \in \mathcal{D}(M_\lambda)\}$ and

(a) *for all $f \in \mathcal{D}(\mathrm{M}_\lambda)$ we have $\mathrm{M}_\lambda f = \lambda \cdot f = \mathcal{U}T\mathcal{U}^\star f$,*

(b) *for all $h \in \mathcal{D}(T)$ it holds $Th = \mathcal{U}^\star \mathrm{M}_\lambda \mathcal{U}h$,*

i.e., T is unitarily equivalent to the multiplication operator M_λ. □

§2.2.43 **Example.** Let $T \in \mathscr{K}(\mathbb{H})$ be an injective and self-adjoint operator with eigenvalue decomposition $T = \mathcal{U}^\star\mathrm{M}_\lambda\mathcal{U}$ where $\mathcal{U} \in \mathscr{L}(\mathbb{H}, \ell^2)$ is unitary, $\mathrm{M}_\lambda \in \mathscr{L}(\ell^2)$ is a multiplication operator and λ a sequence in $\mathbb{R}\backslash\{0\}$ of eigenvalues repeated according to their multiplicities. If $\mathbb{H}$ is not finite dimensional then the range $\mathcal{R}(T)$ of T is dense in $\mathbb{H}$ but not closed. Therefore, there exists an inverse $T^{-1} : \mathcal{R}(T) \to \mathbb{H}$ of T which is densely-defined and self-adjoint but not continuous. In particular, we have $\mathcal{D}(T^{-1}) = \mathcal{R}(T) = \{h : \lambda^{-1}\mathcal{U}h \in \ell^2\}$ (which is called Picard's condition). Consider the multiplication operator $\mathrm{M}_{1/\lambda} : \ell^2 \supset \mathcal{D}(\mathrm{M}_{1/\lambda}) \to \ell$ with $\mathcal{D}(\mathrm{M}_{1/\lambda}) = \{x \in \ell : x/\lambda \in \ell^2\}$, then $\mathcal{D}(T^{-1}) = \{h \in \mathbb{H} : \mathcal{U}h \in \mathcal{D}(\mathrm{M}_{1/\lambda})\}$ and

(a) for all $x \in \mathcal{D}(\mathrm{M}_{1/\lambda})$ we have $\mathrm{M}_{1/\lambda}x = x/\lambda = \mathcal{U}T^{-1}\mathcal{U}^\star x$,

(b) for all $h \in \mathcal{D}(T^{-1})$ it holds $T^{-1}h = \mathcal{U}^\star\mathrm{M}_{1/\lambda}\mathcal{U}h$,

i.e. T^{-1} is unitarily equivalent to the multiplication operator $\mathrm{M}_{1/\lambda}$. We shall emphasise that $h \in \mathcal{D}(T^{-1}) = \mathcal{R}(T)$ if and only if $\|[h]/\lambda\|^2_{\ell^2} = \sum_{j\in\mathcal{J}} |[h]_j/\lambda_j|^2 < \infty$. On the other hand, for any $k \in \mathbb{N}$ we have $T^k = T \cdots T = \mathcal{U}^\star\mathrm{M}_{\lambda^k}\mathcal{U} = \sum_{j\in\mathcal{J}} \lambda_j^k\, u_j \otimes u_j$ which motivates for a function $g : \mathbb{R} \to \mathbb{R}$ to define the operator

$$g(T)h := \mathcal{U}^\star\mathrm{M}_{g(\lambda)}\mathcal{U}h = \sum_{j\in\mathcal{J}} g(\lambda_j)\, u_j \otimes u_j, \quad \text{for all } h \in \mathbb{H} \text{ with } \|g(\lambda)[h]\|_{\ell^2} < \infty.$$

If g is bounded then $g(T) \in \mathscr{L}(\mathbb{H})$ and $\|g(T)\|_{\mathscr{L}} = \sup\{|g(\lambda_j)|, j \in \mathcal{J}\} \leqslant \|g\|_{L^\infty}$. In particular, it allows to define T^s for all $s \in \mathbb{R}$. □

§2.2.44 **Definition** *(Functional calculus)*. Let $T \in \mathscr{L}(\mathbb{H})$ be self-adjoint and hence isometrically equivalent with multiplication by a bounded function λ in some $L^2_\mu(\Omega)$, that is, $T = U^\star \mathrm{M}_\lambda U$. Given a (measurable) function $g : \mathbb{R} \to \mathbb{R}$ define the multiplication operator

$$\mathrm{M}_{g(\lambda)} : L^2_\mu(\Omega) \supset \mathcal{D}(\mathrm{M}_{g(\lambda)}) \to L^2_\mu(\Omega)$$

with $\mathcal{D}(\mathrm{M}_{g(\lambda)}) = \big\{f \in L^2_\mu(\Omega) : g(\lambda)f \in L^2_\mu(\Omega)\big\}$ and an unitarily equivalent operator

$$g(T)h := \mathcal{U}^\star \mathrm{M}_{g(\lambda)}\mathcal{U}h, \quad \forall h \in \mathcal{D}(g(T)) := \{h \in \mathbb{H} : \mathcal{U}h \in \mathcal{D}(\mathrm{M}_{g(\lambda)})\}$$

where $g(T) : \mathscr{L}(\mathbb{H}) \supset \mathcal{D}(g(T)) \to \mathscr{L}(\mathbb{H})$. Moreover, if g is bounded then $g(T) \in \mathscr{L}(\mathbb{H})$ with $\|g(T)\|_{\mathscr{L}} = \sup\{|g(\lambda)|, \lambda \in \sigma(T)\} \leqslant \|g\|_{L^\infty}$. □

§2.2.45 **Property**. *Let $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$. Then $\mathcal{R}(T) = \mathcal{R}((T^\star T)^{1/2})$.*

§2.2.46 **Remark**. Considering an ONB $\{u_j, j \in \mathbb{N}\}$ in $\mathbb{H}$, the associated generalised Fourier series transform $\mathcal{U} \in \mathscr{L}(\mathbb{H}, \ell^2)$ and for a sequence $\mathfrak{v}$ the associated multiplication and diagonal operator $\mathrm{M}_\mathfrak{v} : \ell^2 \supset \mathcal{D}(\mathrm{M}_\mathfrak{v}) \to \ell^2$ and $\nabla_\mathfrak{v} = \mathcal{U}^\star \mathrm{M}_\mathfrak{v} \mathcal{U} : \mathbb{H} \supset \mathcal{D}(\nabla_\mathfrak{v}) \to \mathbb{H}$ defined as in §2.2.4 (iv) and (v), respectively. If $\mathfrak{v}$ is strictly positive then applying the functional calculus we observe that for any $s \in \mathbb{R}$ we have $\nabla^s_\mathfrak{v} = \mathcal{U}^\star \mathrm{M}_{\mathfrak{v}^s} \mathcal{U} = \nabla_{\mathfrak{v}^s}$. Moreover, recall that $\mathbb{H}_{\mathfrak{v}^s}$ denotes the completion of $\mathbb{H}$ w.r.t.. the weighted norm $\|\cdot\|_{\mathfrak{v}^s}$ given by $\|\cdot\|^2_{\mathfrak{v}^s} = \sum_{j\in\mathbb{N}} \mathfrak{v}_j^{2s}|\langle \cdot, u_j\rangle_\mathbb{H}|^2$ where obviously $\|h\|_{\mathfrak{v}^s} = \|\nabla_{\mathfrak{v}^s} h\|_\mathbb{H} = \|\nabla^s_\mathfrak{v} h\|_\mathbb{H}$ for all $h \in \mathcal{D}(\nabla_{\mathfrak{v}^s}) = \mathbb{H}_{\mathfrak{v}^s}$. Introduce further the Hilbert space $(\mathbb{H}_{\mathfrak{v}^s}, \langle\cdot,\cdot\rangle_{\mathfrak{v}^s})$ inner product $\langle\cdot,\cdot\rangle_{\mathfrak{v}^s} = \langle \nabla_{\mathfrak{v}^s}\cdot, \nabla_{\mathfrak{v}^s}\cdot\rangle_\mathbb{H}$. □

§2.2.47 **Definition**. Let $\nabla_\mathfrak{v} : \mathbb{H} \supset \mathcal{D}(\nabla_\mathfrak{v}) \to \mathbb{H}$ be diagonal for an unitary operator $\mathcal{U} \in \mathscr{L}(\mathbb{U}, \ell^2)$ and a monotonically increasing, unbounded sequence $\mathfrak{v}$ with $\mathfrak{v}_1 > 0$. For each $s \in \mathbb{R}$ consider the inner product $\langle\cdot,\cdot\rangle_{\mathfrak{v}^s} = \langle \nabla_{\mathfrak{v}^s}\cdot, \nabla_{\mathfrak{v}^s}\cdot\rangle_\mathbb{H}$ and the norm $\|\cdot\|_{\mathfrak{v}^s} = \|\nabla^s_\mathfrak{v}\cdot\|_\mathbb{H}$. The family $\{(\mathbb{U}_{\mathfrak{v}^s}, \langle\cdot,\cdot\rangle_{\mathfrak{v}^s}), s \in \mathbb{R}\}$ of Hilbert space is called a *Hilbert scale* (see Krein and Petunin [1966] for a rather complete theory). □

§2.2.48 **Properties**. *Let $\{(\mathbb{U}_{\mathfrak{v}^s}, \langle\cdot,\cdot\rangle_{\mathfrak{v}^s}), s \in \mathbb{R}\}$ be a Hilbert scale as introduced in Definition §2.2.47. Then the following assertions hold true:*

(i) *For any $-\infty < s < t < \infty <$ the space $\mathbb{U}_{\mathfrak{v}^t}$ is densely and continuously embedded in $\mathbb{U}_{\mathfrak{v}^s}$.*

(ii) *For $s, t \in \mathbb{R}$ holds $\nabla^{t-s}_\mathfrak{v} = \nabla^t_\mathfrak{v}\nabla^{-s}_\mathfrak{v}$, and in particular, $\nabla^{-1}_{\mathfrak{v}^s} = \nabla_{\mathfrak{v}^{-s}}$.*

(iii) *For $s \geqslant 0$ holds $\mathbb{U}_{\mathfrak{v}^s} = \mathcal{D}(\nabla_{\mathfrak{v}^s})$ and $\mathbb{U}_{\mathfrak{v}^{-s}}$ is the dual space of $\mathbb{U}_{\mathfrak{v}^s}$.*

(iv) *Considering $-\infty < r < s < t < \infty$ for any $h \in \mathbb{U}_{\mathfrak{v}^s}$ the interpolation inequality $\|h\|_{\mathfrak{v}^s} \leqslant \|h\|^{(t-s)/(t-r)}_{\mathfrak{v}^r}\|h\|^{(s-r)/(t-r)}_{\mathfrak{v}^t}$ holds true.* □

§2.2.49 **Example**. Let $T \in \mathscr{K}(\mathbb{H}, \mathbb{G})$ be injective with singular system $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathbb{N}\}$ for some ONB $\{u_j \in \mathbb{N}\}$ in $\mathbb{H}$ and strictly positive, monotonically non-increasing sequence $(\mathfrak{s}_j)_{j\in\mathbb{N}}$ containing each singular value of T repeated according to its multiplicity. Setting $\mathfrak{v} = \mathfrak{s}^{-2}$ the strictly positive definite operator $T^\star T$ admits the spectral representation $T^\star T = \mathcal{U}^\star \mathrm{M}_{\mathfrak{v}^{-1}}\mathcal{U} = \nabla_{\mathfrak{v}^{-1}}$. Obviously, $\mathfrak{v}$ is a monotonically increasing, unbounded sequence with $\mathfrak{v}_1 > 0$. Considering the associated Hilbert scale $\{(\mathbb{H}_{\mathfrak{v}^s}, \langle\cdot,\cdot\rangle_{\mathfrak{v}^s}), s \in \mathbb{R}\}$ it is then an immediate consequence that $\mathbb{H}_{\mathfrak{v}^t} = \mathcal{D}((T^\star T)^t)$ is dense in $\mathbb{H}_{\mathfrak{v}^s} = \mathcal{D}((T^\star T)^s)$ for $0 \leqslant s < t$. We say, a function f satisfies a *source condition*, if $f \in \mathcal{D}((T^\star T)^s)$ for some $s > 0$, i.e., $f = (T^\star T)^s h$ for some $h \in \mathbb{H}$. □

2.2.3 Abstract smoothing condition

§2.2.50 **Definition** *(Link condition)*. Denote by $\mathcal{T}(\mathbb{H})$ or $\mathcal{T}$ for short, the set of all strictly positive definite operator in $\mathscr{K}(\mathbb{H})$. Given an ONB $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and a strictly positive sequence $(\mathfrak{t}_j)_{j\in\mathcal{J}}$ consider the weighted norm $\|\cdot\|_{\mathfrak{t}}^2 = \sum_{j\in\mathcal{J}} \mathfrak{t}_j^2 |\langle\cdot, u_j\rangle_{\mathbb{H}}|^2$. For all $d \geqslant 1$ define the subset $\mathcal{T}_{u\mathfrak{t}}^d := \mathcal{T}_{u\mathfrak{t}}^d(\mathbb{H}) := \{T \in \mathcal{T} : d^{-1}\|h\|_{\mathfrak{t}} \leqslant \|Th\|_{\mathbb{H}} \leqslant d\|h\|_{\mathfrak{t}} \text{ for all } h \in \mathbb{H}\}$. We say, T satisfies the *link condition* $\mathcal{T}_{u\mathfrak{t}}^d$, if $T \in \mathcal{T}_{u\mathfrak{t}}^d$. Define further the subset $\mathcal{E}_{u\mathfrak{t}}^d = \{T \in \mathcal{E}_u : (T^\star T)^{1/2} \in \mathcal{T}_{u\mathfrak{t}}^d\}$ and $\mathcal{S}_{uv\mathfrak{t}}^d = \{T \in \mathcal{S}_{uv} : (T^\star T)^{1/2} \in \mathcal{T}_{u\mathfrak{t}}^d\}$ of $\mathcal{E}_u = \mathcal{E}_u(\mathbb{H})$ and $\mathcal{S}_{uv} = \mathcal{S}_{uv}(\mathbb{H},\mathbb{G})$ (see §2.2.34 and §2.2.39), respectively, containing any diagonal operator T in $\mathcal{E}_u$ and $\mathcal{S}_{uv}$ such that $(T^\star T)^{1/2}$ satisfies the link condition $\mathcal{T}_{u\mathfrak{t}}^d$. □

§2.2.51 **Remark**. We shall emphasise that for $T \in \mathscr{K}(\mathbb{H},\mathbb{G})$ the condition $(T^\star T)^{1/2} \in \mathcal{T}_{u\mathfrak{t}}^d$ is equivalent to $d^{-1}\|h\|_{\mathfrak{t}} \leqslant \|Th\|_{\mathbb{H}} \leqslant d\|h\|_{\mathfrak{t}}$ for all $h \in \mathbb{H}$. Observe further that $T \in \mathcal{S}_{uv}$ admitting a singular system $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathcal{J}'\}$ with $\mathcal{J}' \subseteq \mathcal{J}$ satisfies the link condition $\mathcal{S}_{uv\mathfrak{t}}^d$ if and only if $\mathcal{J}' = \mathcal{J}$ and $d^{-1} \leqslant |\mathfrak{s}_j|/\mathfrak{t}_j \leqslant d$ for all $j \in \mathcal{J}$. Thereby, we have that $T \in \mathcal{S}_{uv\mathfrak{t}}^d(\mathbb{H},\mathbb{G})$ if and only if $T^\star \in \mathcal{S}_{v,u,\mathfrak{t}}^d(\mathbb{G},\mathbb{H})$. We shall emphasise, that there are operators satisfying the link condition $\mathcal{T}_{u\mathfrak{t}}^d$ which do not belong to $\mathcal{E}_u$ (respectively, $\mathcal{S}_{uv}$), i.e., are not equal to ∇_λ for some sequence λ (not diagonal w.r.t. $\mathcal{U}$), that is admitting eigenfunctions which are not contained in the ONS $\{u_j, j \in \mathcal{J}\}$. Let us briefly give a construction of those. We consider a small perturbation of $\nabla_{\mathfrak{t}}$, that is, $T = \nabla_{\mathfrak{t}} + \nabla_{\mathfrak{t}} A \nabla_{\mathfrak{t}}$ where $A \in \mathscr{L}(\mathbb{H})$ is a non-negative definite operator with spectral norm $c := \|\nabla_{\mathfrak{t}} A\|_{\mathscr{L}}$ strictly smaller than one. Obviously, T is strictly positive definite, and $\|Th\|_{\mathbb{H}} \leqslant \|\mathrm{Id}_{\mathbb{H}} + \nabla_{\mathfrak{t}} A\|_{\mathscr{L}} \|\nabla_{\mathfrak{t}} h\|_{\mathbb{H}} \leqslant (1+c)\|h\|_{\mathfrak{t}}$. On the other hand, we have $\|(\mathrm{Id}_{\mathbb{H}} + \nabla_{\mathfrak{t}} A)^{-1}\|_{\mathscr{L}} = \frac{1}{1-\|\nabla_{\mathfrak{t}} A\|_{\mathscr{L}}} = \frac{1}{1-c}$ by the Neumann series argument §2.2.15, which in turn implies $\|h\|_{\mathfrak{t}} = \|\nabla_{\mathfrak{t}} h\|_{\mathbb{H}} = \|(\mathrm{Id}_{\mathbb{H}} + \nabla_{\mathfrak{t}} A)^{-1}\|_{\mathscr{L}} \|Th\|_{\mathbb{H}} \leqslant \frac{1}{1-c}\|Th\|_{\mathbb{H}}$. Combining both bounds the operator T satisfies the link condition $\mathcal{T}_{u\mathfrak{t}}^d$ for all $d \geqslant \max(1+c, \frac{1}{1-c})$ and is obviously not diagonal w.r.t. $\mathcal{U}$. □

§2.2.52 **Property**. *Let $T \in \mathcal{T}_{u\mathfrak{t}}^d$.*

(Inequality of Heinz [1951]) For all $|s| \leqslant 1$ holds $\frac{1}{d^{|s|}}\|h\|_{\mathfrak{t}^s} \leqslant \|T^s h\|_{\mathbb{H}} \leqslant d^{|s|}\|h\|_{\mathfrak{t}^s}$. □

§2.2.53 **Example** *(Example §2.2.49 continued)*. Consider the Hilbert scale $\{(\mathbb{H}_{\mathfrak{v}^s}, \langle\cdot,\cdot\rangle_{\mathfrak{v}^s}), s \in \mathbb{R}\}$ associated with the source condition, i.e., $\mathbb{H}_{\mathfrak{v}^s} = \mathcal{D}((T^\star T)^s)$ and $\|\cdot\|_{\mathfrak{v}^s} = \|(T^\star T)^{-s}\cdot\|_{\mathbb{H}}$ for $s > 0$. Suppose further that $(T^\star T)^{1/2} \in \mathcal{T}_{u\mathfrak{t}}^d$, i.e., T satisfies a link condition for some weighted norm $\|\cdot\|_{\mathfrak{t}}$ defined w.r.t. an ONB $\mathcal{U}$ in $\mathbb{H}$ and a strictly positive sequence $\mathfrak{t}$. Note that in general the two norms $\|\cdot\|_{\mathfrak{t}}$ and $\|\cdot\|_{\mathfrak{v}^s}$ are defined w.r.t. to different orthonormal basis in $\mathbb{H}$. However, rewriting the inequality of Heinz §2.2.52 accordingly it holds $\frac{1}{d^{|s|}}\|\cdot\|_{\mathfrak{t}^s} \leqslant \left\|(T^\star T)^{s/2}\cdot\right\|_{\mathbb{H}} \leqslant d^{|s|}\|\cdot\|_{\mathfrak{t}^s}$ or equivalently $\frac{1}{d^{|s|}}\|\cdot\|_{\mathfrak{t}^s} \leqslant \|\cdot\|_{\mathfrak{v}^{-s/2}} \leqslant d^{|s|}\|\cdot\|_{\mathfrak{t}^s}$. In other words the two norms $\|\cdot\|_{\mathfrak{t}^s}$ and $\|\cdot\|_{\mathfrak{v}^{-s/2}}$ are equivalent for any $|s| \leqslant 1$. Recall that $\mathfrak{v}^{-1/2} = \mathfrak{s}$ equals the sequence of singular values of T. We shall emphasise that the equivalence of $\|\cdot\|_{\mathfrak{t}^s}$ and $\|\cdot\|_{\mathfrak{v}^{-s/2}}$ under a link condition holds generally for all $|s| \leqslant 1$ only. However, if the ONB used to construct the norm $\|\cdot\|_{\mathfrak{t}^s}$ for the link condition coincides with the eigenfunctions of $T^\star T$ then the $\|\cdot\|_{\mathfrak{t}^s}$ and $\|\cdot\|_{\mathfrak{v}^{-s/2}}$ are equivalent for all $s \in \mathbb{R}$. □

§2.2.54 **Corollary**. *Let $T \in \mathcal{T}_{u\mathfrak{t}}^d$ and suppose that $f \in \mathbb{F}_{u\mathfrak{a}}^r$ (see Definition §2.1.18) where the two norms $\|\cdot\|_{\mathfrak{t}}$ and $\|\cdot\|_{1/\mathfrak{a}}$ are constructed w.r.t. the same ONB in $\mathbb{H}$. Assume in addition that there are constants $a, p > 0$ and a sequence $\mathfrak{v}$ such that $\mathfrak{t} = \mathfrak{v}^a$ and $\mathfrak{a} = \mathfrak{v}^p$. If $p \leqslant 2a$ then* *for any* $f \in \mathbb{E}_{u\mathfrak{a}}$ *holds* $f = (T^\star T)^{p/(2a)} h$ *with* $\|h\|_{\mathbb{H}} \leqslant d^{p/a} \|h\|_{1/\mathfrak{a}}$*, and conversely for any* $f = (T^\star T)^{p/(2a)} h$ *with* $\|h\|_{\mathbb{H}} < \infty$ *we have* $f \in \mathbb{E}_{u\mathfrak{a}}$ *with* $\|h\|_{1/\mathfrak{a}} \leqslant d^{p/a} \|h\|_{\mathbb{H}}$.

Proof of Corollary §2.2.54 is given in the lecture. □

§2.2.55 **Lemma**. *Given an ONB* $\{u_j, j \in \mathbb{N}\}$ *in* $\mathbb{H}$ *and a strictly positive non-increasing sequence* $(\mathfrak{t}_j)_{j\in\mathbb{N}}$ *consider the link condition* $\mathcal{T}_{u\mathfrak{t}}^{d}$*. Let* $T \in \mathcal{T}(\mathbb{H})$ *admit* $\{(\lambda_j, \psi_j), j \in \mathbb{N}\}$ *as eigensystem where the strictly positive, monotonically non-increasing sequence* $(\lambda_j)_{j\in\mathbb{N}}$ *contains each eigenvalue of* T *repeated according to its multiplicity and the associated eigenbasis* $\{\psi_j, j \in \mathbb{N}\}$ *does eventually not correspond to the ONB* $\{u_j, j \in \mathbb{N}\}$*. If* $T \in \mathcal{T}_{u\mathfrak{t}}^{d}$*, then we have* $d^{-1} \leqslant \lambda_j/\mathfrak{t}_j \leqslant d$ *for all* $j \in \mathbb{N}$.

Proof of Lemma §2.2.55 is given in the lecture. □

Chapter 3

Regularisation of ill-posed inverse problems

3.1 Ill-posed inverse problems

Let $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ be a linear bounded operator between separable Hilbert spaces $\mathbb{H}$ and $\mathbb{G}$.

§3.1.1 **Definition**. Given $g \in \mathbb{G}$ the reconstruction of a solution $f \in \mathbb{H}$ of the equation $g = Tf$ is called *inverse problem*. □

§3.1.2 **Definition** *(Hadamard [1932])*. An inverse problem $g = Tf$ is called *well-posed* if (i) a solution f *exists*, (ii) the solution f is *unique*, and (iii) the solution depends continuously on g. An inverse problem which is not well-posed is called *ill-posed*. □

For a broader overview on inverse problems we refer the reader to the monograph by Kress [1989] or Engl et al. [2000].

§3.1.3 **Property** *(Existence and identification)*.
There exists an unique solution of the equation $g = Tf$ if and only if

(existence) *g belongs to the range $\mathcal{R}(T)$ of T,*

(identification) *The operator T is injective, i.e., its null space $\mathcal{N}(T) = \{0\}$ is trivial.* □

§3.1.4 **Remark**. If there does not exist a solution typically one might consider a least-square solution which exists if and only if $g \in \mathcal{R}(T) \oplus \mathcal{N}(T^\star)$. A least-square solution with minimal norm, if it exists, could be recovered, in case the solution is not unique. Nevertheless, the main issue is often the stability of the inverse problem. More precisely, if the solution does not depend continuously on g, i.e., the inverse T^{-1} of T is not continuous, a reconstruction $f_n = T^{-1}\widehat{g}$ given a noisy version $\widehat{g}$ of g may be far from the solution f even if the noisy version $\widehat{g}$ is closed to g. □

§3.1.5 **Property**. *Denote by $\Pi_{\overline{\mathcal{R}}(T)}$ the orthogonal projection onto the closure $\overline{\mathcal{R}}(T)$ of the range of T. For each $g \in \mathbb{G}$ the following assertions are equivalent* (i) *f minimises $h \mapsto \|g - Th\|_{\mathbb{G}}$ (least square solution);* (ii) *$\Pi_{\overline{\mathcal{R}}(T)} g = Tf$;* (iii) *$T^\star g = T^\star T f$ (normal equation).* □

§3.1.6 **Remark**. We note that $g \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ implies $\Pi_{\overline{\mathcal{R}}(T)} g \in \mathcal{R}(T)$ and hence the preimage $T^{-1}(\Pi_{\overline{\mathcal{R}}(T)} g)$ is not empty. More precisely, due to the last assertion $T^{-1}(\Pi_{\overline{\mathcal{R}}(T)} g) = \{f \in \mathbb{H} : T^\star g = T^\star T f\}$ is the *set of least square solutions* associated to g. □

In the sequel keep in mind that for each $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ its restriction $T : \mathcal{N}(T)^\perp \to \mathcal{R}(T)$ is bijective and thus has an inverse $T^{-1} : \mathcal{R}(T) \to \mathcal{N}(T)^\perp$.

§3.1.7 **Definition**. For $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ the *Moore-Penrose inverse* (generalised or pseudo inverse) T^+ is the unique linear extension of $T^{-1} : \mathcal{R}(T) \to \mathcal{N}(T)^\perp$ to the domain $\mathcal{D}(T^+) := \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ with $\mathcal{N}(T^+) = \mathcal{R}(T)^\perp$ satisfying $T^+ g := T^{-1}\Pi_{\overline{\mathcal{R}}(T)} g$ for any $g \in \mathcal{D}(T^+)$. □

§3.1.8 **Remark**. We note that $TT^+T = T$, $T^+TT^+ = T^+$, $T^+T = \Pi_{\mathcal{N}(T)^\perp}$ and $TT^+g = \Pi_{\overline{\mathcal{R}(T)}}g$ for any $g \in \mathcal{D}(T^+)$. If T is injective, and hence $T^\star T$, then $T^\star T : \mathbb{H} \to \mathcal{R}(T^\star T)$ is invertible, which in turn, for any $g \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$, implies that $(T^\star T)^{-1}T^\star g$ is the unique solution of the normal equation, and thus $T^{-1}(\Pi_{\overline{\mathcal{R}(T)}}g) = \{T^+g\} = \{(T^\star T)^{-1}T^\star g\}$. More generally, we have $T^+ = (T^\star T)^+T^\star$ and, if T is invertible then $T^+ = T^{-1}$. □

§3.1.9 **Property**. *For each $g \in \mathcal{D}(T^+)$, T^+g belongs to $T^{-1}(\Pi_{\overline{\mathcal{R}(T)}}g)$ and, hence is a least square solution. Moreover, T^+g is the unique least square solution with minimal $\|\cdot\|_{\mathbb{H}}$-norm, that is, $\|T^+g\|_{\mathbb{H}} = \inf\{\|h\|_{\mathbb{H}} : h \in T^{-1}(\Pi_{\overline{\mathcal{R}(T)}}g)\}$.* □

§3.1.10 **Property**. *If $\mathbb{H}$ and $\mathbb{G}$ are infinite dimensional and $T \in \mathscr{K}(\mathbb{H}, \mathbb{G})$ is injective, then $\inf\{\|Th\|_{\mathbb{G}} : \|h\|_{\mathbb{H}} = 1, h \in \mathbb{H}\} = 0$, which in turn implies that $T^{-1} : \mathcal{R}(T) \to \mathbb{H}$ and, hence T^+ is not continuous.* □

3.2 Spectral regularisation

In the sequel, given an infinite dimensional $\mathbb{H}$ let $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ and let T^+ its Moore-Penrose inverse as in Definition §3.1.7.

§3.2.1 **Definition**. A family $\{R_\alpha \in \mathscr{L}(\mathbb{G}, \mathbb{H}), \alpha \in (0, 1)\}$ of operators is called *regularisation* of T^+ if for any $g \in \mathcal{D}(T^+)$ holds $\|R_\alpha g - T^+g\|_{\mathbb{H}} \to 0$ as $\alpha \to 0$. □

§3.2.2 **Remark**. Note that, if T^+ is not bounded, then $\|R_\alpha\|_{\mathscr{L}} \to \infty$ as $\alpha \to 0$. On the other hand side, if $(g_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{G}$ such that $\|g_n - g\|_{\mathbb{G}} \leqslant n^{-1}$ for all $n \in \mathbb{N}$, then there exists a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $(0, 1)$ such that $\|R_{\alpha_n}g_n - T^+g\|_{\mathbb{H}} \to 0$ as $n \to \infty$. □

§3.2.3 **Definition**. The family $\{(T^\star T + \alpha\,\mathrm{Id}_{\mathbb{H}})^{-1}T^\star \in \mathscr{L}(\mathbb{G}, \mathbb{H}), \alpha \in (0, 1)\}$ is called *Tikhonov regularisation* of T^+. □

§3.2.4 **Remark**. Given $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ consider for each $\alpha \in (0, 1)$ the strictly positive definite operator $T_\alpha := T^\star T + \alpha\,\mathrm{Id}_{\mathbb{H}} \in \mathscr{L}(\mathbb{H})$ where $\|T_\alpha h\|_{\mathbb{H}}\,\|h\|_{\mathbb{H}} \geqslant \langle T_\alpha h, h\rangle_{\mathbb{H}} \geqslant \alpha\,\|h\|^2_{\mathbb{H}} > 0$ for any $h \in \mathbb{H}\backslash\{0\}$ by applying the Cauchy-Schwarz inequality §2.1.3 and, hence

$$\inf\{\|T_\alpha h\|_{\mathbb{H}} : \|h\|_{\mathbb{H}} = 1, h \in \mathbb{H}\} \geqslant \alpha > 0. \tag{3.1}$$

Consequently, T_α is injective and moreover, its range is closed. Indeed, if a sequence $(T_\alpha h_n)_{n\in\mathbb{N}}$ converges to $g \in \mathbb{G}$, then $(h_n)_{n\in\mathbb{N}}$ is a Cauchy sequence due to (3.1), and thus converges, say, to $h \in \mathbb{H}$. Since T_α is continuous, it follows $T_\alpha h_n \to T_\alpha h$ and $g = T_\alpha h$. Exploiting that T_α is injective with closed range follows $\mathcal{R}(T_\alpha) = \mathcal{N}(T_\alpha)^\perp = \{0\}^\perp = \mathbb{H}$ which in turn implies T_α is invertible, and due to the open mapping theorem $T_\alpha^{-1} \in \mathscr{L}(\mathbb{H})$ where $\|T_\alpha^{-1}\|_{\mathscr{L}} \leqslant \alpha^{-1}$ employing (3.1) together $\|T_\alpha^{-1}\|_{\mathscr{L}} = \sup\{\|h\|_{\mathbb{H}} / \|T_\alpha h\|_{\mathbb{H}} : h \in \mathbb{H}\backslash\{0\}\}$ since $\mathcal{R}(T_\alpha) = \mathbb{H}$. Consequently, the family $\{(T^\star T + \alpha\,\mathrm{Id}_{\mathbb{H}})^{-1}T^\star \in \mathscr{L}(\mathbb{G}, \mathbb{H}), \alpha \in (0, 1)\}$ is well-defined. □

§3.2.5 **Lemma**. *For each $h \in \mathcal{N}(T)^\perp$ holds $\|\alpha(T^\star T + \alpha\,\mathrm{Id}_{\mathbb{H}})^{-1}h\|_{\mathbb{H}} \to 0$ as $\alpha \to 0$.* □

Proof of Lemma §3.2.5 is given in the lecture. □

§3.2.6 **Remark.** Let $g \in \mathcal{D}(T^+)$. Setting $h = T^+ g$ and $f_\alpha = (T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})^{-1} T^\star g$ we have

$$\begin{aligned}(T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})(h - f_\alpha) &= T^\star T T^+ g + \alpha h - (T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})(T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})^{-1} T^\star g \\ &= T^\star g + \alpha h - T^\star g = \alpha h.\end{aligned}$$

Rewriting the last identity we obtain $(T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})^{-1} T^\star g - T^+ g = -\alpha (T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})^{-1} h$. Consequently, from Lemma §3.2.5 follows $\|(T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})^{-1} T^\star g - T^+ g\|_{\mathbb{H}} \to 0$ as $\alpha \to 0$ since $h = T^+ g \in \mathbb{H}$. Thereby, the Tikhonov family as in §3.2.3 is indeed a regularisation in the sense of Definition §3.2.1. □

§3.2.7 **Lemma.** *For each $C \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ the following statements are equivalent:*

(i) *f minimises the generalised Tikhonov functional $h \mapsto F_\alpha(h) := \frac{1}{2}\|g - Th\|^2_{\mathbb{G}} + \frac{\alpha}{2}\|Ch\|^2_{\mathbb{G}}$*

(ii) *f is solution of the normal equation: $T^\star g = (T^\star T + \alpha C^\star C) f$.*

Proof of Lemma §3.2.7 is given in the lecture. □

§3.2.8 **Remark.** Observe that $\mathcal{N}(T) \cap \mathcal{N}(C) = \mathcal{N}(T^\star T + \alpha C^\star C)$ which in turn implies, that the solution of the generalised Tikhonov functional, if it exists, is unique if and only if $\mathcal{N}(T) \cap \mathcal{N}(C) = \{0\}$. Keep in mind, that the existence of a solution is ensured, for example, if $(T^\star T + \alpha C^\star C)$ has a continuous inverse. □

§3.2.9 **Corollary.** *Given the Tikhonov regularisation $\{(T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})^{-1} T^\star\}$ as in §3.2.3 for each $g \in \mathbb{G}$, $f_\alpha := (T^\star T + \alpha \operatorname{Id}_{\mathbb{H}})^{-1} T^\star g$ is the unique minimiser in $\mathbb{H}$ of the Tikhonov functional $h \mapsto \frac{1}{2}\|g - Th\|^2_{\mathbb{G}} + \frac{\alpha}{2}\|h\|^2_{\mathbb{H}}$.* □

§3.2.10 **Definition.** Given an operator $C \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ satisfying

(i) $\mathcal{R}(C)$ is closed and

(ii) there exists $c > 0$ such that for any $h \in \mathcal{N}(C)$ it holds $\|Th\|_{\mathbb{G}} \geqslant h\,\|h\|_{\mathbb{H}}$,

the family $\{(T^\star T + \alpha C^\star C)^{-1} T^\star \in \mathscr{L}(\mathbb{G}, \mathbb{H}), \alpha \in (0,1)\}$ is called *generalised Tikhonov regularisation* of T^+. □

§3.2.11 **Remark.** Assumption (i) and (ii) ensure together that the generalised Tikhonov regularisation is well-defined. More precisely, introduce inner products $\langle h, h' \rangle_* := \langle Th, Th' \rangle_{\mathbb{G}} + \langle Ch, Ch' \rangle_{\mathbb{G}}$ and $\langle h, h' \rangle_C := \langle h, h' \rangle_{\mathbb{H}} + \langle Ch, Ch' \rangle_{\mathbb{G}}$ on $\mathbb{H}$ with associated norms $\|\cdot\|_*$ and $\|\cdot\|_C$. Since $\mathbb{H}$ is complete w.r.t. both norms (due to (i) and (ii)), it follows from §2.1.2 that $\|\cdot\|_*$ and $\|\cdot\|_C$ are equivalent (keeping in mind that $\|h\|^2_* \leqslant \max(\|T\|^2_{\mathscr{L}}, 1)\,\|h\|^2_C$). Consequently, there is $K > 0$ such that $\|h\|_* \geqslant K\,\|h\|_C$ and thus $\|Th\|^2_{\mathbb{G}} + \|Ch\|^2_{\mathbb{G}} \geqslant K^2(\|h\|^2_{\mathbb{H}} + \|Ch\|^2_{\mathbb{G}})$. Exploiting the last inequality we obtain $\|T^\star Th + \alpha C^\star Ch\|_{\mathbb{H}} \geqslant K^2 \min(1, \alpha)\,\|h\|_{\mathbb{H}}$ for any $h \in \mathbb{H}$. In analogy to the arguments exploiting (3.1) in Remark §3.2.4, $T^\star T + \alpha C^\star C$ is injective with closed range and, thus it has a continuous inverse, i.e., $(T^\star T + \alpha C^\star C)^{-1} \in \mathscr{L}(\mathbb{H})$. Consequently, the generalised Tikhonov regularisation $\{R_\alpha := (T^\star T + \alpha C^\star C)^{-1} T^\star \in \mathscr{L}(\mathbb{G}, \mathbb{H}), \alpha \in (0,1)\}$ is well-defined. Moreover, keeping in mind Lemma §3.2.7 $f_\alpha := R_\alpha g \in \mathbb{H}$ is obviously a solution of the normal equation, and thus the unique minimiser of the generalised Tikhonov functional. □

§3.2.12 **Corollary.** *Consider the generalised Tikhonov regularisation as in §3.2.10. For each $g \in \mathbb{G}$, $f_\alpha := (T^\star T + \alpha C^\star C)^{-1} T^\star g$ is the unique minimiser in $\mathbb{H}$ of the generalised Tikhonov functional $h \mapsto \frac{1}{2}\|g - Th\|^2_{\mathbb{G}} + \frac{\alpha}{2}\|Ch\|^2_{\mathbb{G}}$.* □

§3.2.13 **Remark.** Introduce further the adjoint $T_*^\star$ and $C_*^\star$ of T and C, respectively, w.r.t. the inner product $\langle\cdot,\cdot\rangle_*$, i.e., $\langle Th, g\rangle_{\mathbb{G}} = \langle h, T_*^\star g\rangle_*$ and $\langle Ch, g\rangle_{\mathbb{G}} = \langle h, C_*^\star g\rangle_*$ for all $h \in \mathbb{H}$ and $g \in \mathbb{G}$. In particular, for each $g \in \mathbb{G}$ and $h \in \mathbb{H}$ we have

(a) $T_*^\star g = (T^\star T + C^\star C)^{-1} T^\star g$,

(b) $C_*^\star g = (T^\star T + C^\star C)^{-1} C^\star g$ and

(c) $(T_*^\star T + C_*^\star C)h = h$ (i.e., $T_*^\star T + C_*^\star C = \mathrm{Id}_{\mathbb{H}}$).

We note that $\mathcal{N}(T_\star^\star) = \mathcal{N}(T^\star)$ and $\overline{\mathcal{R}}(T_\star^\star) = \mathcal{N}(T)^{\perp_*}$ where $\mathcal{N}(T)^{\perp_*}$ denotes the orthogonal complement of $\mathcal{N}(T)$ in $(\mathbb{H}, \langle\cdot,\cdot\rangle_*)$. □

Consider the restriction of T as bijective map from $\mathcal{N}(T)^{\perp_*}$ to $\mathcal{R}(T)$ and denote its inverse by $T_*^{-1} : \mathcal{R}(T) \to \mathcal{N}(T)^{\perp_*}$. Given the orthogonal projection $\Pi_{\overline{\mathcal{R}}(T)}$ onto $\overline{\mathcal{R}}(T)$ its associated Moore-Penrose inverse T_*^+ (see §3.1.7) defined on $\mathcal{D}(T_*^+) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp = \mathcal{D}(T^+)$ is given by $T_*^+ := T_*^{-1}\Pi_{\overline{\mathcal{R}}(T)}$.

§3.2.14 **Proposition.** *Consider the generalised Tikhonov regularisation $\{(T^\star T + \alpha C^\star C)^{-1} T^\star\}$ as in §3.2.10. Under the conditions (i) and (ii) of Definition §3.2.10 for $g \in \mathbb{G}$ and $f_\alpha = (T^\star T + \alpha C^\star C)^{-1} T^\star g$ the following statements are equivalent:*

(I) *$g \in \mathcal{D}(T_*^+) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp = \mathcal{D}(T^+)$;*

(II) *there is $f_o \in \mathbb{H}$ such that $\|f_\alpha - f_o\|_* \to 0$ as $\alpha \to 0$.*

Moreover, under the equivalent conditions holds $f_o = T_*^+ g$.

Proof of Proposition §3.2.14 is given in the lecture. □

§3.2.15 **Remark.** Due to the last proposition the generalised Tikhonov family as in §3.2.10 is indeed a regularisation in the sense of Definition §3.2.1. Moreover, we shall emphasise that $\|f_\alpha - f_o\|_* \to 0$ if and only if $\|Tf_\alpha - Tf_o\|_{\mathbb{G}} \to 0$ and $\|Cf_\alpha - Cf_o\|_{\mathbb{G}} \to 0$, which in turn implies $\|f_\alpha - f_o\|_{\mathbb{H}} \to 0$. Keep further in mind that $T_*^\star g = T_*^\star T f$ holds if and only if $T^\star g = T^\star T f$ is true, since $T^\star T + C^\star C$ is continuously invertible. Thereby, for each $g \in \mathcal{D}(T^+)$ the set of least squares solution $T^{-1}(\Pi_{\overline{\mathcal{R}}(T)} g)$ satisfies $T^{-1}(\Pi_{\overline{\mathcal{R}}(T)} g) = \{f \in \mathbb{H} : T^\star T f = T^\star g\} = \{f \in \mathbb{H} : T_*^\star T f = T_*^\star g\} = \{f_o\} + \mathcal{N}(T)$ with $f_o = T_*^+ g$. Each $f \in T^{-1}(\Pi_{\overline{\mathcal{R}}(T)} g)$ can thus be written as $f = f_o + u$ for some $u \in \mathcal{N}(T)$ with $f_o \in (\mathcal{N}(T))^{\perp_*}$, and hence, $Tf = Tf_*$ and $\|f_o\|_*^2 \leqslant \|f_o\|_*^2 + \|u\|_*^2 = \|f\|_*^2$, which together implies $\|Cf_o\|_{\mathbb{G}}^2 + \|Cf\|_{\mathbb{G}}^2$ for any $f \in T^{-1}(\Pi_{\overline{\mathcal{R}}(T)} g)$. In other words, f_o is the unique least squares solution with minimal $\|C\bullet\|_{\mathbb{G}}$-norm. □

§3.2.16 **Definition.** Given a family $\{r_\alpha, \alpha \in (0,1)\}$ of real-valued (piecewise) continuous functions defined on $[0, \|T\|_{\mathscr{L}}^2]$ the family $\{r_\alpha(T^\star T)T^\star \in \mathscr{L}(\mathbb{G}, \mathbb{H}), \alpha \in (0,1)\}$ of operators is called *spectral regularisation* of T^+ if

(i) for all $\lambda \in (0, \|T\|_{\mathscr{L}}^2]$ holds $|1 - \lambda r_\alpha(\lambda)| \to 0$ as $\alpha \to 0$, and

(ii) there is $K > 0$ such that $|\lambda r_\alpha(\lambda)| \leqslant K$ for all $\lambda \in [0, \|T\|_{\mathscr{L}}^2]$ and $\alpha \in (0,1)$. □

§3.2.17 **Remark.** Given $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ consider a spectral regularisation $\{R_\alpha = r_\alpha(T^\star T)T^\star\}$ as in Definition §3.2.16. The operator $T^\star T \in \mathscr{L}(\mathbb{H})$ is isometrically equivalent with multiplication in some $L_\mu^2(\Omega)$ by a strictly positive function λ bounded by $\|T\|_{\mathscr{L}}^2$. Applying the functional calculus we have $\|r_\alpha(T^\star T)T^\star\|_{\mathscr{L}} \leqslant \sup\{|r_\alpha(\lambda)\sqrt{\lambda}|, \lambda \in [0, \|T\|_{\mathscr{L}}^2]\} < \infty$ since r_α is piecewise continuous on the compact interval $[0, \|T\|_{\mathscr{L}}^2]$. Consequently, $R_\alpha \in \mathscr{L}(\mathbb{G}, \mathbb{H})$

for all $\alpha \in (0,1)$, i.e., the family is well-defined. Moreover, $\|(r_\alpha(T^\star T)T^\star T - \mathrm{Id}_{\mathbb{H}})h\|_{\mathbb{H}}^2 = \|U^\star \mathrm{M}_{r_\alpha(\lambda)\lambda-1} Uh\|_{\mathbb{H}}^2 = \|(1-\lambda r_\alpha(\lambda))Uh\|_{L^2_\mu}^2 = \mu(|1-\lambda r_\alpha(\lambda)|^2|Uh|^2)$ holds for $h \in \mathcal{N}(T)^\perp$. From §3.2.16 (ii) follows $|1-\lambda r_\alpha(\lambda)| \leqslant 1+K$ for all $\alpha \in (0,1)$. Since $Uh \in L^2_\mu$ employing the dominated convergence theorem from §3.2.16 (i) follows $\|(r_\alpha(T^\star T)T^\star T - \mathrm{Id}_{\mathbb{H}})h\|_{\mathbb{H}}^2 \to 0$ as $\alpha \to 0$ for all $h \in \mathcal{N}(T)^\perp$. Since for all $g \in \mathcal{D}(T^+)$ with $h := T^+g \in \mathcal{N}(T)^\perp$ holds $R_\alpha g - h = (r_\alpha(T^\star T)T^\star T - \mathrm{Id}_{\mathbb{H}})h$ we have $\|R_\alpha g - T^+ g\|_{\mathbb{H}} \to 0$ as $\alpha \to 0$, and a continuous spectral regularisation as in §3.2.16 is indeed a regularisation in the sense of Definition §3.2.1. We shall emphasise that for any $g \notin \mathcal{D}(T^+)$ it can be shown that $\|r_\alpha(T^\star T)T^\star g\|_{\mathbb{H}} \to \infty$ as $\alpha \to 0$. □

§3.2.18 **Proposition**. *Let $\{r_\alpha(T^\star T)T^\star \in \mathscr{L}(\mathbb{G},\mathbb{H}), \alpha \in (0,1)\}$ be a continuous spectral regularisation of T^+ defined in Definition §3.2.16. Assume that in addition to §3.2.16 (i) and (ii) for any $s \in [0, s_o]$ for some $s_o > 0$ there is a constant $c_s < \infty$ such that*

(iii) *for all $\lambda \in [0, \|T\|_{\mathscr{L}}^2]$ and $\alpha \in (0,1)$ holds $\lambda^s|1-\lambda r_\alpha(\lambda)| \leqslant c_s \alpha^s$.*

Consider $f_\alpha := r_\alpha(T^\star T)T^\star g$ and let $g \in \mathcal{D}(T^+)$ and, thus $f := T^+ g \in \mathbb{H}$.

(a) *If there are $s \in [0, s_o]$ and $h \in \mathbb{H}$ such that $f \in \mathcal{R}\big((T^\star T)^s\big)$ (source condition as in Example §2.2.49), then for all $\alpha \in (0,1)$ holds $\|f_\alpha - f\|_{\mathbb{H}} \leqslant c_s \alpha^s\, \|h\|_{\mathbb{H}}$.*

(b) *If $T \in \mathcal{T}^d_{\mathfrak{ut}}$ (link condition as in Definition §2.1.18) and $f \in \mathbb{F}_{\mathfrak{ua}}$ (abstract smoothness condition as in Definition §2.1.18) where $\mathfrak{t} = \mathfrak{v}^a$ and $\mathfrak{a} = \mathfrak{v}^p$ for some constants $0 < p \leqslant 2a$ and a sequence $\mathfrak{v}$, then $\|f_\alpha - f\|_{\mathbb{H}} \leqslant c_s\, d^{p/a}\, r\, \alpha^{p/(2a)}$ for all $\alpha \in (0,1)$.*

Proof of Proposition §3.2.18 is given in the lecture. □

§3.2.19 **Examples**. Let us discuss certain special continuous regularisation methods satisfying in addition §3.2.18 (iii).

(i) Tikhonov regularisation as defined in §3.2.3 is given by $r_\alpha(\lambda) = (\lambda+\alpha)^{-1}$ and satisfies §3.2.18 (iii) with $s_o = 1$ and $c_s = s^s(1-s)^{1-s}$.

(ii) Spectral cut-off given by the piecewise continuous function $r_\alpha(\lambda) = \frac{1}{\lambda}\mathbb{1}_{\{\lambda \geqslant \alpha\}}$ is a continuous regularisation methods satisfying §3.2.16 (i) and (ii) with $K = 1$. Moreover, §3.2.18 (iii) holds with $s_o = \infty$ and $c_s = 1$.

(iii) A special iterative regularisation method is the Landweber iteration. This method is based on a transformation of the normal equation into an equivalent fixed point equation $f = f + \omega T^\star(g - Tf)$ with $0 < \omega \leqslant \|T\|_{\mathscr{L}}^{-2}$. Then the corresponding fixed point operator $\mathrm{Id}_{\mathbb{H}} - \omega T^\star T$ is nonexpansive and f may be approximated by f_k determined by $f_{n+1} = f_n + \omega T^\star(g - Tf_n)$, $n = [\![0, k-1]\!]$, $f_0 = 0$. Note, that without loss of generality, we can assume $\|T\|_{\mathscr{L}} \leqslant 1$ and drop the parameter ω. By induction the iterate f_k can be expressed non-recursively through $f_k = \sum_{n=0}^{k-1}(\mathrm{Id}_{\mathbb{H}} - T^\star T)^j T^\star g$ and thus $r_\alpha(\lambda) = \sum_{n=0}^{k-1}(1-\lambda)^j$ where $1 - \lambda r_\alpha(\lambda) = (1-\lambda)^k$. Under the assumption $\|T\|_{\mathscr{L}} \leqslant 1$, the Landweber iteration is thus a continuous regularisation methods with $\alpha = 1/k$ satisfying §3.2.16 (i) and (ii) with $K = 1$. Moreover, §3.2.18 (iii) holds with $s_o = \infty$ and $c_s = s^s e^{-s}$. □

3.3 Regularisation by dimension reduction

Here and subsequently, we consider a class of functions $\mathbb{F}^r_{\mathfrak{ua}} \subset \mathbb{U}_{1/\mathfrak{a}}$ as given in §2.1.18 w.r.t. an ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and a strictly positive sequence $\mathfrak{a} = (\mathfrak{a}_j)_{j\in\mathcal{J}}$. We shall frequently exploit that $\{(\mathbb{U}_{\mathfrak{a}^s}, \langle\cdot,\cdot\rangle_{\mathfrak{a}^s}), s\in\mathbb{R}\}$ eventually forms a Hilbert scale w.r.t. $\nabla_{\mathfrak{a}}$ which is diagonal w.r.t.. the generalised Fourier transform U associated to $\mathcal{U}$. Moreover, we assume a nested sieve $(\mathcal{J}_m)_{m\in\mathcal{M}}$ in $\mathcal{J}$ and its associated nested sieve $(\mathbb{U}_m)_{m\in\mathcal{M}}$ in $\mathbb{U}$ (see §2.1.12). For $f\in\mathbb{U}$ we introduce a theoretical approximation $f_m\in\mathbb{U}_m$. On the one hand we consider the orthogonal projection $f_m=\Pi_{\mathbb{U}_m}f=\sum_{j\in\mathcal{J}}([f]_j\mathbb{1}_{\mathcal{J}_m}(j))u_j=U^\star([f]\mathbb{1}_{\mathcal{J}_m})$ of f onto $\mathbb{U}_m$ by using the sequence of indicators $\mathbb{1}_{\mathcal{J}_m}:=(\mathbb{1}_{\mathcal{J}_m}(j))_{j\in\mathcal{J}}$. On the other hand the construction of f_m is motivated by a linear Galerkin approach introduced below. We shall measure the accuracy of the approximation f_m of f by its distance $\mathfrak{d}_{\text{ist}}(f_m,f)$ where $\mathfrak{d}_{\text{ist}}(\cdot,\cdot)$ is a certain semi metric. Note that in general $\mathfrak{d}_{\text{ist}}(f_m,f)$ is not monotone in $m\in\mathcal{M}$ and hence we define $\operatorname{bias}_m(f):=\sup\{\mathfrak{d}_{\text{ist}}(f,f_k),k\in\mathcal{M}\cap[\![m,\infty[\![\}$ as the approximation error. We are particularly interested in the following two cases.

§3.3.1 **Definition**. Let $f_m\in\mathbb{U}_m$ be a theoretical approximation of $f\in\mathbb{U}_{1/\mathfrak{a}}$, and hence $\Pi_{\mathbb{U}^\perp}f=0$. Keep in mind that $\mathbb{U}^\perp$ and $\mathbb{U}_m^\perp$ denotes the orthogonal complement of $\mathbb{U}$ and $\mathbb{U}_m$ in $\mathbb{H}$ and $\mathbb{U}$, respectively.

(global) Given the ONS $\mathcal{U}$ and a strictly positive sequence $\mathfrak{v}$ consider the completion $\mathbb{U}_{\mathfrak{v}}$ of $\mathbb{U}$ w.r.t. the weighted norm $\|\cdot\|_{\mathfrak{v}}$. If $\mathbb{U}_{1/\mathfrak{a}}\subset\mathbb{U}_{\mathfrak{v}}$, then $\mathfrak{d}^{\mathfrak{v}}_{\text{ist}}(h_1,h_2):=\|h_1-h_2\|_{\mathfrak{v}}$, $h_1,h_2\in\mathbb{U}_{\mathfrak{v}}$ defines a *global distance* on $\mathbb{U}_{\mathfrak{v}}$. For $f\in\mathbb{F}^r_{u\mathfrak{a}}$ and $m\in\mathcal{M}$ we denote by $\operatorname{bias}^{\mathfrak{v}}_m(f):=\|\Pi_{\mathbb{U}_m}f-f\|_{\mathfrak{v}}=\left\|\Pi_{\mathbb{U}_m^\perp}f\right\|_{\mathfrak{v}}=\sup\{\mathfrak{d}^{\mathfrak{v}}_{\text{ist}}(f,f_k),k\in\mathcal{M}\cap[\![m,\infty[\![\}$ the *global approximation error*.

(local) Let Φ be a linear functional and $\mathbb{U}_{1/\mathfrak{a}}\subset\mathcal{D}(\Phi)$, then $\mathfrak{d}^{\Phi}_{\text{ist}}(h_1,h_1):=|\Phi(h_1-h_2)|$, $h_1,h_2\in\mathcal{D}(\Phi)$, defines a *local distance*. For $f\in\mathbb{F}^r_{u\mathfrak{a}}$ and $m\in\mathcal{M}$ we denote by $\operatorname{bias}^{\Phi}_m(f):=\sup\{|\Phi(\Pi_{\mathbb{U}_k^\perp}f)|,k\in\mathcal{M}\cap[\![m,\infty[\![\}=\sup\{\mathfrak{d}^{\Phi}_{\text{ist}}(f,f_k),k\in\mathcal{M}\cap[\![m,\infty[\![\}$ the *local approximation error*. □

§3.3.2 **Remark**. We shall emphasise, if $\|\mathfrak{a}\mathfrak{v}\|_{\ell^\infty}=\sup\{\mathfrak{a}_j\mathfrak{v}_j: j\in\mathcal{J}\}<\infty$, then $\|h\|_{\mathfrak{v}}\leqslant\|\mathfrak{a}\mathfrak{v}\|_{\ell^\infty}\|h\|_{1/\mathfrak{a}}$ for all $h\in\mathbb{U}_{1/\mathfrak{a}}$, and hence $\mathbb{U}_{1/\mathfrak{a}}\subset\mathbb{U}_{\mathfrak{v}}$. On the other hand side, if $\|[\Phi]\|_{\ell^2_{\mathfrak{a}}}<\infty$, i.e., $\Phi\in\mathcal{L}_{\mathfrak{a}}$, then $\mathbb{U}_{1/\mathfrak{a}}\subset\mathcal{D}(\Phi)$. □

Keep in mind that in case of an orthogonal projection $f_m=\Pi_{\mathbb{U}_m}f$, $m\in\mathcal{M}$, we have $\operatorname{bias}^{\mathfrak{v}}_m(f)=\|\Pi_{\mathbb{U}_m}f-f\|_{\mathfrak{v}}=\left\|\Pi_{\mathbb{U}_m^\perp}f\right\|_{\mathfrak{v}}$ and $\operatorname{bias}^{\Phi}_m(f)=\sup\{|\Phi(\Pi_{\mathbb{U}_k^\perp}f)|,k\in\mathcal{M}\cap[\![m,\infty[\![\}$ where $\mathbb{U}_m^\perp$ denotes the orthogonal complement of $\mathbb{U}_m$ in $\mathbb{H}$.

§3.3.3 **Lemma**. *Consider the orthogonal projection $f_m=\Pi_{\mathbb{U}_m}f\in\mathbb{U}_m$ as theoretical approximation of $f\in\mathbb{F}^r_{u\mathfrak{a}}$. Given $\|\mathfrak{a}\mathfrak{v}\|_{\ell^\infty}<\infty$ for each $m\in\mathcal{M}$ let $(\mathfrak{a}\mathfrak{v})_{(m)}:=\|\mathfrak{a}\mathfrak{v}\mathbb{1}_{\mathcal{J}_m^c}\|_{\ell_\infty}=\sup\{\mathfrak{a}_j\mathfrak{v}_j,j\in\mathcal{J}_m^c\}\leqslant\|\mathfrak{a}\mathfrak{v}\|_{\ell^\infty}<\infty$, then $\operatorname{bias}^{\mathfrak{v}}_m(f)\leqslant r\,(\mathfrak{a}\mathfrak{v})_{(m)}$. On the other hand if $\Phi\in\mathcal{L}_{\mathfrak{a}}$ as in §2.2.8, then for each $m\in\mathcal{M}$, $\sum_{j\in\mathcal{J}_m^c}|[\Phi]_j|^2\mathfrak{a}_j^2=\|[\Phi]\mathbb{1}_{\mathcal{J}_m^c}\|^2_{\ell^2_{\mathfrak{a}}}\leqslant\|[\Phi]\|^2_{\ell^2_{\mathfrak{a}}}<\infty$ and $(\operatorname{bias}^{\Phi}_m(f))^2\leqslant r^2\,\|[\Phi]\mathbb{1}_{\mathcal{J}_m^c}\|^2_{\ell^2_{\mathfrak{a}}}$.*

Proof of Lemma §3.3.3 is given in the lecture. □

§3.3.4 **Definition** *(Linear Galerkin approach)*. Let $T\in\mathcal{T}(\mathbb{H})$, i.e., a compact and strictly positive definite operator in $\mathscr{L}(\mathbb{H})$, and $g\in\mathbb{H}$. An element $f_m\in\mathbb{U}_m$ satisfying

$$\langle f_m,Tf_m\rangle_{\mathbb{H}}-2\langle f_m,g\rangle_{\mathbb{H}}\leqslant\langle h,Th\rangle_{\mathbb{H}}-2\langle h,g\rangle_{\mathbb{H}}\quad\text{for all } h\in\mathbb{U}_m$$

is called a *Galerkin solution* in $\mathbb{U}_m$ of the equation $g=Tf$. □

§3.3.5 **Notations**.

(i) For $f \in \mathbb{H}$ considering the sequence of generalised Fourier coefficients $[f]$ as in §2.1.16 introduce its sub-vector $[f]_{\underline{m}} := ([f]_j)_{j \in \mathcal{J}_m}$, where $[\Pi_{\mathbb{U}_m} f]_{\underline{m}} = [f]_{\underline{m}}$.

(ii) For $T \in \mathscr{L}(\mathbb{H})$ denote by $[T]$ the (infinite) matrix with generic entries $[T]_{k,j} := \langle u_k, T u_j \rangle_{\mathbb{H}}$. For $m \in \mathcal{M}$, let $[T]_{\underline{m}}$ denote the $(|\mathcal{J}_m| \times |\mathcal{J}_m|)$-sub-matrix of $[T]$ given by $[T]_{\underline{m}} := ([T]_{k,j})_{j,k \in \mathcal{J}_m}$. Note that $[T^\star]_{\underline{m}} = [T]^t_{\underline{m}}$. Clearly, if we restrict $\Pi_{\mathbb{U}_m} T \Pi_{\mathbb{U}_m}$ to an operator from $\mathbb{U}_m$ to itself, then it can be represented by the matrix $[T]_{\underline{m}}$.

(iii) Given the identity $\mathrm{Id} \in \mathscr{L}(\mathbb{H})$ the $|\mathcal{J}_m|$-dimensional identity matrix is denoted by $[\mathrm{Id}]_{\underline{m}}$.

(iv) Let $\nabla_{\mathfrak{v}} = U^\star \mathrm{M}_{\mathfrak{v}} U : \mathbb{H} \supset \mathcal{D}(\nabla_{\mathfrak{v}}) \to \mathbb{H}$ be diagonal w.r.t. an unitary $U \in \mathscr{L}(\mathbb{H}, \ell(\mathcal{J}))$ (e.g., §2.2.4 (iii)) and multiplication operator $\mathrm{M}_{\mathfrak{v}} : \mathbb{K}^{\mathcal{J}} \to \mathbb{K}^{\mathcal{J}}$. Denote by $[\nabla_{\mathfrak{v}}]_{\underline{m}}$ the $|\mathcal{J}_m|$-dimensional diagonal matrix with diagonal entries $(\mathfrak{v}_j)_{j \in \mathcal{J}_m}$. Note that, $[\nabla_{\mathfrak{v}}]^s_{\underline{m}} = [\nabla_{\mathfrak{v}^s}]_{\underline{m}}$, $s \in \mathbb{R}$.

(v) Keep in mind the Euclidean norm $\|\cdot\|$ of a vector and the weighted norm $\|\cdot\|_{\mathfrak{v}}$ w.r.t. an ONS $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$. For all $f \in \mathbb{U}_m$ we have $\|f\|^2_{\mathfrak{v}} = [f]^t_{\underline{m}} [\nabla_{\mathfrak{v}^2}]_{\underline{m}} [f]_{\underline{m}} = \|[\nabla_{\mathfrak{v}}]_{\underline{m}} [f]_{\underline{m}}\|^2$.

(vi) Given a matrix M, let $\|M\|_s := \sup\{\|Mx\| : \|x\| \leqslant 1\}$ be its spectral norm then $\|\Pi_{\mathbb{U}_m} T \Pi_{\mathbb{U}_m}\|_{\mathscr{L}} = \|[T]_{\underline{m}}\|_s$ and hence $\|\Pi_{\mathbb{U}_m} \nabla^s_{\mathfrak{v}} \Pi_{\mathbb{U}_m}\|_{\mathscr{L}} = \max\{\mathfrak{v}^s_j, j \in \mathcal{J}_m\}$. □

§3.3.6 **Lemma**. *Let $T \in \mathcal{T}(\mathbb{H})$.* (i) *For all $m \in \mathbb{N}$ the matrix $[T]_{\underline{m}}$ is strictly positive definite.* (ii) *The Galerkin solution $f_m \in \mathbb{U}_m$ is uniquely determined by $[f_m]_{\underline{m}} = [T]^{-1}_{\underline{m}} [g]_{\underline{m}}$ and $[f_m]_j = 0$ for all $j \in \mathcal{J}^c_m$.* (iii) *The Galerkin solution f_m minimises in $\mathbb{U}_m$ the functional $F(h) = \|T^{1/2}(h - f)\|^2_{\mathbb{H}}$.*

Proof of Lemma §3.3.6 is given in the lecture. □

§3.3.7 **Remark**. Consider the orthogonal projection $\Pi_{\mathbb{U}_m} f$ and $\Pi_{\mathbb{U}^\perp_m} f$ of f onto the subspace $\mathbb{U}_m$ and $\mathbb{U}^\perp_m$, respectively, then the approximation error $\|\Pi_{\mathbb{U}_m} f - f\|_{\mathbb{H}} = \|\Pi_{\mathbb{U}^\perp_m} f\|_{\mathbb{H}}$ converges to zero as $m \to \infty$ by Lebesgue's dominated convergence theorem. On the other hand, the Galerkin solution $f_m \in \mathbb{U}_m$ satisfies $[\Pi_{\mathbb{U}_m} f - f_m]_{\underline{m}} = -[T]^{-1}_{\underline{m}} [T \Pi_{\mathbb{U}^\perp_m} f]_{\underline{m}}$ and, hence does generally not correspond to the orthogonal projection $\Pi_{\mathbb{U}_m} f$. Moreover, the approximation error $\sup\{\|f_m - f\|_{\mathbb{H}} : m \in [\![n, \infty[\![\cap \mathcal{M}\}$ does generally not converge to zero as $n \to \infty$. However, if $C := \{\|[T]^{-1}_{\underline{m}} [T \Pi_{\mathbb{U}^\perp_m} f]_{\underline{m}}\| : \|f\|_{\mathbb{H}} = 1, f \in \mathbb{H}, m \in \mathcal{M}\} < \infty$, then $\|f_m - f\|_{\mathbb{H}} \leqslant (1 + C) \|\Pi_{\mathbb{U}^\perp_m} f\|_{\mathbb{H}}$ which in turn implies $\lim_{n \to \infty} \sup\{\|f_m - f\|_{\mathbb{H}} : m \in [\![n, \infty[\![\cap \mathcal{M}\} = 0$. Here and subsequently, we will restrict ourselves to classes $\mathbb{F}$ and $\mathcal{T}$ of solutions and operators respectively which ensure the convergence. Obviously, this is a minimal regularity condition for us if we aim to estimate the Galerkin solution. □

§3.3.8 **Lemma**. *Given an ONB $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$, a nested sieve $(\mathcal{J}_m)_{m \in \mathcal{M}}$ in $\mathcal{J}$ and a strictly positive sequence $\mathfrak{t}$ consider the link condition $T \in \mathcal{T}^d_{\mathfrak{u}\mathfrak{t}}$ as in §2.2.50. Let $\mathfrak{t}$ be monotonically non-increasing, that is, $\min\{\mathfrak{t}_j, j \in \mathcal{J}_m\} \geqslant \sup\{\mathfrak{t}_j, j \in \mathcal{J}^c_m\} =: \mathfrak{t}_{(m)}$ for all $m \in \mathcal{M}$, then for all $0 \leqslant s \leqslant 1$ we have* (i) $\sup\{\mathfrak{t}^s_{(m)} \|[T]^{-s}_{\underline{m}}\|_s : m \in \mathcal{M}\} \leqslant \{d(d+2)\}^s \leqslant \{3d^2\}^s$, (ii) $\sup\{\|[T]^{-s}_{\underline{m}} [\nabla_{\mathfrak{t}}]^s_{\underline{m}}\|_s : m \in \mathcal{M}\} \leqslant \{d(d+2)\}^s \leqslant \{3d^2\}^s$ *and* (iii) $\sup\{\|[T]^s_{\underline{m}} [\nabla_{\mathfrak{t}}]^{-s}_{\underline{m}}\|_s : m \in \mathcal{M}\} \leqslant d^s$.

Proof of Lemma §3.3.8 is given in the lecture. □

§3.3.9 **Lemma** *(Bias of the Galerkin solution)*. *Given a strictly positive, monotonically non-increasing sequence $\mathfrak{t}$ consider $T \in \mathcal{T}^d_{\mathfrak{u}\mathfrak{t}}$ as in Lemma §3.3.8. Let in addition $f \in \mathbb{F}^r_{\mathfrak{u}\mathfrak{a}}$ with strictly* *positive, monotonically non-increasing sequence* $\mathfrak{a}$*, i.e.,* $\min\{\mathfrak{a}_j, j \in \mathcal{J}_m\} \geqslant \sup\{\mathfrak{a}_j, j \in \mathcal{J}_m^c\} =: \mathfrak{a}_{(m)}$ *for all* $m \in \mathcal{M}$*. If* f_m *denotes a Galerkin solution of* $g = Tf$ *then for each strictly positive sequence* $\mathfrak{v}$ *such that* $\mathfrak{a}\mathfrak{v}$ *is monotonically non-increasing, that is,* $\min\{\mathfrak{a}_j\mathfrak{v}_j, j \in \mathcal{J}_m\} \geqslant \sup\{\mathfrak{a}_j\mathfrak{v}_j, j \in \mathcal{J}_m^c\} =: (\mathfrak{a}\mathfrak{v})_{(m)}$ *for all* $m \in \mathcal{M}$*, we obtain for any* $m \in \mathcal{M}$ *and* $0 \leqslant s \leqslant 1$*,*

$$\|f - f_m\|_{\mathfrak{v}} \leqslant 4d^3\,(\mathfrak{v}\mathfrak{a})_{(m)} \max\big(1, (\mathfrak{t}/\mathfrak{v})_{(m)} \max\{\mathfrak{v}_j/\mathfrak{t}_j, j \in \mathcal{J}_m\}\big)\, \big\|\Pi_{\mathbb{U}_m^\perp} f\big\|_{1/\mathfrak{a}},$$
$$\|f_m\|_{1/\mathfrak{a}} \leqslant 3d^3\,\|f\|_{1/\mathfrak{a}}, \quad \text{and} \quad \|T^s(f - f_m)\|_{\mathbb{H}} \leqslant 4d^{3+s}(\mathfrak{a}\mathfrak{t}^s)_{(m)}\,\big\|\Pi_{\mathbb{U}_m^\perp} f\big\|_{1/\mathfrak{a}}. \quad (3.2)$$

Furthermore, for any $\Phi \in \mathcal{L}_{1/\mathfrak{a}}$ *we have*

$$|\Phi(f_m - f)|^2 \leqslant (4d^3)^2\,\big\|\Pi_{\mathbb{U}_m^\perp} f\big\|_{1/\mathfrak{a}}^2 \max\Big\{\sum_{j \in \mathcal{J}_m^c} |[\Phi]_j|^2\mathfrak{a}_j^2, (\mathfrak{t}^s\mathfrak{a})_{(m)}^2 \sum_{j \in \mathcal{J}_m^c} |[\Phi]_j|^2\mathfrak{t}_j^{-2s}\Big\}. \quad (3.3)$$

Proof of Lemma §3.3.9 is given in the lecture. □

§3.3.10 **Notations.** Let $\{u_j, j \in \mathcal{J}\}$, and $\{v_j, j \in \mathcal{J}\}$ be an ONS in $\mathbb{H}$ and $\mathbb{G}$, respectively, and let $(\mathcal{J}_m)_{m\in\mathcal{M}}$ be a nested sieve in $\mathcal{J}$.

(i) For $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ denote by $[T]$ the (infinite) matrix with generic entries $[T]_{k,j} := \langle v_k, Tu_j\rangle_{\mathbb{G}}$. For $m \in \mathcal{M}$, let $[T]_{\underline{m}} := ([T]_{k,j})_{k,j\in\mathcal{J}_m}$ denote the $(|\mathcal{J}_m| \times |\mathcal{J}_m|)$-sub-matrix of $[T]$. Note that $[T^\star]_{\underline{m}} = [T]_{\underline{m}}^t$.

(ii) Let $\mathbb{U}_m := \overline{lin}\,\{u_j, j \in \mathcal{J}_m\}$ and $\mathbb{V}_m := \overline{lin}\,\{v_j, j \in \mathcal{J}_m\}$ denote the linear subspaces of $\mathbb{H}$ and $\mathbb{G}$ spanned by the functions $\{u_j\}_{j\in\mathcal{J}_m}$ and $\{v_j\}_{j\in\mathcal{J}_m}$, respectively. Clearly, if we restrict $\Pi_{\mathbb{V}_m} T \Pi_{\mathbb{U}_m}$ to an operator from $\mathbb{U}_m$ to $\mathbb{V}_m$, then it can be represented by the matrix $[T]_{\underline{m}}$. □

§3.3.11 **Definition** *(Generalised linear Galerkin approach).* Given an ONS $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$, an ONS $\{v_j, j \in \mathcal{J}\}$ in $\mathbb{G}$, and a nested sieve $(\mathcal{J}_m)_{m\in\mathcal{M}}$ in $\mathcal{J}$ consider $T \in \mathscr{K}(\mathbb{H}, \mathbb{G})$ and $g \in \mathbb{G}$. Any element $f_m \in \mathbb{U}_m$ satisfying $[T]_{\underline{m}}[f_m]_{\underline{m}} = [g]_{\underline{m}}$ is called a *generalised Galerkin solution* in $\mathbb{U}_m$ of the equation $g = Tf$. □

§3.3.12 **Remark.** Throughout this note $[T]_{\underline{m}}$ is assumed to be non-singular for each $m \in \mathcal{M}$, so that $[T]_{\underline{m}}^{-1}$ always exists. We shall emphasise that it is a non-trivial problem to determine when such an assumption holds (cf. Efromovich and Koltchinskii [2001] and references therein). However, if $[T]_{\underline{m}}$ is non-singular, then the *generalised Galerkin solution* in $\mathbb{U}_m$ of the equation $g = Tf$ is unique and given by $[f_m]_{\underline{m}} = [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$. □

§3.3.13 **Definition** *(Generalised link condition).* Given an ONB $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and a strictly positive sequence $(\mathfrak{t}_j)_{j\in\mathcal{J}}$ consider the weighted norm $\|\cdot\|_{\mathfrak{t}} = \|\nabla_{\mathfrak{t}}\cdot\|_{\mathbb{H}}$ in $\mathbb{H}$. For all $d \geqslant 1$ define the subset $\mathcal{K}_{u,\mathfrak{t}}^d(\mathbb{H}, \mathbb{G}) := \Big\{T \in \mathscr{K}(\mathbb{H}, \mathbb{G}) : (T^\star T)^{1/2} \in \mathcal{T}_{u\mathfrak{t}}^d(\mathbb{H})\Big\}$. Given in addition an ONS $\{v_j, j \in \mathcal{J}\}$ in $\mathbb{G}$ and a nested Sieve $(\mathcal{J}_m)_{m\in\mathcal{M}}$ in $\mathcal{J}$ for $D \geqslant d$ we define $\mathcal{K}_{uv\mathfrak{t}}^{dD}(\mathbb{H}, \mathbb{G}) := \big\{T \in \mathcal{K}_{u,\mathfrak{t}}^d(\mathbb{H}, \mathbb{G}) : \big\|[\nabla_{\mathfrak{t}}]_{\underline{m}}[T]_{\underline{m}}^{-1}\big\| \leqslant D$ for all $m \in \mathcal{M}\big\}$ or $\mathcal{K}_{uv\mathfrak{t}}^{dD}$ for short. We say $T \in \mathscr{K}(\mathbb{H}, \mathbb{G})$ satisfies the *generalised link condition* $\mathcal{K}_{uv\mathfrak{t}}^{dD}$, if $T \in \mathcal{K}_{uv\mathfrak{t}}^{dD}$. □

§3.3.14 **Remark.** We shall emphasise that $\mathcal{K}_{uv\mathfrak{t}}^{dD}$ contains the subset $\mathcal{S}_{uv\mathfrak{t}}^d$ of all diagonal operator $\mathcal{S}_{uv}$ satisfying the link condition $\mathcal{K}_{u,\mathfrak{t}}^d$ (see §2.2.50), i.e., $\mathcal{S}_{uv\mathfrak{t}}^d = \mathcal{S}_{uv} \cap \mathcal{K}_{u,\mathfrak{t}}^d \subset \mathcal{K}_{uv\mathfrak{t}}^{dD}(\mathbb{H}, \mathbb{G})$. Keeping in mind Remark §2.2.51 an operator $T \in \mathcal{S}_{uv}$ admitting singular values $(\mathfrak{s}_j)_{j\in\mathcal{J}}$ satisfies the link condition $\mathcal{S}_{uv\mathfrak{t}}^d$ if and only if $d^{-1} \leqslant |\mathfrak{s}_j|/\mathfrak{t}_j \leqslant d$ for all $j \in \mathcal{J}$. Thereby, for any $m \in \mathcal{M}$ we have

$\big\|[\nabla_{\mathfrak{t}}]_{\underline{m}}[T]_{\underline{m}}^{-1}\big\| = \sup\{\mathfrak{t}_j/|\mathfrak{s}_j|, j \in \mathcal{J}_m\} \leqslant d \leqslant D$ and hence $\mathcal{S}^d_{uv\mathfrak{t}}(\mathbb{H},\mathbb{G}) \subset \mathcal{K}^{dD}_{uv\mathfrak{t}}(\mathbb{H},\mathbb{G})$. Moreover, there are operators in $\mathcal{K}^d_{u,\mathfrak{t}}(\mathbb{H},\mathbb{G})$ which do not belong to $\mathcal{S}^d_{uv\mathfrak{t}}$, i.e., they are not not diagonal w.r.t. $\mathcal{U}$ and $\mathcal{V}$ (see Remark §2.2.51). Furthermore, for each pre-specified ONB $(u_j)_{j\in\mathcal{J}}$ in $\mathbb{H}$ and $T \in \mathcal{K}^d_{u,\mathfrak{t}}(\mathbb{H},\mathbb{G})$ we can theoretically construct an ONS $(v_j)_{j\in\mathcal{J}}$ such that $\big\|[\nabla_{\mathfrak{t}}]_{\underline{m}}[T]_{\underline{m}}^{-1}\big\| \leqslant D$ holds for all $m \in \mathcal{M}$ and sufficiently large constant D. To be more precise, if $T \in \mathcal{K}^d_{u,\mathfrak{t}}(\mathbb{H},\mathbb{G})$, which involves only the ONB $(u_j)_{j\in\mathcal{J}}$, then the fundamental inequality of Heinz [1951] as given in §2.2.52 implies $\big\|(T^\star T)^{-1/2}u_j\big\|_{\mathbb{H}} \leqslant d\mathfrak{t}_j^{-1} < \infty$ for each $j \in \mathcal{J}$. Thereby, the function $(T^\star T)^{-1/2}u_j$ is an element of $\mathbb{H}$ and, hence $v_j := T(T^\star T)^{-1/2}u_j$, $j \in \mathcal{J}$ belongs to $\mathbb{G}$. Then it is easily checked that $(v_j)_{j\in\mathcal{J}}$ is an ONB of the closure of the range of T which may be completed to an ONB of $\mathbb{G}$. Keeping in mind that $\langle Tu_j, v_l\rangle_{\mathbb{G}} = \langle (T^\star T)^{1/2}u_j, u_l\rangle_{\mathbb{H}}$ for all $j,l \in \mathcal{J}$ it is obvious, that $[T]_{\underline{m}}$ is symmetric and moreover, strictly positive definite. Since $(T^\star T)^{1/2} \in \mathcal{T}^d_{u\mathfrak{t}}(\mathbb{H})$ from Lemma §3.3.8 (i) it follows $\big\|[\nabla_{\mathfrak{t}}]_{\underline{m}}[T]_{\underline{m}}^{-1}\big\|_s = \big\|[T]_{\underline{m}}^{-1}[\nabla_{\mathfrak{t}}]_{\underline{m}}\big\|_s \leqslant 3d^2$ for each $m \in \mathcal{M}$, which implies $T \in \mathcal{K}^{dD}_{uv\mathfrak{t}}(\mathbb{H},\mathbb{G})$ for all $D \geqslant 3d^2$. □

§3.3.15 **Lemma** *(Bias of the generalised Galerkin solution). Given an ONB $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$, an ONS $\{v_j, j \in \mathcal{J}\}$ in $\mathbb{G}$, a nested sieve $(\mathcal{J}_m)_{m\in\mathcal{M}}$ in $\mathcal{J}$, and a strictly positive, monotonically non-increasing sequence $\mathfrak{t}$ consider $T \in \mathcal{K}^{dD}_{uv\mathfrak{t}}$ as in §3.3.13. Let in addition $f \in \mathbb{F}^r_{u\mathfrak{a}}$ with strictly positive, monotonically non-increasing sequence $\mathfrak{a}$, i.e., $\min\{\mathfrak{a}_j, j \in \mathcal{J}_m\} \geqslant \sup\{\mathfrak{a}_j, j \in \mathcal{J}_m^c\} =: \mathfrak{a}_{(m)}$ for all $m \in \mathcal{M}$. If f_m denotes a generalised Galerkin solution of $g = Tf$ then for each strictly positive sequence $\mathfrak{v}$ such that $\mathfrak{a}\mathfrak{v}$ is monotonically non-increasing, that is, $\min\{\mathfrak{a}_j\mathfrak{v}_j, j \in \mathcal{J}_m\} \geqslant \sup\{\mathfrak{a}_j\mathfrak{v}_j, j \in \mathcal{J}_m^c\} =: (\mathfrak{a}\mathfrak{v})_{(m)}$ for all $m \in \mathcal{M}$, we obtain for any $m \in \mathcal{M}$ and $0 \leqslant s \leqslant 1$,*

$$\begin{aligned} \|f - f_m\|_{\mathfrak{v}} &\leqslant 2Dd\,(\mathfrak{v}\mathfrak{a})_{(m)} \max\big(1, (\mathfrak{t}/\mathfrak{v})_{(m)} \max\{\mathfrak{v}_j/\mathfrak{t}_j, j \in \mathcal{J}_m\}\big) \big\|\Pi_{\mathbb{U}_m^\perp} f\big\|_{1/\mathfrak{a}}, \\ \|f_m\|_{1/\mathfrak{a}} &\leqslant Dd\,\|f\|_{1/\mathfrak{a}}, \quad \textit{and} \quad \big\|(T^\star T)^{s/2}(f - f_m)\big\|_{\mathbb{H}} \leqslant 2Dd^{1+s}(\mathfrak{a}\mathfrak{t}^s)_{(m)} \big\|\Pi_{\mathbb{U}_m^\perp} f\big\|_{1/\mathfrak{a}}. \end{aligned} \tag{3.4}$$

Furthermore, for any $\Phi \in \mathcal{L}_{1/\mathfrak{a}}$ we have

$$|\Phi(f_m - f)|^2 \leqslant (2dD)^2 \big\|\Pi_{\mathbb{U}_m^\perp} f\big\|^2_{1/\mathfrak{a}} \max\Big\{\sum_{j\in\mathcal{J}_m^c} |[\Phi]_j|^2\mathfrak{a}_j^2, (\mathfrak{t}^s\mathfrak{a})^2_{(m)} \sum_{j\in\mathcal{J}_m^c} |[\Phi]_j|^2\mathfrak{t}_j^{-2s}\Big\}. \tag{3.5}$$

Proof of Lemma §3.3.15 is given in the lecture. □

Chapter 4

Statistical inverse problem

Throughout this note we consider the reconstruction of a functional parameter of interest f satsifying an equation $g = Tf$ based on noisy versions of g and T. In the sequel we formalise the meaning of a noisy version. First we consider the direct problem, that is, $T = \mathrm{Id}_{\mathbb{H}}$. Secondly, we assume the operator T is known in advance. In the last subsection the operator T is unknown and we introduce its noisy version.

4.1 Stochastic process on Hilbert spaces

In the sequel, $(\Omega, \mathscr{A}, \mathbb{P})$ is a probability space, where Ω will be interpreted as the set of elementary random events, $\mathscr{A}$ is a σ-algebra of subsets of Ω and $\mathbb{P}$ is a probability measure over $\mathscr{A}$. Here and subsequently, $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ and $\mathcal{U}$ denotes a separable Hilbert space and a subset of $\mathbb{H}$, respectively. Considering the product spaces $\mathbb{K}^{\mathbb{H}} = \times_{h\in\mathbb{H}} \mathbb{K}$ and $\mathbb{K}^{\mathcal{U}} = \times_{u\in\mathcal{U}} \mathbb{K}$ the mapping $\Pi_{\mathcal{U}} : \mathbb{K}^{\mathbb{H}} \to \mathbb{K}^{\mathcal{U}}$ given by $y = (y_h, h \in \mathbb{H}) \mapsto (y_u, u \in \mathcal{U}) =: \Pi_{\mathcal{U}} y$ is called canonical projection and for each $h \in \mathbb{H}$ in particular $\Pi_h : \mathbb{K}^{\mathbb{H}} \to \mathbb{K}$ given by $y = (y_{h'}, h' \in \mathbb{H}) \mapsto y_h =: \Pi_h y$ is called coordinate map. Moreover, $\mathscr{B}$ denotes the Borel-σ-algebra on $\mathbb{K}$ and $\mathbb{K}^{\mathbb{H}}$ is equipped with the product Borel-σ-algebra $\mathscr{B}^{\otimes\mathbb{H}} := \bigotimes_{h\in\mathbb{H}} \mathscr{B}$. Recall that $\mathscr{B}^{\otimes\mathbb{H}}$ equals the smallest σ-algebra such that all coordinate maps Π_h, $h \in \mathbb{H}$ are measurable. i.e., $\mathscr{B}^{\otimes\mathbb{H}} = \sigma(\Pi_h, h \in \mathbb{H})$.

§4.1.1 **Definition** *(Stochastic process on $\mathbb{H}$)*. Let $\{Y_h, h \in \mathbb{H}\}$ be a family of $\mathbb{K}$-valued r.v.'s on a common probability space $(\Omega, \mathscr{A}, \mathbb{P})$, that is, $Y_h : \Omega \to \mathbb{K}$ is a $\mathscr{A}$-$\mathscr{B}$-measurable mapping for each $h \in \mathbb{H}$. Consider the $\mathbb{K}^{\mathbb{H}}$-valued r.v. $Y := (Y_h, h \in \mathbb{H})$ where $Y : \Omega \to \mathbb{K}^{\mathbb{H}}$ is a $\mathscr{A}$-$\mathscr{B}^{\otimes\mathbb{H}}$-measurable mapping given by $\omega \mapsto (Y_h(\omega), h \in \mathbb{H}) =: Y(\omega)$. Y is called a *stochastic process* on $\mathbb{H}$. Its *distribution* $\mathbb{P}^Y := \mathbb{P} \circ Y^{-1}$ is the image probability measure of $\mathbb{P}$ under the map Y. Further, denote by $\mathbb{P}^{\Pi_{\mathcal{U}} Y}$ the distribution of the stochastic process $\Pi_{\mathcal{U}} Y = (Y_u, u \in \mathcal{U})$ on $\mathcal{U}$. The family $\{\mathbb{P}^{\Pi_{\mathcal{U}} Y}, \mathcal{U} \subset \mathbb{H} \text{ finite}\}$ is called family of the finite-dimensional distributions of Y or $\mathbb{P}^Y$. In particular, $\mathbb{P}^{Y_h} := \mathbb{P}^{\Pi_h Y}$ denotes the distribution of $Y_h = \Pi_h Y$. We write $\mathbb{E}(Y_h)$ and $\mathbb{V}\mathrm{ar}(Y_h) := \mathbb{E}((Y_h - \mathbb{E}(Y_h))\overline{(Y_h - \mathbb{E}(Y_h))})$, if it is exists, for the expectation and the variance of Y_h w.r.t. $\mathbb{P}^{Y_h}$, respectively. If Y_h has mean zero and variance then write $Y_h \sim \mathfrak{L}(0,1)$ for short. Furthermore, let $\mathbb{C}\mathrm{ov}(Y_h, Y_{h'}) := \mathbb{E}((Y_h - \mathbb{E}(Y_h))\overline{(Y_{h'} - \mathbb{E}(Y_{h'}))})$, if it is exists, for the covariance of Y_h and $Y_{h'}$ w.r.t. $\mathbb{P}^{\Pi_{\{h,h'\}} Y}$. □

§4.1.2 **Definition**. Let $Y := (Y_h, h \in \mathbb{H})$ be a stochastic process on $\mathbb{H}$. If $\mathbb{E}|Y_h| < \infty$ for each $h \in \mathbb{H}$ then the functional $\mu : \mathbb{H} \to \mathbb{K}$ with $h \mapsto \mathbb{E}(Y_h) =: \mu(h)$ is called *mean function* of Y. If the mean function μ is in addition linear and bounded, that is, $\mu \in \mathscr{L}(\mathbb{H}, \mathbb{K})$, then due to the Fréchet-Riesz representation theorem §2.2.6 there exists $\mu_Y \in \mathbb{H}$ such that $\mu(h) = \langle \mu_Y, h\rangle_{\mathbb{H}}$ for all $h \in \mathbb{H}$. The element $\mathbb{E}(Y) := \mu_Y$ is called *mean* or *expectation* of Y or $\mathbb{P}^Y$. If $\mathbb{E}|Y_h|^2 < \infty$ for each $h \in \mathbb{H}$ then the mapping $\mathrm{cov} : \mathbb{H} \times \mathbb{H} \to \mathbb{K}$ with $(h, h') \mapsto \mathbb{C}\mathrm{ov}(Y_h, Y_{h'}) =: \mathrm{cov}(h, h')$ is called *covariance function* of Y. If the covariance function cov is in addition a bounded bilinear form, then there is $\Gamma_Y \in \mathscr{L}(\mathbb{H})$ such that $\mathrm{cov}(h, h') = \langle \Gamma_Y h, h'\rangle_{\mathbb{H}} = \langle h, \Gamma_Y h'\rangle_{\mathbb{H}}$ for all $h, h' \in \mathbb{H}$. The operator Γ_Y is called *covariance operator* of Y or $\mathbb{P}^Y$. If Y admits a mean function μ and a covariance function cov then we write shortly $Y \sim \mathfrak{L}(\mu, \operatorname{cov})$. Analogously, $Y \sim \mathfrak{L}(\mu_Y, \Gamma)$ if there is an expectation $\mu_Y \in \mathbb{H}$ and a covariance operator $\Gamma_Y \in \mathscr{L}(\mathbb{H})$. □

§4.1.3 **Property**. *A covariance operator $\Gamma_Y \in \mathscr{L}(\mathbb{H})$ associated with a stochastic process Y on $\mathbb{H}$ is self-adjoint and non-negative definite.* □

§4.1.4 **Example** *(Non-parametric density estimation)*. Let X be a r.v. taking its values in the interval $[0,1]$ with distribution $\mathbb{P}$, c.d.f. $\mathbb{F}$ and admitting a Lebesgue-density $\mathbb{p} = \mathrm{d}\mathbb{P}/\mathrm{d}\lambda$. Given $h \in L^1_X$ as introduced in §2.1.4 (v) denote by $\mathbb{E}_{\mathbb{p}}(h(X)) = \mathbb{P}h = \lambda(h\mathbb{p})$ the expectation of $h(X)$ w.r.t. $\mathbb{P}$. For convenience we suppose that the density $\mathbb{p}$ is square integrable, i.e., $\mathbb{p}$ belongs to the real Hilbert space $L^2 := L^2([0,1])$ equipped with its usual inner product $\langle\cdot,\cdot\rangle_{L^2}$ (compare §2.1.4 (iv)). Thereby, for any $h \in L^2$ we have $\langle \mathbb{p}, h\rangle_{L^2} = \lambda(\mathbb{p}h) = \mathbb{P}h = \mathbb{E}_{\mathbb{p}}(h(X))$. Assuming an i.i.d. sample $X_i \sim \mathbb{p}$, $i \in [\![1,n]\!]$ we denote by $\mathbb{P}^{\otimes n}$ its joint product probability measure. Let $Y = (Y_h, h \in L^2)$ be the stochastic process on L^2 defined for each $h \in L^2$ by $Y_h := \overline{\mathbb{P}}_n h := \frac{1}{n}\sum_{i=1}^n h(X_i)$. Obviously, the mean function μ of Y satisfies $\mu(h) = \mathbb{E}(Y_h) = \mathbb{P}^{\otimes n}(\overline{\mathbb{P}}_n h) = \mathbb{P}h = \langle \mathbb{p}, h\rangle_{L^2}$ and hence, $Y_h = \langle \mathbb{p}, h\rangle_{L^2} + \frac{1}{\sqrt{n}}\dot{W}_h$ with $\dot{W}_h := n^{1/2}(\overline{\mathbb{P}}_n h - \mathbb{P}h)$. Moreover, the stochastic process $\dot{W} := (\dot{W}_h, h \in L^2)$ of error terms admits a covariance function given for all $h, h' \in L^2$ by $\mathbb{C}\mathrm{ov}(\dot{W}_h, \dot{W}_{h'}) = \mathbb{P}(hh') - \mathbb{P}h\mathbb{P}h' = \mathbb{P}((h - \mathbb{P}h)(h' - \mathbb{P}h')) = \mathbb{C}\mathrm{ov}(h(X), h(X'))$. Observe that $\mathbb{P}h\mathbb{P}h' = \langle \mathrm{M}_{\mathbb{p}}h, \mathbb{1}_{[0,1]}\rangle_{L^2}\langle \mathbb{1}_{[0,1]}, \mathrm{M}_{\mathbb{p}}h'\rangle_{L^2} = \langle \Pi_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_{\mathbb{p}}h, \mathrm{M}_{\mathbb{p}}h'\rangle_{L^2}$ and $\mathbb{P}(hh') - \mathbb{P}h\mathbb{P}h' = \langle \Gamma_{\mathbb{p}}h, h'\rangle_{L^2}$ with $\Gamma_{\mathbb{p}} = \mathrm{M}_{\mathbb{p}} - \mathrm{M}_{\mathbb{p}}\Pi_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_{\mathbb{p}}$, and thus, $\dot{W} \sim \mathfrak{L}(0, \Gamma_{\mathbb{p}})$ and consequently, $Y = \mathbb{p} + \frac{1}{n}\dot{W} \sim \mathfrak{L}(\mathbb{p}, \frac{1}{n}\Gamma_{\mathbb{p}})$. □

§4.1.5 **Example** *(Non-parametric regression)*. Let (X, Z) obey a non-parametric regression model $\mathbb{E}_f(X|Z) = f(Z)$ satisfying the Assumptions: (i) the regressor Z is uniformly distributed on the interval $[0,1]$, i.e., $Z \sim \mathfrak{U}[0,1]$; (ii) the centred error term $\varepsilon := X - f(Z)$, i.e., $\mathbb{E}_f(\varepsilon) = 0$, has a finite second moment $\sigma_\varepsilon^2 := \mathbb{E}_f(\varepsilon^2) < \infty$; (iii) ε and Z are independent; (iv) the regression function f is square integrable, i.e., $f \in L^2 := L^2([0,1])$. Given $h \in L^2$ denote by $\mathbb{E}_f(Xh(Z)) = \mathbb{P}_f[\mathrm{Id}\otimes h]$ with $[\mathrm{Id}\otimes h](X, Z) := Xh(Z)$ the expectation of $Xh(Z) = \{f(Z) + \varepsilon\}h(Z)$ w.r.t. the joint distribution $\mathbb{P}_f$ of (X, Z), where $\mathbb{E}_f[\varepsilon h(Z)] = 0$ and hence, $\mathbb{E}_f[Xh(Z)] = \mathbb{E}_f[f(Z)h(Z)] = \lambda(fh) = \langle f, h\rangle_{L^2}$. Assuming an i.i.d. sample (X_i, Z_i), $i \in [\![1,n]\!]$, from $\mathbb{P}_f$ we denote by $\mathbb{P}_f^{\otimes n}$ its joint product probability measure. Let $Y = (Y_h)_{h\in L^2}$ be the stochastic process on L^2 given for each $h \in L^2$ by $Y_h := \overline{\mathbb{P}}_n[\mathrm{Id}\otimes h] := n^{-1}\sum_{i=1}^n X_i h(Z_i)$. Obviously, the mean function μ of Y satisfies $\mu(h) = \mathbb{E}(Y_h) = \mathbb{E}_f[Xh(Z)] = \langle f, h\rangle_{L^2}$ and hence, $Y_h = \langle f, h\rangle_{L^2} + \frac{1}{\sqrt{n}}\dot{W}_h$ where $\dot{W}_h := n^{1/2}(\overline{\mathbb{P}}_n[\mathrm{Id}\otimes h] - \mathbb{P}_f[\mathrm{Id}\otimes h])$ is centred. The stochastic process $\dot{W} := (\dot{W}_h, h \in L^2)$ of error terms admits a covariance function given for $h, h' \in L^2$ by $\mathbb{C}\mathrm{ov}(\dot{W}_h, \dot{W}_{h'}) = \mathbb{P}_f([\mathrm{Id}\otimes h][\mathrm{Id}\otimes h']) - \mathbb{P}_f[\mathrm{Id}\otimes h]\mathbb{P}_f[\mathrm{Id}\otimes h'] = \mathbb{C}\mathrm{ov}(Xh(Z), Xh'(Z)) = \sigma_\varepsilon^2\langle h, h'\rangle_{L^2} + \langle \mathrm{M}_f h, \mathrm{M}_f h'\rangle_{L^2} - \langle \Pi_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_f h, \mathrm{M}_f h'\rangle_{L^2} = \sigma_\varepsilon^2\langle h, h'\rangle_{L^2} + \langle \mathrm{M}_f \Pi^\perp_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_f h, h'\rangle_{L^2} = \langle \Gamma_f h, h'\rangle_{L^2}$ with $\Gamma_f = \sigma_\varepsilon^2 \mathrm{Id}_{L^2} + \mathrm{M}_f\Pi^\perp_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_f$, and hence, $\dot{W} \sim \mathfrak{L}(0, \Gamma_f)$ and consequently, $Y = f + \frac{1}{n}\dot{W} \sim \mathfrak{L}(f, \frac{1}{n}\Gamma_f)$. □

§4.1.6 **Definition** *(White noise process on $\mathbb{H}$)*. Let $Y := (Y_h, h \in \mathbb{H})$ be a stochastic process on $\mathbb{H}$. If $\{Y_u, u \in \mathcal{U}\}$ for an ONS $\mathcal{U}$ in $\mathbb{H}$ is a family of $\mathbb{K}$-valued, independent and identically $\mathfrak{L}(0,1)$-distributed r.v.'s, i.e., $\mathbb{P}^{\Pi_{\mathcal{U}}Y} = \otimes_{u\in\mathcal{U}}\mathbb{P}^{Y_u} = \otimes_{u\in\mathcal{U}}\mathfrak{L}(0,1) =: \mathfrak{L}^{\otimes\mathcal{U}}(0,1)$, then we write shortly $\Pi_{\mathcal{U}}Y \sim \mathfrak{L}^{\otimes\mathcal{U}}(0,1)$ and call $\Pi_{\mathcal{U}}Y$ a *white noise process* on $\mathcal{U}$. If $\Pi_{\mathcal{U}}Y$ for any ONS $\mathcal{U}$ is a *white noise process* on $\mathcal{U}$ then we call Y a *white noise process* on $\mathbb{H}$. □

§4.1.7 **Remark**. Considering in example §4.1.4 or §4.1.5 the centred stochastic process $\dot{W} := (\dot{W}_h, h \in L^2)$ of error terms we note that generally there does not exists an ONB $\mathcal{U}$ in L^2 such that $\Pi_{\mathcal{U}}\dot{W}$ is a white noise process on $\mathcal{U}$. □

§4.1.8 **Property**. *Let $Y := (Y_h, h \in \mathbb{H})$ be a stochastic process on $\mathbb{H}$ admitting an expectation $\mu_Y \in \mathbb{H}$ and a covariance operator $\Gamma \in \mathscr{L}(\mathbb{H})$, i.e., $Y \sim \mathfrak{L}(\mu_Y, \Gamma)$. If there exists an ONB $\mathcal{U}$ in $\mathbb{H}$ such that $\Pi_{\mathcal{U}}Y$ is a white noise process on $\mathcal{U}$, i.e., $\Pi_{\mathcal{U}}Y \sim \mathfrak{L}^{\otimes\mathcal{U}}(0,1)$. Then we have $\mu_Y = 0 \in \mathbb{H}$ and $\Gamma = \mathrm{Id}_{\mathbb{H}}$ since $\mu_Y = \sum_{u\in\mathcal{U}}\langle\mu_Y, u\rangle_{\mathbb{H}} u = \sum_{u\in\mathcal{U}} \mathbb{E}(Y_u)u = 0$ and $\langle\Gamma\cdot,\cdot\rangle_{\mathbb{H}} = \sum_{u,u'\in\mathcal{U}}\langle u,\cdot\rangle_{\mathbb{H}}\langle\Gamma u, u'\rangle_{\mathbb{H}}\overline{\langle u',\cdot\rangle_{\mathbb{H}}} = \sum_{u,u'\in\mathcal{U}}\langle u,\cdot\rangle_{\mathbb{H}}\langle u, u'\rangle_{\mathbb{H}}\overline{\langle u',\cdot\rangle_{\mathbb{H}}} = \langle\cdot,\cdot\rangle_{\mathbb{H}}$. Consequently, for each ONB $\mathcal{V}$ in $\mathbb{H}$ the r.v.'s $\{Y_v, v \in \mathcal{V}\}$ are pairwise uncorrelated.* □

§4.1.9 **Definition** *(Gaussian process on $\mathbb{H}$)*. A stochastic process $Y = (Y_h, h \in \mathbb{H})$ on $\mathbb{H}$ with mean function μ and covariance function cov is called a *Gaussian process* on $\mathbb{H}$, if the family of finite-dimensional distributions $\left\{\mathbb{P}^{\Pi_{\mathcal{U}}Y}, \mathcal{U} \subset \mathbb{H} \text{ finite}\right\}$ of Y consists of normal distributions, that is, $\Pi_{\mathcal{U}}Y = (Y_u)_{u\in\mathcal{U}}$ is normally distributed with mean vector $(\mu(u))_{u\in\mathcal{U}}$ and covariance matrix $(\mathrm{cov}(u,u'))_{u,u'\in\mathcal{U}}$. We write shortly $Y \sim \mathfrak{N}(\mu, \mathrm{cov})$ or $Y \sim \mathfrak{N}(\mu_Y, \Gamma)$, if in addition there exist an expectation $\mu_Y \in \mathbb{H}$ and a covariance operator $\Gamma \in \mathscr{L}(\mathbb{H})$ associated with Y. The Gaussian process $Y \sim \mathfrak{N}(0, \langle\cdot,\cdot\rangle_{\mathbb{H}})$, or equivalently $Y \sim \mathfrak{N}(0, \mathrm{Id}_{\mathbb{H}})$, with mean $0 \in \mathbb{H}$ and covariance operator $\mathrm{Id}_{\mathbb{H}}$ is called *iso-Gaussian process* or *Gaussian white noise process* on $\mathbb{H}$. □

§4.1.10 **Property**. *Let $Y := (Y_h, h \in \mathbb{H})$ be a Gaussian process on $\mathbb{H}$ admitting an expectation $\mu_Y \in \mathbb{H}$ and a covariance operator $\Gamma \in \mathscr{L}(\mathbb{H})$, i.e., $Y \sim \mathfrak{N}(\mu_Y, \Gamma)$. If there exists an ONB $\mathcal{U}$ in $\mathbb{H}$ such that $\Pi_{\mathcal{U}}Y$ is a Gaussian white noise process on $\mathcal{U}$, i.e., $\Pi_{\mathcal{U}}Y \sim \mathfrak{N}^{\otimes\mathcal{U}}(0,1)$, then due to §4.1.8 we have $Y \sim \mathfrak{N}(0, \mathrm{Id}_{\mathbb{H}})$ and for each ONS $\mathcal{V}$ in $\mathbb{H}$ the standard normally distributed r.v.'s $\{Y_v, v \in \mathcal{V}\}$ are pairwise uncorrelated, and hence, independent, i.e., $\Pi_{\mathcal{V}}Y \sim \mathfrak{N}^{\otimes\mathcal{V}}(0,1)$.* □

§4.1.11 **Definition** *(Random function in $\mathbb{H}$)*. Let $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ be an Hilbert space equipped with its Borel-σ-algebra $\mathscr{B}_{\mathbb{H}}$, which is induced by its topology. An $\mathscr{A}$-$\mathscr{B}_{\mathbb{H}}$-measurable map $Y : \Omega \to \mathbb{H}$ is called an $\mathbb{H}$-valued r.v. or a *random function* in $\mathbb{H}$. □

§4.1.12 **Lemma**. *Let $\mathcal{U} = \{u_j, j \in \mathbb{N}\}$ be an ONS in $\mathbb{H}$. There does not exist a random function Y in $\mathbb{H}$ such that $\Pi_{\mathcal{U}}Y$ is a Gaussian white noise process on $\mathcal{U}$.*

Proof of Lemma §4.1.12 is given in the lecture. □

§4.1.13 **Properties**. *Let Y be a random function in $\mathbb{H}$.*

(i) *For each $h \in \mathbb{H}$, the map $\langle\cdot, h\rangle_{\mathbb{H}} : \mathbb{H} \to \mathbb{K}$ is continuous and hence, $\langle Y, h\rangle_{\mathbb{H}}$ a $\mathbb{K}$-valued r.v.. Thereby, $\langle Y, \bullet\rangle_{\mathbb{H}} := \{\langle Y, h\rangle_{\mathbb{H}}, h \in \mathbb{H}\}$ is a stochastic process on $\mathbb{H}$. If $\langle Y, \bullet\rangle_{\mathbb{H}}$ admits a mean function μ and a covariance function cov, then it is, respectively, linear, i.e., $\mu(ah + h') = \mathbb{E}(\langle Y, ah+h'\rangle_{\mathbb{H}}) = a\mu(h)+\mu(h')$, and bilinear. If in addition μ and cov are bounded, then there exists an expectation $\mathbb{E}(Y) \in \mathbb{H}$ and a covariance operator $\Gamma \in \mathscr{L}(\mathbb{H})$ such that $\mathbb{E}(\langle Y, h\rangle_{\mathbb{H}}) = \langle\mathbb{E}(Y), h\rangle_{\mathbb{H}}$ and $\mathbb{C}\mathrm{ov}(\langle Y, h\rangle_{\mathbb{H}}, \langle Y, h'\rangle_{\mathbb{H}}) = \langle\Gamma h, h'\rangle_{\mathbb{H}}$ for all $h, h' \in \mathbb{H}$.*

(ii) *If $\mathbb{E}(\|Y\|_{\mathbb{H}}) < \infty$, then $\mathbb{E}|\langle Y, h\rangle_{\mathbb{H}}| \leqslant \|h\|_{\mathbb{H}}\,\mathbb{E}(\|Y\|_{\mathbb{H}})$ for each $h \in \mathbb{H}$ due to the Cauchy-Schwarz-inequality §2.1.3, which in turn implies, that $\langle Y, \bullet\rangle_{\mathbb{H}}$ admits a bounded linear mean function μ and hence, there exists an expectation $\mathbb{E}(Y) \in \mathbb{H}$.*

(iii) *If $\mathbb{E}(\|Y\|_{\mathbb{H}}^2) < \infty$, then $\mathbb{V}\mathrm{ar}(\langle Y, h\rangle_{\mathbb{H}}) \leqslant \mathbb{E}|\langle Y, h\rangle_{\mathbb{H}}|^2 \leqslant \|h\|_{\mathbb{H}}^2\,\mathbb{E}(\|Y\|_{\mathbb{H}}^2)$ which in turn implies $|\mathbb{C}\mathrm{ov}(\langle Y, h\rangle_{\mathbb{H}}, \langle Y, h'\rangle_{\mathbb{H}})| \leqslant [\mathbb{V}\mathrm{ar}(\langle Y, h\rangle_{\mathbb{H}})\,\mathbb{V}\mathrm{ar}(\langle Y, h'\rangle_{\mathbb{H}})]^{1/2} \leqslant \|h\|_{\mathbb{H}}\,\|h'\|_{\mathbb{H}}\,\mathbb{E}(\|Y\|_{\mathbb{H}}^2)$.*

Thereby, $\langle Y, \bullet\rangle_{\mathbb{H}}$ admits a bounded, bilinear covariance function cov *and hence, there exists a covariance operator $\Gamma \in \mathscr{L}(H)$. Moreover, $\Gamma \in \mathscr{N}(\mathbb{H})$ since for any ONB $\mathcal{U}$ in $\mathbb{H}$ we have $\sum_{u\in\mathcal{U}}\langle\Gamma u, u\rangle_{\mathbb{H}} = \sum_{u\in\mathcal{U}} \mathbb{V}\mathrm{ar}(\langle Y, u\rangle_{\mathbb{H}}) = \mathbb{E}\sum_{u\in\mathcal{U}} |\langle Y - \mathbb{E}(Y), u\rangle_{\mathbb{H}}|^2 = \mathbb{E}\,\|Y - \mathbb{E}(Y)\|^2_{\mathbb{H}}$.* □

§4.1.14 **Notation.** Let Y be a random function in $\mathbb{H}$. If the associated stochastic process $\langle Y, \bullet\rangle_{\mathbb{H}}$ admits an expectation $\mu_Y \in \mathbb{H}$ and a covariance operator $\Gamma \in \mathscr{L}(\mathbb{H})$, then we write $Y \sim \mathfrak{L}(\mu_Y, \Gamma)$ with a slight abuse of notations. □

§4.1.15 **Example.** Let X be a random function in a real Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ having a finite second moment, i.e., $\mathbb{E}\,\|X\|^2_{\mathbb{H}} < \infty$. We say that X is centred if for all $h \in \mathbb{H}$ the real valued random variable $\langle X, h\rangle_{\mathbb{H}}$ has mean zero. Moreover, the linear operator $\Gamma : \mathbb{H} \to \mathbb{H}$ defined by $\langle\Gamma h_1, h_2\rangle_{\mathbb{H}} := \mathbb{E}[\langle h_1, X\rangle_{\mathbb{H}}\langle X, h_2\rangle_{\mathbb{H}}]$ for all $h_1, h_2 \in \mathbb{H}$ belongs to $\mathscr{N}(\mathbb{H})$ and satisfies $\mathrm{tr}(\Gamma) = \mathbb{E}\,\|X\|^2_{\mathbb{H}}$. Obviously, if the random function X is centred then $X \sim \mathfrak{L}(0, \Gamma)$, i.e., Γ is the covariance operator associated with X. In this situation the eigenvectors $\{u_j, j \in \mathcal{J}\}$ of T associated with the strictly positive eigenvalues $\{\lambda_j, j \in \mathcal{J}\}$ form an ONB in $\mathbb{H}$, and hence the corresponding generalised Fourier series transform $\mathcal{U}f = [f]$ is unitary. Furthermore, given the ONB of eigenfunctions the (infinite) matrix representation $[\Gamma] = [\nabla_\lambda]$ is diagonal, i.e., for all $m \in \mathcal{M}$, $[\Gamma]_{\underline{m}} = [\nabla_\lambda]_{\underline{m}}$ is a $|\mathcal{J}_m|$-dimensional diagonal matrix with entries $(\lambda_j)_{j\in\mathcal{J}_m}$. □

§4.1.16 **Notation.** Let $Y = \big(Y_{(h,g)}, h \in \mathbb{H}, g \in \mathbb{G}\big)$ be a *stochastic process on* $\mathbb{H} \times \mathbb{G}$, that is, a family $\big\{Y_{(h,g)}, h \in \mathbb{H}, g \in \mathbb{G}\big\}$ of $\mathbb{K}$-valued r.v.'s on a common probability space $(\Omega, \mathscr{A}, \mathbb{P})$. We call Y *centred* if $\mathbb{E}(Y_{(h,g)}) = 0$ for all $h \in \mathbb{H}$ and $g \in \mathbb{G}$. Moreover, if Y admits a *covariance function*, i.e., $\mathrm{cov}((h_1, g_1), (h_2, g_2)) = \mathbb{C}\mathrm{ov}(Y_{(h_1,g_1)}, Y_{(h_2,g_2)})$, for all $h_1, h_2 \in \mathbb{H}$ and $g_1, g_2 \in \mathbb{G}$, then we write $Y \sim \mathfrak{L}(0, \mathrm{cov})$, for short. Furthermore, if $\Pi_{(\mathcal{U}\times\mathcal{V})}Y = \big(Y_{(h,g)}, h \in \mathcal{U}, g \in \mathcal{V}\big)$ for an ONS $\mathcal{U}$ and $\mathcal{V}$ in $\mathbb{H}$ and $\mathbb{G}$, respectively, consists of $\mathbb{K}$-valued, independent and identically $\mathfrak{L}(0, 1)$-distributed r.v.'s, i.e., $\mathbb{P}^{\Pi_{(\mathcal{U}\times\mathcal{V})}Y} = \otimes_{u\in\mathcal{U}} \otimes_{v\in\mathcal{V}} \mathbb{P}^{Y_{(u,v)}} = \otimes_{u\in\mathcal{U}} \otimes_{v\in\mathcal{V}} \mathfrak{L}(0, 1) =: \mathfrak{L}^{\otimes(\mathcal{U}\times\mathcal{V})}(0, 1)$, then we write shortly $\Pi_{(\mathcal{U}\times\mathcal{V})}Y \sim \mathfrak{L}^{\otimes(\mathcal{U}\times\mathcal{V})}(0, 1)$ and call $\Pi_{(\mathcal{U}\times\mathcal{V})}Y$ a *white noise process* on $\mathcal{U} \times \mathcal{V}$. If $\Pi_{(U\times\mathcal{V})}Y$ for any ONS $\mathcal{U}$ in $\mathbb{H}$ and $\mathcal{V}$ in $\mathbb{G}$ is a *white noise process* on $\mathcal{U} \times \mathcal{V}$ then we call Y a *white noise process on* $\mathbb{H} \times \mathbb{G}$. Note that for a white noise process $Y \sim \mathfrak{L}(0, \mathrm{cov})$ on $\mathbb{H} \times \mathbb{G}$ holds $\mathrm{cov}((h_1, g_1), (h_2, g_2)) = \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\langle h_1, u_j\rangle_{\mathbb{H}}\langle u_j, h_2\rangle_{\mathbb{H}}\langle g_1, v_k\rangle_{\mathbb{G}}\langle v_k, g_2\rangle_{\mathbb{G}} = \langle h_1, h_2\rangle_{\mathbb{H}}\langle g_1, g_2\rangle_{\mathbb{G}}$ for any $h_1, h_2 \in \mathbb{H}$ and $g_1, g_2 \in \mathbb{G}$ and we write $Y \sim \mathfrak{L}(0, \langle\cdot,\cdot\rangle_{\mathbb{H}}\langle\cdot,\cdot\rangle_{\mathbb{G}})$. A centred stochastic process $Y \sim \mathfrak{L}(0, \mathrm{cov})$ on $\mathbb{H} \times \mathbb{G}$ is called a *Gaussian process* on $\mathbb{H} \times \mathbb{G}$, if the family of finite-dimensional distributions $\big\{\mathbb{P}^{\Pi_{(\mathcal{U}\times\mathcal{V})}Y}, \mathcal{U} \subset \mathbb{H}, \mathcal{V} \subset \mathbb{G} \text{ finite}\big\}$ of Y consists of normal distributions, that is, $\Pi_{(\mathcal{U}\times\mathcal{V})}Y = \big(Y_{(u,v)}, u \in \mathcal{U}, v \in \mathcal{V}\big)$ is normally distributed with mean vector zero and covariance matrix $(\mathrm{cov}((u, v), (u', v')))_{u,u'\in\mathcal{U},v,v'\in\mathcal{V}}$. We write shortly $Y \sim \mathfrak{N}(0, \mathrm{cov})$. If in addition $\mathrm{cov} = \langle\cdot,\cdot\rangle_{\mathbb{H}}\langle\cdot,\cdot\rangle_{\mathbb{G}}$, then Y is a white noise process and we call $Y \sim \mathfrak{N}(0, \langle\cdot,\cdot\rangle_{\mathbb{H}}\langle\cdot,\cdot\rangle_{\mathbb{G}})$ a *Gaussian white noise process* on $\mathbb{H} \times \mathbb{G}$. □

4.2 Statistical direct problem

Given a pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ we base our estimation procedure on the expansion of the function of interest $f \in \mathbb{U} = \overline{lin}(\mathcal{U})$. The choice of an adequate ONS is determined by the presumed information on the function of interest f formalised by the abstract smoothness conditions given in §2.1.18. However, the statistical selection of a basis from a family of bases (c.f. Birgé and Massart [1997]) is complicated, and its discussion is far beyond the scope of this lecture.

§4.2.1 **Definition** *(Sequence space model (SSM)).* Let $\dot{W} = (\dot{W}_h, h \in \mathbb{H})$ be a centred stochastic process on $\mathbb{H}$, and let $n \in \mathbb{N}$ be a sample size. The stochastic process $\widehat{f} = f + \frac{1}{\sqrt{n}}\dot{W}$ on $\mathbb{H}$ is called a noisy version of $f \in \mathbb{H}$ and we denote by $\mathbb{P}^n_f$ its distribution. If $\dot{W}$ admits a covariance operator (possibly depending on f), say Γ_f, then we eventually write $\widehat{f} \sim \mathfrak{L}(f, \frac{1}{n}\Gamma_f)$, or $\widehat{f} \sim \mathfrak{L}^n_f$ for short. Given the pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ considering the family of $\mathbb{K}$-valued r.v.'s $\{[\dot{W}]_j := \dot{W}_{u_j}, j \in \mathcal{J}\}$ the observable quantities take the form

$$[\widehat{f}]_j = \langle f, u_j\rangle_{\mathbb{H}} + \tfrac{1}{\sqrt{n}}\dot{W}_{u_j} = [f]_j + \tfrac{1}{\sqrt{n}}[\dot{W}]_j, \quad j \in \mathcal{J}. \tag{4.1}$$

We denote by $\mathbb{P}^n_{[f]}$, or $\mathfrak{L}([f], \frac{1}{n}[\Gamma_f])$, the distribution of the observable stochastic process $[\widehat{f}] = ([\widehat{f}]_j)_{j\in\mathcal{J}}$ on $\mathbb{U}$ which obviously is determined by the distribution $\mathbb{P}^n_f$, or $\mathfrak{L}(f, \frac{1}{n}\Gamma_f)$, of $\widehat{f}$. The reconstruction of the sequence $[f] = ([f]_j)_{j\in\mathcal{J}}$ and whence the function $f = U^\star[f] \in \mathbb{U}$ from the noisy version $[\widehat{f}] \sim \mathbb{P}^n_{[f]}$ is called a *(direct) sequence space model (SSM)*. □

§4.2.2 **Example** *(Gaussian sequence space model (GSSM)).* Given a Gaussian white noise process $\dot{W} = (\dot{W}_h, h \in \mathbb{H}) \sim \mathfrak{N}(0, \mathrm{Id}_{\mathbb{H}})$ on $\mathbb{H}$ as defined in §4.1.9 consider a noisy version $\widehat{f} = f + \frac{1}{\sqrt{n}}\dot{W} \sim \mathfrak{N}(f, \frac{1}{n}\mathrm{Id}_{\mathbb{H}}) = \mathfrak{N}^n_f$ of a function $f \in \mathbb{H}$. Given a pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ the observable quantities take the form $[\widehat{f}]_j = [f]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j$, $j \in \mathcal{J}$, where the error terms $[\dot{W}]_j, j \in \mathcal{J}$, are independent and $\mathfrak{N}(0,1)$-distributed, i.e., $[\dot{W}] = ([\dot{W}]_j)_{j\in\mathcal{J}} \sim \mathfrak{N}^{\otimes\mathcal{J}}(0,1) = \mathfrak{N}(0, \mathrm{Id}_{\mathcal{J}})$, and thus, $[\widehat{f}] = ([\widehat{f}]_j)_{j\in\mathcal{J}}$ is a sequence of independent Gaussian random variables having mean $[f]_j$ and variance n^{-1}, i.e., $[\widehat{f}] \sim \mathfrak{N}([f], \frac{1}{n}\mathrm{Id}_{\mathcal{J}}) = \mathfrak{N}^n_{[f]}$. The reconstruction of the sequence $[f]$ and whence the function $f = U^\star[f] \in \mathbb{U}$ from a noisy version $[\widehat{f}] \sim \mathfrak{N}^n_{[f]}$ is called a *Gaussian (direct) sequence space model (GSSM)*. □

§4.2.3 **Example** *(Non-parametric density estimation §4.1.4 continued).* For $n \in \mathbb{N}$ consider an i.i.d. sample $X_i \sim \mathbb{P}$, $i \in [\![1, n]\!]$, where $\mathbb{P}$ admits a Lebesgue-density $\mathbb{p} \in L^2 = L^2([0,1])$ and $\mathbb{P}^{\otimes n}$ denotes the associated joint product distribution. Consider the centred stochastic process $\dot{W} = (\dot{W}_h, h \in L^2) \sim \mathfrak{L}(0, \Gamma_{\mathbb{p}})$ of error terms with $\Gamma_{\mathbb{p}} = \mathrm{M}_{\mathbb{p}} - \mathrm{M}_{\mathbb{p}}\Pi_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_{\mathbb{p}}$ as introduced in §4.1.4. The non-parametric estimation of a density $\mathbb{p} \in L^2$ from an i.i.d. sample of size n may thus be based on the noisy version $\widehat{\mathbb{p}} = \mathbb{p} + \frac{1}{\sqrt{n}}\dot{W} \sim \mathfrak{L}(\mathbb{p}, \frac{1}{n}\Gamma_{\mathbb{p}})$ of the density of interest $\mathbb{p}$. In other words, given a pre-specified ONS $\{u_j, j \in \mathcal{J}\}$ the observable quantity $[\widehat{\mathbb{p}}] = ([\widehat{\mathbb{p}}]_j)_{j\in\mathcal{J}} \sim \mathbb{P}^n_{[\mathbb{p}]}$ takes for each $j \in \mathcal{J}$ with $[\dot{W}]_j := \dot{W}_{u_j}$ the form $[\widehat{\mathbb{p}}]_j = [\mathbb{p}]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j = \overline{\mathbb{P}}_n u_j$. Consequently, non-parametric estimation of a density can be covered by a sequence space model, where the error process $\dot{W}$, however, is generally not a white noise process. For convenient notations let $\{\mathbb{1}_{[0,1]}\} \cup \{u_j, j \in \mathbb{N}\}$ be an ONB of L^2 for some ONS $\mathcal{U} = \{u_j, j \in \mathbb{N}\}$. Keeping in mind that $\mathbb{p}$ is a density, it admits an expansion $\mathbb{p} = \mathbb{1}_{[0,1]} + U^\star[\mathbb{p}] = \mathbb{1}_{[0,1]} + \sum_{j\in\mathbb{N}}[\mathbb{p}]_j u_j$ where $[\mathbb{p}] = U\mathbb{p} = ([\mathbb{p}]_j)_{j\in\mathbb{N}}$ with $[\mathbb{p}]_j = \mathbb{E}_{\mathbb{p}}(u_j(X))$ for $j \in \mathbb{N}$ is a sequence of unknown coefficients, and hence, $f := \Pi_{\mathbb{U}}\mathbb{p} = U^\star[\mathbb{p}]$ is the function of interest. Given the pre-specified ONS $\mathcal{U}$ the observable quantity $[\widehat{\mathbb{p}}] = ([\widehat{\mathbb{p}}]_j)_{j\in\mathbb{N}} \sim \mathbb{P}^n_{[\mathbb{p}]}$ takes for each $j \in \mathbb{N}$ the form $[\widehat{\mathbb{p}}]_j = \overline{\mathbb{P}}_n u_j$. Note that the distribution $\mathbb{P}^n_{[\mathbb{p}]}$ of the observable quantity $[\widehat{\mathbb{p}}]$ is determined by the distribution $\mathbb{P}^{\otimes n}$ of the sample $X_1, \dots, X_n$. □

§4.2.4 **Example** *(Non-parametric regression §4.1.5 continued).* Consider $(X, Z) \sim \mathbb{P}_f$ obeying $\mathbb{E}_f(X|Z) = f(Z)$ and $Z \sim \mathfrak{U}[0,1]$ with $f \in L^2 = L^2([0,1])$. Given an i.i.d.. sample $(X_i, Z_i) \sim$

$\mathbb{P}_f$, $i \in [\![1,n]\!]$, their joint distribution is denoted by $\mathbb{P}_f^{\otimes n}$. Consider the centred stochastic process $\dot{W} = (\dot{W}_h, h \in L^2) \sim \mathfrak{L}(0, \Gamma_f)$ of error terms as introduced in §4.1.5. The non-parametric estimation of a regression function $f \in L^2$ from an i.i.d. sample of size n may thus be based on the noisy version $\widehat{f} = f + \frac{1}{\sqrt{n}}\dot{W} \sim \mathfrak{L}(f, \frac{1}{n}\Gamma_f)$ of the regression function f. In other words, given a pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ the observable quantity $[\widehat{f}] = ([\widehat{f}]_j)_{j\in\mathcal{J}} \sim \mathbb{P}^n_{[f]}$ takes for each $j \in \mathcal{J}$ the form $[\widehat{f}]_j = \overline{\mathbb{P}}_n[\mathrm{Id}\otimes u_j]$. Consequently, non-parametric regression can also be covered by a sequence space model, where the error process $\dot{W}$, however, is generally not a white noise process. □

4.3 Statistical inverse problem: known operator

Consider the reconstruction of a solution $f \in \mathbb{H}$ of an equation $g = Tf$ where the linear operator $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$ is known in advance. For ease of presentation we restrict ourselves to two cases only. First, we suppose $T \in \mathcal{T}(\mathbb{H}) \subset \mathscr{L}(\mathbb{H})$, i.e., T is compact and strictly positive definite, which is a rather mild assumption keeping in mind that f is a solution of the normal equation $T^\star g = T^\star T f$ and that $T^\star T$ is strictly positive definite and compact if T is injective and compact. Secondly, we assume $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G}) \subset \mathscr{L}(\mathbb{H}, \mathbb{G})$ admitting a singular system $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathcal{J}\}$ with eigenfunctions given by an ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ and $\mathcal{V} = \{v_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and $\mathbb{G}$, respectively. In both cases the same pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ is used to formalise the smoothing properties of the known operator T and the presumed information on the function of interest f given by an abstract smoothness condition, $f \in \mathbb{F}^r_{u\mathfrak{a}}$ as in Definition §2.1.18. In the first case the smoothing properties of the known operator T are characterised by a link condition, $T \in \mathcal{T}^d_{u\mathfrak{t}}$, as in Definition §2.2.50. We shall stress, that in this case T is generally not diagonal w.r.t. $\mathcal{U}$, in other words, T does generally not belong to $\mathcal{E}_u$ (see Definition §2.2.34). In the second case the choice of the ONS $\mathcal{U}$ and $\mathcal{V}$ is determined by the spectral decomposition of $T \in \mathcal{S}^d_{uv\mathfrak{t}}$, as in Definition §2.2.50.

§4.3.1 **Definition**. Given $T \in \mathcal{T}(\mathbb{H})$ consider the reconstruction of a solution $f \in \mathbb{H}$ from $g = Tf \in \mathbb{H}$. Let $\dot{W} = (\dot{W}_h, h \in \mathbb{H})$ be a centred stochastic process on $\mathbb{H}$, and let $n \in \mathbb{N}$ be a sample size. The stochastic process $\widehat{g} = Tf + \frac{1}{\sqrt{n}}\dot{W}$ on $\mathbb{H}$ is called a noisy version of $g = Tf \in \mathbb{H}$ and we denote by $\mathbb{P}^n_{Tf}$ its distribution. Keeping in mind that T is known in advance we may suppress the dependence of $\mathbb{P}^n_{Tf}$ on T and write $\mathbb{P}^n_f$, for short. If $\dot{W}$ admits a covariance operator (possibly depending on $g = Tf$), say Γ_{Tf}, then we eventually write $\widehat{g} \sim \mathfrak{L}(Tf, \frac{1}{n}\Gamma_f)$, or $\widehat{g} \sim \mathfrak{L}^n_{Tf}$ for short. The reconstruction of $f \in \mathbb{H}$ from a noisy version $\widehat{g} \sim \mathbb{P}^n_{Tf}$ is called a *statistical inverse problem*. Given the pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ considering the family of $\mathbb{K}$-valued random variables $\{[\dot{W}]_j := \dot{W}_{u_j}, j \in \mathcal{J}\}$ the observable quantities take the form

$$[\widehat{g}]_j = \langle Tf, u_j\rangle_{\mathbb{H}} + \tfrac{1}{\sqrt{n}}\dot{W}_{u_j} = [Tf]_j + \tfrac{1}{\sqrt{n}}[\dot{W}]_j. \quad j \in \mathcal{J}. \tag{4.2}$$

We denote by $\mathbb{P}^n_{[g]}$, or $\mathfrak{L}([g], \frac{1}{n}[\Gamma_g])$, the distribution of the observable stochastic process $[\widehat{g}] = ([\widehat{g}]_j)_{j\in\mathcal{J}}$ on $\mathbb{U}$ which obviously is determined by the distribution $\mathbb{P}^n_{Tf}$, or $\mathfrak{L}(Tf, \frac{1}{n}\Gamma_{Tf})$, of $\widehat{g}$. □

§4.3.2 **Example** *(Non-parametric inverse regression)*. Given $T \in \mathcal{T}(L^2([0,1]))$ let the dependence of a real r.v. X on the variation of an explanatory random variable Z be characterised by $X = [Tf](Z) + \varepsilon$, where f is an unknown function of interest and ε is an error term. The reconstruction of f from a sample of (X, Z) is called *non-parametric inverse regression*. For convenience, we assume that (i) the regressor Z is uniformly distributed on the interval $[0, 1]$, i.e., $Z \sim \mathfrak{U}[0, 1]$; (ii) the centred error term $\varepsilon := X - [Tf](Z)$, i.e., $\mathbb{E}_{Tf}(\varepsilon) = 0$, has a finite second moment $\sigma_\varepsilon^2 := \mathbb{E}_{Tf}(\varepsilon^2) < \infty$; (iii) ε and Z are independent; (iv) the inverse regression function f is square integrable, i.e., $f \in L^2 := L^2([0, 1])$, and hence $g := Tf \in L^2$. Given $h \in L^2$ denote by $\mathbb{E}_g(Xh(Z)) = \mathbb{P}_{Tf}[\mathrm{Id}\otimes h]$ with $[\mathrm{Id}\otimes h](X, Z) := Xh(Z)$ the expectation of $Xh(Z) = \{[Tf](Z) + \varepsilon\}h(Z)$ w.r.t. the joint distribution $\mathbb{P}_{Tf}$ of (X, Z), where $\mathbb{E}_g[\varepsilon h(Z)] = 0$ and hence, $\mathbb{E}_g[Xh(Z)] = \mathbb{E}_g[g(Z)h(Z)] = \lambda(gh) = \langle g, h\rangle_{L^2} = \langle Tf, h\rangle_{L^2}$. Assuming an i.i.d. sample (X_i, Z_i), $i \in [\![1, n]\!]$, from $\mathbb{P}_{Tf}$ we denote by $\mathbb{P}_{Tf}^{\otimes n}$ its joint product probability measure. Consider as noisy version of $g = Tf$ the stochastic process $\widehat{g}$ on L^2 given for each $h \in L^2$ by $\widehat{g}_h := \overline{\mathbb{P}}_{Tf}^n[\mathrm{Id}\otimes h] := n^{-1}\sum_{i=1}^n X_i h(Z_i)$. Obviously, the mean function μ of $\widehat{g}$ satisfies $\mu(h) = \mathbb{E}(\widehat{g}_h) = \mathbb{E}_g[Xh(Z)] = \langle Tf, h\rangle_{L^2}$ and hence, $\widehat{g}_h = \langle Tf, h\rangle_{L^2} + \frac{1}{\sqrt{n}}\dot{W}_h$ where $\dot{W}_h := n^{1/2}(\overline{\mathbb{P}}_{Tf}^n[\mathrm{Id}\otimes h] - \mathbb{P}_{Tf}[\mathrm{Id}\otimes h])$ is centred. Keeping in mind Example §4.1.5 the stochastic process $\dot{W} := (\dot{W}_h, h \in L^2)$ of error terms admits a covariance function given for $h, h' \in L^2$ by $\mathbb{C}\mathrm{ov}(\dot{W}_h, \dot{W}_{h'}) = \langle \Gamma_{Tf}h, h'\rangle_{L^2}$ with $\Gamma_{Tf} = \sigma_\varepsilon^2 \mathrm{Id}_{L^2} + \mathrm{M}_{Tf}\Pi^\perp_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_{Tf}$, i.e., $\dot{W} \sim \mathfrak{L}(0, \Gamma_{Tf})$ and consequently, $\widehat{g} = Tf + \frac{1}{n}\dot{W} \sim \mathfrak{L}(Tf, \frac{1}{n}\Gamma_{Tf}) = \mathfrak{L}_{Tf}^n$. Note that the error terms $\{\dot{W}_h, h \in L^2\}$ are centred, and generally not identically distributed. In other words, the reconstruction of f leads to a statistical inverse problem, where the error process $\dot{W}$ is generally not a white noise process. Given a pre-specified ONB $\mathcal{U}$ in L^2 and the $\mathbb{R}$-valued random variables $[\dot{W}]_j := \dot{W}_{u_j}, j \in \mathcal{J}$, the observable quantities take for each $j \in \mathcal{J}$ the form $[\widehat{g}]_j = \langle Tf, u_j\rangle_{L^2} + \frac{1}{\sqrt{n}}\dot{W}_{u_j} = [Tf]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j = \overline{\mathbb{P}}_{Tf}^n[\mathrm{Id}\otimes u_j]$ and we denote by $\mathfrak{L}([Tf], \frac{1}{n}[\Gamma_{Tf}])$ the joint distribution of the observable quantity $[\widehat{g}]$ which is obviously determined by the distribution $\mathbb{P}_{Tf}^{\otimes n}$ of the i.i.d. sample (X_i, Z_i), $i \in [\![1, n]\!]$. □

§4.3.3 **Example** *(Gaussian non-parametric inverse regression)*. Consider a Gaussian white noise process $\dot{W} = (\dot{W}_h, h \in \mathbb{H}) \sim \mathfrak{N}(0, \mathrm{Id}_{\mathbb{H}})$ on $\mathbb{H}$ as defined in §4.1.9. Given $T \in \mathcal{T}(\mathbb{H})$ the reconstruction of a function $f \in \mathbb{H}$ based on a noisy version $\widehat{g} = Tf + \frac{1}{\sqrt{n}}\dot{W} \sim \mathfrak{N}(Tf, \frac{1}{n}\mathrm{Id}_{\mathbb{H}}) = \mathfrak{N}_{Tf}^n$ is called *Gaussian non-parametric inverse regression*. Considering the projection onto an ONB $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ of $\mathbb{H}$ the observable quantities take consequently the form $[\widehat{g}]_j = [Tf]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j$, $j \in \mathcal{J}$, where the error terms $[\dot{W}]_j, j \in \mathcal{J}$, are independent and $\mathfrak{N}(0, 1)$-distributed, i.e., $[\dot{W}] = ([\dot{W}]_j)_{j\in\mathcal{J}} \sim \mathfrak{N}^{\otimes\mathcal{J}}(0, 1) = \mathfrak{N}(0, \mathrm{Id}_{\mathcal{J}})$, and thus, $[\widehat{g}] = ([\widehat{g}]_j)_{j\in\mathcal{J}}$ is a sequence of independent Gaussian random variables having mean $[Tf]_j$ and variance n^{-1}, i.e., $[\widehat{g}] \sim \mathfrak{N}_{[Tf]}^n = \mathfrak{N}([Tf], \frac{1}{n}\mathrm{Id}_{\mathcal{J}})$. □

§4.3.4 **Definition**. Given $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G})$ admitting a singular system $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathcal{J}\}$ consider the reconstruction of $f \in \mathbb{H}$ from a noisy version $\widehat{g} = Tf + \frac{1}{\sqrt{n}}\dot{W} \sim \mathbb{P}_{Tf}^n$, which is a statistical inverse problem as in Definition §4.3.1. A projection onto the ONS of eigenfunctions $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ and $\mathcal{V} = \{v_j, j \in \mathcal{J}\}$ allows to write $[g]_j = [Tf]_j = \langle Tf, v_j\rangle_{\mathbb{G}} = \mathfrak{s}_j\langle f, u_j\rangle_{\mathbb{H}} = \mathfrak{s}_j[f]_j$ for all $j \in \mathcal{J}$. Considering the family of $\mathbb{K}$-valued random variables $\{[\dot{W}]_j := \dot{W}_{v_j}, j \in \mathcal{J}\}$ the observable quantities take the form

$$[\widehat{g}]_j = \mathfrak{s}_j[f]_j + \tfrac{1}{\sqrt{n}}[\dot{W}]_j, \quad j \in \mathcal{J}. \tag{4.3}$$

We denote by $\mathbb{P}_{\mathfrak{s}[f]}^n$, or $\mathfrak{L}(\mathfrak{s}[f], \frac{1}{n}[\Gamma_f])$, the distribution of the observable stochastic process $[\widehat{g}] = ([\widehat{g}]_j)_{j\in\mathcal{J}}$ on $\mathbb{V} = \overline{lin}(\mathcal{V})$ which obviously is determined by the distribution $\mathbb{P}_{Tf}^n$, or $\mathfrak{L}(Tf, \frac{1}{n}\Gamma_{Tf})$, of $\widehat{g}$. The reconstruction of the sequence $[f] = ([f]_j)_{j\in\mathcal{J}}$ and whence the function $f = U^\star[f] \in$

$\mathbb{U} = \overline{lin}(\mathcal{U})$ from a noisy version $[\widehat{g}] \sim \mathbb{P}^n_{\mathfrak{s}[f]}$ is called an *indirect sequence space model (iSSM)*. Recall that it is called a *(direct) sequence space model* (see §4.2.1), if the sequence of singular values $\mathfrak{s}$ is equal to one, i.e., $\mathfrak{s}_j = 1$, for all $j \in \mathcal{J}$. In particular, if $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G})$ then the sequence $\mathfrak{s}$ has zero as an accumulation point and hence, the indirect sequence space model is *ill-posed*. □

§4.3.5 **Example** *(Gaussian indirect sequence space model (GiSSM)).* Given a Gaussian white noise process $\dot{W} = (\dot{W}_g, g \in \mathbb{G}) \sim \mathfrak{N}(0, \mathrm{Id}_{\mathbb{G}})$ on $\mathbb{G}$ as defined in §4.1.9 consider a noisy version $\widehat{g} = Tf + \frac{1}{\sqrt{n}}\dot{W} \sim \mathfrak{N}(Tf, \frac{1}{n}\mathrm{Id}_{\mathbb{G}}) = \mathfrak{N}^n_{Tf}$ of $g = Tf \in \mathbb{G}$. Given $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G})$ the reconstruction of the sequence $[f] = ([f]_j)_{j\in\mathcal{J}}$ and whence the function $f = U^\star[f] \in \mathbb{U}$ from observable quantities (4.3), where the error terms $\{[\dot{W}]_j, j \in \mathcal{J}\}$ are independent and $\mathfrak{N}(0,1)$-distributed, i.e., $[\dot{W}] = ([\dot{W}]_j)_{j\in\mathcal{J}} \sim \mathfrak{N}^{\otimes\mathcal{J}}(0,1)$, is called a *Gaussian indirect sequence space model (GiSSM)*. Recall that it is called a *Gaussian (direct) sequence space model (GSSM)* (see Example §4.2.2), if the sequence of singular values $\mathfrak{s}$ is equal to one, i.e., $\mathfrak{s}_j = 1$, for all $j \in \mathcal{J}$. □

§4.3.6 **Example** *(Circular deconvolution with known error density).* Let X be a circular random variable whose density $\mathbb{p}$ we are interested in, and ε an independent additive circular error with known density ϕ. Denote by $Y = X + \varepsilon$ the contaminated observation of X and by g its density. We will identify the circle with the unit interval $[0,1)$, for notational convenience. Thus, X and ε take their values in $[0,1)$. Let $\lfloor\cdot\rfloor$ be the floor function. Taking into account the circular nature of the data, the model can be written as $Y = X + \varepsilon - \lfloor X + \varepsilon\rfloor$ or equivalently $Y = X + \varepsilon \mod [0,1)$. Then, we have $g = \mathbb{p} \circledast \phi$ where $\circledast$ denotes *circular convolution* as in Examples §2.2.4 (ix) and, hence $g = C_\phi \mathbb{p}$ where the *convolution operator* $C_\phi \in \mathcal{L}(L^2([0,1)))$ is compact (see §2.2.35). If the error density ϕ and thus the operator C_ϕ are known in advance then the reconstruction of the density $\mathbb{p}$ given a sample from g is called *circular deconvolution with known error density*. Consider the *exponential basis* $\{e_j\}_{j\in\mathbb{Z}}$ in $L^2([0,1))$ introduced in §2.1.6 (ii) and let $[h]_j = \langle h, e_j\rangle_{L^2}$, $j \in \mathbb{Z}$, denote the Fourier coefficients of $h \in L^2([0,1))$. Applying the convolution theorem (see §2.2.35) we have $[g]_j = [\phi]_j[\mathbb{p}]_j$ with $[g]_j = \mathbb{E}_g e_j(-Y)$, $[\phi]_j = \mathbb{E}_\phi e_j(-\varepsilon)$ and $[\mathbb{p}]_j = \mathbb{E}_{\mathbb{p}} e_j(-X)$ for all $j \in \mathbb{Z}$. Assuming an iid. sample $Y_i \sim g$, $i = 1, \dots, n$, as in Example §4.2.3 consider a noisy version $\widehat{g} = g + \frac{1}{\sqrt{n}}\dot{W} \sim \mathfrak{L}(g, \frac{1}{n}\Gamma_g)$ of the density g with $\Gamma_g = \mathrm{M}_g - \mathrm{M}_g \Pi_{\{\mathbb{1}_{[0,1]}\}} \mathrm{M}_g$ as introduced in §4.1.4 where $\widehat{g}_h = \overline{\mathbb{P}}^n_g \overline{h} = \frac{1}{n}\sum_{i=1}^n \overline{h(Y_i)}$ for any $h \in L^2$. Given an arbitrary ONS $\{u_j, j \in \mathcal{J}\}$ the observable quantity $[\widehat{g}] = ([\widehat{g}]_j)_{j\in\mathcal{J}} \sim \mathbb{P}^n_{[g]}$ takes for each $j \in \mathcal{J}$ with $[\dot{W}]_j := \dot{W}_{u_j}$ the form $[\widehat{g}]_j = [g]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j = \overline{\mathbb{P}}^n_g \overline{u_j}$. Consequently, given the pre-specified exponential ONB $\{e_j, j \in \mathbb{Z}\}$ and the noisy version $\widehat{g}$ of $g = C_\phi \mathbb{p}$ the observable quantities are of the form $[\widehat{g}]_j = [\phi]_j[\mathbb{p}]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j$ for all $j \in \mathbb{Z}$, and thus, the reconstruction of $\mathbb{p}$ is an *ill-posed indirect sequence space model* where the error process $\dot{W}$, however, is generally not a white noise process. For convenient notatations let $\mathbb{Z}_o := \mathbb{Z}\setminus\{0\}$ and $\mathcal{U} = \{e_j, j \in \mathbb{Z}_o\}$ where $\{e_0 = \mathbb{1}_{[0,1]}\} \cup \{e_j, j \in \mathbb{Z}_o\}$ is the exponential ONB in L^2. Keeping in mind, that $\mathbb{p}$ is a density, it admits an expansion $\mathbb{p} = \mathbb{1}_{[0,1]} + U^\star[\mathbb{p}] = \mathbb{1}_{[0,1]} + \sum_{j\in\mathbb{Z}_o}[\mathbb{p}]_j e_j$ where $[\mathbb{p}] = U\mathbb{p} = ([\mathbb{p}]_j)_{j\in\mathbb{Z}_o}$ with $[\mathbb{p}]_j = \mathbb{E}_{\mathbb{p}} e_j(-X)$ for $j \in \mathbb{Z}_o$ is a sequence of unknown coefficients, and hence, $f := \Pi_{\mathbb{U}}\mathbb{p} = U^\star[\mathbb{p}]$ is the function of interest. Given the pre-specified ONS $\mathcal{U}$ the observable quantity $[\widehat{g}] = ([\widehat{g}]_j)_{j\in\mathbb{Z}_o} \sim \mathbb{P}^n_{[g]}$ takes for each $j \in \mathbb{Z}_o$ the form $[\widehat{g}]_j = \overline{\mathbb{P}}^n_g \overline{e_j}$. Note that the distribution $\mathbb{P}^n_{[g]}$ of the observable quantity $[\widehat{g}]$ is determined by the distribution $\mathbb{P}^{\otimes n}_g$ of the sample $Y_1, \dots, Y_n$. However, if the error density ϕ is known in advance, then $\mathbb{P}^n_{[g]}$ and $\mathbb{P}^{\otimes n}_g$ are uniquely determined by $\mathbb{p}$. □

4.4 Statistical inverse problems: partially known operator

Consider the reconstruction of a solution $f \in \mathbb{H}$ of an equation $g = Tf$ where the linear operator T belongs to $\mathcal{S}_{uv}(\mathbb{H}, \mathbb{G}) \subset \mathscr{L}(\mathbb{H}, \mathbb{G})$ for some pre-specified ONS of eigenfunctions $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ and $\mathcal{V} = \{v_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and $\mathbb{G}$, respectively. In other words the operator T admitts a singular system $\{(\mathfrak{s}_j, u_j, v_j), j \in \mathcal{J}\}$ where the eigenfunctions are known in advanced. However, there is only a noisy version $\widehat{\mathfrak{s}} = (\widehat{\mathfrak{s}}_j)_{j \in \mathcal{J}}$ of the sequence of the singular values $\mathfrak{s}$ available, and hence, the operator T is called partially known. In this situation the same pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ is again used to formalise the smoothing properties of the known operator T by a link condition, $T \in \mathcal{S}^d_{uvt}$, as in Definition §2.2.50, and the presumed information on the function of interest f given by an abstract smoothness condition, $f \in \mathbb{F}^r_{u\mathfrak{a}}$ as in Definition §2.1.18.

§4.4.1 **Definition**. Assume a statistical inverse problem $\widehat{g} = Tf + \frac{1}{\sqrt{n}}\dot{W}$ for some centred stochastic process $\dot{W} = (\dot{W}_h, h \in \mathbb{H})$ on $\mathbb{H}$, and sample size $n \in \mathbb{N}$, i.e., $\widehat{g} \sim \mathbb{P}^n_{Tf}$ or $\widehat{g} \sim \mathfrak{L}(Tf, \frac{1}{n}\Gamma_{Tf})$ if $\dot{W}$ admits a covariance operator Γ_{Tf}. Suppose further that $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G}) \subset \mathscr{L}(\mathbb{H}, \mathbb{G})$ for some pre-specified ONS of eigenfunctions $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ and $\mathcal{V} = \{v_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and $\mathbb{G}$, respectively. Given a centred stochastic process $[\dot{B}] = ([\dot{B}]_j)_{j \in \mathcal{J}}$ on $\ell^2(\mathcal{J})$ and a sample size $m \in \mathbb{N}$ for $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G})$ admitting a sequence of singular values $\mathfrak{s}$ the stochastic process $\widehat{\mathfrak{s}} = (\widehat{\mathfrak{s}}_j)_{j \in \mathcal{J}} = \mathfrak{s} + \frac{1}{\sqrt{m}}[\dot{B}] \sim \mathbb{P}^m_{\mathfrak{s}}$ is called a noisy version of $\mathfrak{s}$. If $[\dot{B}]$ admits a covariance function (possibly depending on $\mathfrak{s}$), say $\operatorname{cov}_{\mathfrak{s}}$, then we eventually write $\widehat{\mathfrak{s}} \sim \mathfrak{L}(\mathfrak{s}, \frac{1}{n}\operatorname{cov}_{\mathfrak{s}})$, or $\widehat{\mathfrak{s}} \sim \mathfrak{L}^n_{\mathfrak{s}}$ for short. The reconstruction of a solution $f \in \mathbb{H}$ from $g = Tf \in \mathbb{G}$ given a noisy version $\widehat{g} \sim \mathbb{P}^n_{Tf}$ and $\widehat{\mathfrak{s}} \sim \mathbb{P}^m_{\mathfrak{s}}$ of g and of the singular values $\mathfrak{s}$ of $T \in \mathcal{S}_{uv}(\mathbb{H}, \mathbb{G})$, respectively, is called *statistical inverse problem with partially known operator*. Projecting the inverse problem onto the pre-specified ONS $\mathcal{U}$ and $\mathcal{V}$ and hence obtaining $\mathbb{K}$-valued random variables $\{[\dot{W}]_k := \dot{W}_{v_k}, k \in \mathcal{K}\}$ the observable quantities take the form

$$[\widehat{g}]_j = \mathfrak{s}_j [f]_j + \tfrac{1}{\sqrt{n}}[\dot{W}]_j \quad \text{and} \quad \widehat{\mathfrak{s}}_j = \mathfrak{s}_j + \tfrac{1}{\sqrt{m}}[\dot{B}]_j, \quad j \in \mathcal{J}. \tag{4.4}$$

We denote by $\mathbb{P}^n_{\mathfrak{s}[f]}$, or $\mathfrak{L}(\mathfrak{s}[f], \frac{1}{n}[\Gamma_{Tf}])$, the distribution of the observable stochastic process $[\widehat{g}] = ([\widehat{g}]_j)_{j \in \mathcal{J}}$ on $\mathbb{V}$ which obviously is determined by the distribution $\mathbb{P}^n_{Tf}$. The reconstruction of the sequence $[f] = ([f]_j)_{j \in \mathcal{J}}$ and whence the function $f = U^\star[f] \in \mathbb{U} = \overline{\operatorname{lin}}(\mathcal{U})$ from the noisy versions $[\widehat{g}] \sim \mathbb{P}^n_{\mathfrak{s}[f]}$ and $\widehat{\mathfrak{s}} \sim \mathbb{P}^m$ is called an *indirect sequence space model with noise in the operator*. □

§4.4.2 **Example**. Consider as in §4.1.9 Gaussian white noise processes $\dot{W} = (\dot{W}_g, g \in \mathbb{G}) \sim \mathfrak{N}(0, \langle \cdot, \cdot \rangle_{\mathbb{G}})$ and $[\dot{B}] = ([\dot{B}]_j, j \in \mathcal{J}) \sim \mathfrak{N}^{\otimes \mathcal{J}}(0, 1)$ on $\mathbb{G}$ and ℓ^2, respectively. Given $T \in \mathcal{S}_{uv}(\mathbb{H})$ the reconstruction of a function $f = U^\star[f] \in \mathbb{U} = \overline{\operatorname{lin}}(\mathcal{U})$ based on observable quantities $[\widehat{g}] = \mathfrak{s}[f] + \frac{1}{\sqrt{n}}[\dot{W}] \sim \mathfrak{N}(\mathfrak{s}[f], \frac{1}{n}\operatorname{Id}_{\mathcal{J}}) = \mathfrak{N}^n_{\mathfrak{s}[f]}$ and $\widehat{\mathfrak{s}} = \mathfrak{s} + \frac{1}{\sqrt{m}}[\dot{W}] \sim \mathfrak{N}(\mathfrak{s}, \frac{1}{m}\operatorname{Id}_{\mathcal{J}}) = \mathfrak{N}^m_{\mathfrak{s}}$ is called *Gaussian indirect sequence space model with noise in the operator*. □

§4.4.3 **Example** *(Circular deconvolution with unknown error density)*. Consider a circular deconvolution problem §4.3.6 where neither the density $g = C_\phi \mathbb{p} = \phi \circledast \mathbb{p}$ of the contaminated observations, nor the error density ϕ is known in advance. The reconstruction of the density $\mathbb{p}$ based on two independent samples of independent and identically distributed random variables $Y_i \sim g$, $i \in [\![1, n]\!]$, and $\varepsilon_i \sim \phi$, $i \in [\![1, m]\!]$, of size $n \in \mathbb{N}$ and $m \in \mathbb{N}$, respectively, is called a *circular deconvolution problem with unknown error density*. Consider a noisy version

$\widehat{g} \sim \mathfrak{L}(g, \frac{1}{n}\Gamma_g)$ of $g = C_\phi \mathbb{p}$ as defined in §4.3.6, where $\widehat{g}_h = \overline{\mathbb{P}}_g^n \overline{h} = \frac{1}{n}\sum_{i=1}^n \overline{h(Y_i)}$ for any $h \in L^2$. In addition, given the i.i.d. sample $\varepsilon_i \sim \phi$, $i \in [\![1, m]\!]$, introduce as in Example §4.2.3 a noisy version $\widehat{\phi} = \phi + \frac{1}{\sqrt{m}}\dot{B} \sim \mathfrak{L}(\phi, \frac{1}{m}\Gamma_\phi)$ of the density ϕ with $\Gamma_\phi = \mathrm{M}_\phi - \mathrm{M}_\phi \Pi_{\{\mathbb{1}_{[0,1]}\}}\mathrm{M}_\phi$ as introduced in §4.1.4 where $\widehat{\phi}_h = \overline{\mathbb{P}}_\phi^m \overline{h} = \frac{1}{m}\sum_{i=1}^m \overline{h(\varepsilon_i)}$ for any $h \in L^2$. Keeping Example §2.2.35 in mind the *convolution operator* C_ϕ belongs to $\mathcal{E}_e(L^2([0,1)))$ w.r.t. the *exponential basis* $\{e_j, j \in \mathbb{Z}\}$ in $L^2([0,1))$ introduced in §2.1.6 (ii). In other words, any *convolution operator* C_ϕ has an eigen system $\{([\phi]_j, e_j), j \in \mathbb{Z}\}$ and for $j \in \mathbb{Z}$ we denote by $[\widehat{\phi}]_j := \overline{\mathbb{P}}_\phi^m \overline{e_j}$, the noisy version of $[\phi]_j = \mathbb{E}_\phi e_j(-\varepsilon)$ associated with $\widehat{\phi}$. Consequently, given the pre-specified exponential ONB $\{e_j, j \in \mathbb{Z}\}$ and the noisy version $\widehat{g}$ and $\widehat{\phi}$ of $g = C_\phi \mathbb{p}$ and ϕ, respectively, the observable quantities are of the form $[\widehat{g}]_j = [\phi]_j [\mathbb{p}]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j$ and $[\widehat{\phi}]_j = [\phi]_j + \frac{1}{\sqrt{m}}[\dot{B}]_j$ for all $j \in \mathbb{Z}$, and thus, the reconstruction of $\mathbb{p}$ is an *ill-posed indirect sequence space model with partially known operator*, where the error processes $\dot{W}$ and $\dot{B}$, however, are generally not white noise processes. For convenient notatations let $\mathbb{Z}_o := \mathbb{Z}\backslash\{0\}$ and $\mathcal{U} = \{e_j, j \in \mathbb{Z}_o\}$ where $\{e_0 = \mathbb{1}_{[0,1)}\} \cup \{e_j, j \in \mathbb{Z}_o\}$ is the exponential ONB in L^2. Keeping in mind, that $\mathbb{p}$ and ϕ are densities, they admit an expansion $\mathbb{p} = \mathbb{1}_{[0,1)} + U^\star[\mathbb{p}] = \mathbb{1}_{[0,1)} + \sum_{j \in \mathbb{Z}_o}[\mathbb{p}]_j e_j$ and $\phi = \mathbb{1}_{[0,1)} + U^\star[\phi]$ where $[\mathbb{p}] = U\mathbb{p} = ([\mathbb{p}]_j)_{j \in \mathbb{Z}_o}$ with $[\mathbb{p}]_j = \mathbb{E}_{\mathbb{p}} e_j(-X)$ for $j \in \mathbb{Z}_o$ is a sequence of unknown coefficients, and hence, $f := \Pi_{\mathbb{U}}\mathbb{p} = U^\star[\mathbb{p}] = U^\star([g]/[\phi])$ is the function of interest. Given the pre-specified ONS $\mathcal{U}$ the observable quantity $[\widehat{g}] = ([\widehat{g}]_j)_{j \in \mathbb{Z}_o} \sim \mathbb{P}_{[g]}^n$ and $[\widehat{\phi}] = ([\widehat{\phi}]_j)_{j \in \mathbb{Z}_o} \sim \mathbb{P}_{[\phi]}^m$, respectively, takes for each $j \in \mathbb{Z}_o$ the form $[\widehat{g}]_j = \overline{\mathbb{P}}_g^n \overline{e_j}$ and $[\widehat{\phi}]_j = \overline{\mathbb{P}}_\phi^m \overline{e_j}$. Note that the distribution $\mathbb{P}_{[g]}^n$ and $\mathbb{P}_{[\phi]}^m$ of the observable quantity $[\widehat{g}]$ and $[\widehat{\phi}]$ is determined, respectively, by the distribution $\mathbb{P}_g^{\otimes n}$ and $\mathbb{P}_\phi^{\otimes m}$ of the sample $Y_1, \dots, Y_n$ and $\varepsilon_1, \dots, \varepsilon_m$. □

4.5 Statistical inverse problems: unknown operator

Given a linear operator T belonging to $\mathscr{L}(\mathbb{H}, \mathbb{G})$ consider the reconstruction of a solution $f \in \mathbb{H}$ of an equation $g = Tf$ based on a noisy version $\widehat{g}$ and $\widehat{T}$ of g and T, repsectively, which we formalise next. In this situation the same pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ is again used to characterise the smoothing properties of the unknown operator T by a link condition, $T \in \mathcal{T}_{ut}^d$ as in Definition §2.2.50, or its generalisation, $T \in \mathcal{K}_{uvt}^{dD}$, as in Definition §3.3.13, and the presumed information on the function of interest f given by an abstract smoothness condition, $f \in \mathbb{F}_{ua}^r$ as in Definition §2.1.18.

§4.5.1 **Definition.** Given a centred stochastic process $\dot{B} = \big(\dot{B}_{(h,g)}, h \in \mathbb{H}, g \in \mathbb{G}\big)$ on $\mathbb{H} \times \mathbb{G}$ and a sample size $m \in \mathbb{N}$ the stochastic process on $\mathbb{H} \times \mathbb{G}$ for $h \in \mathbb{H}$ and $g \in \mathbb{G}$ satisfying $\widehat{T}_{(h,g)} = \langle g, Th\rangle_{\mathbb{G}} + \frac{1}{\sqrt{m}}\dot{B}_{(h,g)}$, or $\widehat{T} = T + \frac{1}{\sqrt{m}}\dot{B}$ for short, is called a noisy version of $T \in \mathscr{L}(\mathbb{H}, \mathbb{G})$. We denote its distribution by $\mathbb{P}_T^m$. If $\dot{B}$ admits a covariance function (possibly depending on T), say cov_T, then we eventually write $\widehat{T} \sim \mathfrak{L}(T, \frac{1}{n}\mathrm{cov}_T)$, or $\widehat{T} \sim \mathfrak{L}_T^n$ for short. The reconstruction of a solution $f \in \mathbb{H}$ from $g = Tf \in \mathbb{G}$ given a noisy version $\widehat{g} = g + \frac{1}{\sqrt{n}}\dot{W} \sim \mathbb{P}_{Tf}^n$ of g and a noisy version $\widehat{T} = T + \frac{1}{\sqrt{m}}\dot{B} \sim \mathbb{P}_T^n$ of T is called *statistical inverse problem with unknown operator*. Given a pre-specified ONS $\mathcal{U} = \{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ and $\mathcal{V} = \{v_k, k \in \mathcal{K}\}$ in $\mathbb{G}$ considering the families of $\mathbb{K}$-valued random variables $\big\{[\dot{W}]_k := \dot{W}_{v_k}, k \in \mathcal{K}\big\}$ and $\big\{[\dot{B}]_{k,j} := \dot{B}_{(u_j, v_k)}, k \in \mathcal{K}, j \in \mathcal{J}\big\}$ the observable quantities take the form

$$[\widehat{g}]_k = \langle Tf, v_k\rangle_{\mathbb{H}} + \tfrac{1}{\sqrt{n}}\dot{W}_{v_k} = [Tf]_k + \tfrac{1}{\sqrt{n}}[\dot{W}]_k \quad \text{and}$$
$$[\widehat{T}]_{k,j} = \langle v_k, Tu_j\rangle_{\mathbb{G}} + \tfrac{1}{\sqrt{m}}\dot{B}_{(u_j,v_k)} = [T]_{k,j} + \tfrac{1}{\sqrt{m}}[\dot{B}]_{k,j}, \quad j \in \mathcal{J}, k \in \mathcal{K}. \quad \square \tag{4.5}$$

We denote by $\mathbb{P}^n_{[Tf]}$, or $\mathfrak{L}([Tf], \frac{1}{n}[\Gamma_{Tf}])$, and $\mathbb{P}^m_{[T]}$, or $\mathfrak{L}([T], \frac{1}{m}[\mathrm{cov}_T])$, the distribution of the observable stochastic process $[\widehat{g}] = ([\widehat{g}]_k)_{k\in\mathcal{K}}$ on $\mathbb{V}$ and $[\widehat{T}] = ([\widehat{T}]_{k,j})_{j\in\mathcal{J},k\in\mathcal{K}}$ on $\mathbb{U}\times\mathbb{V}$ which obviously is determined by the distribution $\mathbb{P}^n_{Tf}$ and $\mathbb{P}^m_T$ of $\widehat{g}$ and $\widehat{T}$, respectively. □

§4.5.2 **Example**. Let $T \in \mathcal{T}(\mathbb{H})$ and $\{u_j, j \in \mathbb{N}\}$ be an ONB in $\mathbb{H}$ not necessarily corresponding to the eigenfunctions of T. The reconstruction of a function $f \in \mathbb{H}$ based on noisy versions $\widehat{g} = Tf + \frac{1}{\sqrt{n}}\dot{W}$ and $\widehat{T} = T + \frac{1}{\sqrt{m}}\dot{B}$ of $g = Tf \in \mathbb{H}$ and T, respectively, where $\dot{W} \sim \mathfrak{N}(0, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ and $\dot{B} \sim \mathfrak{N}(0, \langle\cdot,\cdot\rangle_{\mathbb{H}}\langle\cdot,\cdot\rangle_{\mathbb{G}})$ are Gaussian white noise processes on $\mathbb{H}$ and $\mathbb{H}\times\mathbb{H}$, is called *Gaussian non-parametric inverse regression with unknown operator*. Projecting onto $\{u_j, j \in \mathcal{J}\}$ the observable quantities take the form $[\widehat{g}]_j = [g]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j$ and $[\widehat{T}]_{j,k} = [T]_{j,k} + \frac{1}{\sqrt{m}}[\dot{B}]_{j,k}$, for $j, k \in \mathcal{J}$, where the error terms $\{[\dot{W}]_j, [\dot{B}]_{j,k}, j, k \in \mathbb{N}\}$ are independent and $\mathfrak{N}(0,1)$-distributed. □

§4.5.3 **Example** *(Non-parametric functional linear regression)*. Let X be a random function taking its values in a separable Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$. For convenient notations we assume that $X \sim \mathfrak{L}(0,\Gamma)$ with $\mathrm{tr}(\Gamma) = \mathbb{E}\,\|X\|^2_{\mathbb{H}} < \infty$ (see Example §4.1.15). The linear relationship between a real random variable Y and the variation of X is expressed by the equation $Y = \langle f, X\rangle_{\mathbb{H}} + \varepsilon$, with an unknown slope function $f \in \mathbb{H}$ and a real-valued and centred error term ε. The reconstruction of the slope parameter f given a sample of (Y, X) is called *non-parametric functional linear regression*. We suppose that the regressor X is uncorrelated to the random error ε in the sense that $\mathbb{E}(\varepsilon\langle X, h\rangle_{\mathbb{H}}) = 0$ for all $h \in \mathbb{H}$. Multiplying both sides in the model equation by X and taking the expectation leads for any $h \in \mathbb{H}$ to the normal equation $\langle g, h\rangle_{\mathbb{H}} := \mathbb{E}(Y\langle X, h\rangle_{\mathbb{H}}) = \mathbb{E}(\langle f, X\rangle_{\mathbb{H}}\langle X, h\rangle_{\mathbb{H}}) = \langle \Gamma f, h\rangle_{\mathbb{H}}$, or $g = \mathbb{E}(YX) = \mathbb{E}(\langle f, X\rangle_{\mathbb{H}} X) = \mathbb{E}(X \otimes X)f = \Gamma f$, for short, where the cross-correlation function g belongs to $\mathbb{H}$. Let us denote by $\mathbb{P}_{f,\Gamma}$ the distribution of (Y, X). Assuming an iid. sample $\{(Y_i, X_i), i = 1, \dots, n\}$ of (Y, X), it is natural to consider the estimators $\widehat{g} := \frac{1}{n}\sum_{i=1}^n Y_i X_i$ and $\widehat{\Gamma} := \frac{1}{n}\sum_{i=1}^n X_i \otimes X_i$ of g and Γ respectively. Note that $\widehat{g} = g + \frac{1}{\sqrt{n}}\dot{W}$ with $\dot{W} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i X_i - g)$ and $\widehat{\Gamma} = \Gamma + \frac{1}{\sqrt{n}}\dot{B}$ with $\dot{B} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i \otimes X_i - \Gamma)$ is a noisy version of g and Γ, where $\dot{W}$ and $\dot{B}$ are centred but generally not white noise processes. We denote by $\mathfrak{L}^n_{\Gamma f}$ and $\mathfrak{L}^n_{\Gamma}$ the distribution of $\widehat{g}$ and $\widehat{\Gamma}$, respectively. Given the noisy versions $\widehat{g}$ of $g = \Gamma f$ and $\widehat{\Gamma}$ of Γ the reconstruction of f is hence a *statistical inverse problem with unknown operator* where the observable quantities given an ONB $\{u_j, j \in \mathcal{J}\}$ in $\mathbb{H}$ take the form $[\widehat{g}]_k = [\Gamma f]_k + \frac{1}{\sqrt{n}}[\dot{W}]_k$ and $[\widehat{\Gamma}]_{k,j} = [\Gamma]_{k,j} + \frac{1}{\sqrt{n}}[\dot{B}]_{k,j}$ with $[\dot{W}]_k := \frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i[X_i]_k - [\Gamma f]_k\}$ and $[\dot{B}]_{k,j} := \frac{1}{\sqrt{n}}\sum_{i=1}^n \{[X_i]_j[X_i]_k - [\Gamma]_{k,j}\}$ for all $j, k \in \mathcal{J}$. □

§4.5.4 **Example**. A structural function f characterises the dependency of a real response Y on the variation of an $\mathbb{R}^p$-valued endogenous explanatory random variable X by $Y = f(X) + U$ where $\mathbb{E}[U|X] \neq 0$ for some error term U. In other words, the structural function equals not the conditional mean function of Y given X. In *non-parametric instrumental regression*, however, a sample from (Y, X, Z) is available, where Z is an additional $\mathbb{R}^q$-valued random vector of exogenous *instruments* such that $\mathbb{E}[U|Z] = 0$. It is convenient to rewrite the model equations in terms of an operator between Hilbert spaces. Therefore, let us first recall the Hilbert spaces $(L^2_X, \langle\cdot,\cdot\rangle_{L^2_X})$ and $(L^2_Z, \langle\cdot,\cdot\rangle_{L^2_Z})$ defined in §2.1.4 (v). Taking the conditional expectation w.r.t. the instrument Z on both sides in the model equation yields $g := \mathbb{E}[Y|Z] = \mathbb{E}[f(X)|Z] =: Kf$ where the regression function g belongs to L^2_Z and K is the conditional expectation of X given Z assumed to be an element of $\mathscr{K}(L^2_X, L^2_Z)$ (compare §2.2.4 (vii)). Keep in mind that for $u \in L^2_X$ and $v \in L^2_Z$ we have $\langle g, v\rangle_{L^2_Z} = \mathbb{E}(Yv(Z)) = \mathbb{P}_{Kf}[\mathrm{Id}\otimes v]$ and $\langle v, Ku\rangle_{L^2_Z} = \mathbb{E}(u(X)v(Z)) = \mathbb{P}_K[u \otimes v]$ where $[u \otimes v](X, Z) := u(X)v(Z)$. Assuming an iid. sample $\{(Y_i, X_i, Z_i), i = 1, \dots, n\}$ of (Y, X, Z), it is natural to consider a noisy version $\widehat{g}$ and $\widehat{K}$ of g and K, respectively, for $u \in L^2_X$ and $v \in L^2_Z$ given by $\widehat{g}_v = \overline{\mathbb{P}}^n_{Kf}[\mathrm{Id}\otimes v] := n^{-1}\sum_{i=1}^n Y_i v(Z_i) = \langle Kf, v\rangle_{L^2_Z} + \frac{1}{\sqrt{n}}\dot{W}_v$ and $(\widehat{K})_{u,v} = \overline{\mathbb{P}}^n_K[u \otimes v] := n^{-1}\sum_{i=1}^n u(X_i)v(Z_i) = \langle v, Ku\rangle_{L^2_Z} + \frac{1}{\sqrt{n}}\dot{B}_{u,v}$ where $\dot{W}_v := n^{1/2}(\overline{\mathbb{P}}^n_{Kf}[\mathrm{Id}\otimes v] - \mathbb{P}_{Kf}[\mathrm{Id}\otimes v])$ and $\dot{B}_{u,v} := n^{1/2}(\overline{\mathbb{P}}^n_K[u\otimes v] - \mathbb{P}_K[u\otimes v])$ are centred. Note that $\dot{W}$ and $\dot{B}$ are centred but generally not white noise processes. Given the noisy versions $\widehat{g}$ of $g = Kf$ and $\widehat{K}$ of K only the reconstruction of f is a *statistical inverse problem with unknown operator* where the observable quantities given an ONB $\mathcal{U} = \{u_j, j \in \mathbb{N}\}$ in L^2_X and $\mathcal{V} = \{v_j, j \in \mathbb{N}\}$ in L^2_Z take the form $[\widehat{g}]_j = [Kf]_j + \frac{1}{\sqrt{n}}[\dot{W}]_j$ and $[\widehat{K}]_{j,k} = [K]_{j,k} + \frac{1}{\sqrt{n}}[\dot{B}]_{j,k}$ with $[\dot{W}]_j = \dot{W}_{v_j}$ and $[\dot{B}]_{j,k} = \dot{B}_{u_k,v_j}$ for all $j, k \in \mathbb{N}$. □

Bibliography

L. Birgé and P. Massart. From model selection to adaptive estimation. Pollard, David (ed.) et al., Festschrift for Lucien Le Cam: research papers in probability and statistics. New York, NY: Springer. 55-87, 1997.

N. Dunford and J. T. Schwartz. *Linear Operators, Part I: General Theory*. Wiley Classics Library. John Wiley & Sons Ltd, New York, 1988a.

N. Dunford and J. T. Schwartz. *Linear operators. Part II: Spectral theory, self adjoint operators in Hilbert space.* Wiley Classics Library. John Wiley & Sons Ltd, New York, 1988b.

N. Dunford and J. T. Schwartz. *Linear operators. Part III, Spectral Operators.* Wiley Classics Library. John Wiley & Sons Ltd, New York, 1988c.

S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems.* Kluwer Academic, Dordrecht, 2000.

J. Hadamard. *Le Problème de Cauchy et les Équations aux Dérivées Partielles Linéaires Hyperboliques*. Paris, Hermann, 1932.

P. R. Halmos. What does the spectral theorem say? *Amer. Math. Monthly*, 70:241–247, 1963.

E. Heinz. Beiträge zur Störungstheorie der Spektralzerlegung. *Mathematische Annalen*, 123: 415–438, 1951.

T. Kawata. *Fourier analysis in probability theory.* Academic Press, New York, 1972.

S. Krein and Y. I. Petunin. Scales of banach spaces. In *Russian Math. Surveys*, volume 21, pages 85–169, 1966.

R. Kress. *Linear integral equations*, volume 82 of *Applied Mathematical Sciences.* Springer, New York, NY, 2 edition, 1989.

A. B. Tsybakov. *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York, 2009.

D. Werner. *Funktionalanalysis*. Springer-Lehrbuch, 2011.

Index