Empirical Economics and Econometrics

Jean-Pierre Florens *

2024 Spring

I have a question!

*Notes by FU Zixuan, last compiled on March 1, 2024.

Contents

Recall that our problem is to solve the following equation as an inverse problem

$$\mathcal{A}(F, \phi) = 0 \Rightarrow K_F \phi = r_F$$

where K_F is a linear operator and r_F is a known function. We will solve this equation in a functional space. We will use the following notation

1 Linear equation in functional space

We will introduce the following terminology in the context of functional spaces:

- Spaces
- Linear operator
- Solution of linear equations

1.1 Spaces

Let's define a space (of functions) $\mathcal{E}$ on $\mathbb{R}$ as a linear space if it satisifies the following properties:

- $\forall f, g \in \mathcal{E}$, $f + g \in \mathcal{E}$
- $\forall f \in \mathcal{E}$, $\forall \alpha \in \mathbb{R}$, $\alpha f \in \mathcal{E}$

Now let's define a norm on $\mathcal{E}$ as a function $\|\cdot\| : \mathcal{E} \to \mathbb{R}$ such that

- $\|f\| \geq 0$ and $\|f\| = 0$ if and only if $f = 0$
- $\|\alpha f\| = |\alpha| \, \|f\|$
- $\|f + g\| \leq \|f\| + \|g\|$

Definition 1.1 (complete space)**.** A space $\mathcal{E}$ is called a complete space if every Cauchy sequence in $\mathcal{E}$ converges to a limit in $\mathcal{E}$.

Definition 1.2 (Banach space)**.** A space $\mathcal{E}$ is called a Banach space if it is a complete space with respect to the norm $\|\cdot\|$.

Definition 1.3 (scalar product)**.** A scalar product on $\mathcal{E}$ is a function $\langle \cdot, \cdot \rangle : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ such that

- $\langle f, g\rangle = \langle g, f\rangle$
- $\langle \alpha f, g\rangle = \alpha \langle f, g\rangle$
- $\langle f + g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$
- $\langle f, f\rangle \geq 0$ and $\langle f, f\rangle = 0$ if and only if $f = 0$

If $\mathcal{E}$ is equipped with a scalar product, then it is a Hilbert space.

Definition 1.4 (Hilbert space)**.** A space $\mathcal{E}$ is called a Hilbert space if it is a complete space with respect to the norm $\|\cdot\|$ induced by the scalar product $\langle \cdot, \cdot\rangle$.

The relationship between the norm and the scalar product is given by the following equation:

$$\|f\| = \sqrt{\langle f, f\rangle}$$

Remark. A Banach space B is a complete normed vector space. In terms of generality, it lies somewhere in between a metric space M (that has a metric, but no norm) and a Hilbert space H (that has an inner-product, and hence a norm, that in turn induces a metric). See the summary in Table 1.

spaces	metric	norm	inner product	complete
metric space	✓			
normed space	✓	✓		
inner product space	✓	✓	✓	
Banach space	✓	✓		✓
Hilbert space	✓	✓	✓	✓

Table 1: Summary of linear spaces/vector spaces

Example 1.1. $L^p(\Omega, \mathcal{F}, \mu)$ is a space of functions such that $\int |f|^p < \infty$. It is a Banach space with the norm $\|f\|_p = \left(\int |f|^p\right)^{1/p}$. Also if μ is a probability measure, then we have the inclusion $L^p(\Omega, \mathcal{F}, \mu) \subset L^q(\Omega, \mathcal{F}, \mu)$ for $p \geq q$.

Definition 1.5 (Sobolev space)**.** Let $\Omega \subset \mathbb{R}^d$ be an open set. The Sobolev space $W^{k,p}(\Omega)$ is the space of functions $f : \Omega \to \mathbb{R}$ such that

$$\|f\|_{W^{k,p}} = \left(\sum_{|\alpha| \leq k} \int_\Omega |\partial^\alpha f|^p\right)^{1/p} < \infty$$

where α is a multi-index and $\partial^\alpha f$ is a partial derivative of order $|\alpha|$.

1.1.1 Subspaces

Definition 1.6 (subspace)**.** Let $\mathcal{E}$ be a space and $\mathcal{H}$ be a subspace of $\mathcal{E}$. Then $\mathcal{H}$ is a subspace of $\mathcal{E}$ if it satisfies the following properties:

- $\forall f, g \in \mathcal{H}$, $f + g \in \mathcal{H}$
- $\forall f \in \mathcal{H}$, $\forall \alpha \in \mathbb{R}$, $\alpha f \in \mathcal{H}$

Proposition 1.1. $\mathcal{H}$ is closed if for every sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $f_n \to f$ in $\mathcal{E}$, we have $f \in \mathcal{H}$.

Remark. In a finite dimensional space, every subspace is closed. However, in an infinite dimensional space, a subspace can be closed or not.

Definition 1.7 (Orthogonal subspace)**.** Let $\mathcal{E}$ be a space and $\mathcal{H}$ be a subspace of $\mathcal{E}$. Then $\mathcal{H}^\perp$ is the orthogonal subspace of $\mathcal{H}$ if

$$\mathcal{H}^\perp = \{f \in \mathcal{E} : \langle f, g\rangle = 0, \forall g \in \mathcal{H}\}$$

Remark. The orthogonal subspace of a subspace is always closed.

Definition 1.8 (Dual of a space)**.** Let $\mathcal{E}$ be a Banach space. The dual space of $\mathcal{E}$, denoted by $\mathcal{E}^*$, is the space of all linear functionals on $\mathcal{E}$. A linear functional is a linear map from $\mathcal{E}$ to $\mathbb{R}$.

Definition 1.9 (Riesz representation theorem)**.** Let $\mathcal{E}$ be a Banach space. Then, for every $f(\cdot) \in \mathcal{E}^*$ (f is linear form by definition), there exists a $\psi \in \mathcal{E}$ such that

$$f(\phi) = \langle \phi, \psi\rangle, \quad \forall \phi \in \mathcal{E}$$

Also, $L^p(\Omega, \mathcal{F}, \mu)^* = L^q(\Omega, \mathcal{F}, \mu)$ for $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$.

1.1.2 Basis

Definition 1.10 (Basis of a Hilber space)**.** Let $\mathcal{E}$ be a Hilbert space. A set $\{\phi_j\}_{j\in J}$ is a basis of $\mathcal{E}$ if every $f \in \mathcal{E}$ can be written as

$$f = \sum_{j\in J} c_j \phi_j$$

where $c_j \in \mathbb{R}$ and $\sum_{j\in J} |c_j|^2 < \infty$.

The space is separable if it has a countable basis.

Theorem 1.1. *In a Hilbert separable space, there exists a countable orthonormal basis. That is for every $f \in \mathcal{E}$, we have*

$$f = \sum_{j=1}^{\infty} c_j \phi_j$$

which is called the Fourier series decomposition.

The norm of a function in a Hilbert separable space can be written as

$$\|f\|^2 = \sum_{j=1}^{\infty} |c_j|^2 \quad \text{where} \quad c_j = \langle f, \phi_j \rangle$$

1.1.3 Projection

Projection in probability space

1.2 Linear operator

Definition 1.11 (linear operator)**.** Let $\mathcal{E}$ and $\mathcal{H}$ be two Hilbert spaces (equipped with scalar product). A linear operator $K : \mathcal{E} \to \mathcal{H}$ is a function such that

$$K(\alpha_1\phi_1 + \alpha_2\phi_2) = \alpha_1 K\phi_1 + \alpha_2 K\phi_2$$

for all $\phi_1, \phi_2 \in \mathcal{E}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$.

Example 1.2. Let X, Y, Z be three random variables defined on the $(\Omega, \mathcal{F}, \mathbb{P})$. Let $X = Y \times Z$. We construct three L^2 spaces L^2_X, L^2_Y, L^2_Z. Define a linear operator K such that $K\phi = \mathbb{E}(\phi(y) \mid Z = z)$. Then K is a linear operator from L^2_Y to L^2_Z.

For a linear operator, we define the corresponding subspaces defined as domain, range and kernel as follows:

Definition 1.12 (domain, range and kernel)**.** Let $\mathcal{E}$ and $\mathcal{H}$ be two Hilbert spaces and $K : \mathcal{E} \to \mathcal{H}$ be a linear operator. Then

- The domain of K is defined as

$$\mathcal{D}(K) = \{\phi \in \mathcal{E} : K\phi \in \mathcal{H}\}$$

- The range of K is defined as

$$\mathcal{R}(K) = \{K\phi : \phi \in \mathcal{M}(K)\}$$

- The kernel of K is defined as

$$\mathcal{N}(K) = \{\phi \in \mathcal{M}(K) : K\phi = 0\}$$

For completeness, we define injection and surjection as follows:

Definition 1.13 (injection and surjection)**.** Let $\mathcal{E}$ and $\mathcal{H}$ be two Hilbert spaces and $K : \mathcal{E} \to \mathcal{H}$ be a linear operator. Then

- K is called an injection if for all $\phi_1, \phi_2 \in \mathcal{E}$, $K\phi_1 = K\phi_2$ implies $\phi_1 = \phi_2$.
- K is called a surjection if for all $h \in \mathcal{H}$, there exists $\phi \in \mathcal{E}$ such that $K\phi = h$.

1.2.1 Boundness and continuity

Definition 1.14 (boundness)**.** Let $\mathcal{E}$ and $\mathcal{H}$ be two Hilbert spaces and $K : \mathcal{E} \to \mathcal{H}$ be a linear operator. Then K is called bounded if there exists a constant $C > 0$ such that

$$\|K\phi\|_{\mathcal{H}} \leq C \|\phi\|_{\mathcal{E}}$$

for all $\phi \in \mathcal{E}$.

Definition 1.15 (continuity)**.** Let $\mathcal{E}$ and $\mathcal{H}$ be two Hilbert spaces and $K : \mathcal{E} \to \mathcal{H}$ be a linear operator. Then K is called continuous if for all $\phi_n \to \phi$ in $\mathcal{E}$, we have $K\phi_n \to K\phi$ in $\mathcal{H}$.

We will look at an example where k is not continuous.

Example 1.3. Let $\mathcal{E} = \mathcal{C}^0_{[0,1]}$ be the space of continuous functions on $[0, 1]$. Define a linear operator $K : \mathcal{E} \to \mathbb{R}$ such that $K\phi = \phi(x_0)$.

1.2.2 Adjoint operator

Similar to the transpose of a matrix, we define the adjoint operator of a linear operator.

Definition 1.16 (adjoint operator)**.** Let $\mathcal{E}$ and $\mathcal{H}$ be two Hilbert spaces and $K : \mathcal{E} \to \mathcal{H}$ be a linear operator. Then the adjoint operator $K^* : \mathcal{H} \to \mathcal{E}$ is defined as

$$\langle K\phi, h\rangle_{\mathcal{H}} = \langle \phi, K^* h\rangle_{\mathcal{E}}$$

for all $\phi \in \mathcal{E}$ and $h \in \mathcal{H}$.

Remark. We can show that matrix transpose is a special case of adjoint operator. For example, let A be a matrix of dimension $m \times n$ and x, y be vectors of dimension n, m respectively. Then we have $\langle Ax, y \rangle = (Ax)^\top y = x^\top A^\top y = \langle x, A^* y \rangle$.

Example 1.4 (integral operator as a self-adjoint operator)**.** Let $\mathcal{E} = \mathcal{C}^0([0,1]) = \mathcal{F}$. Define a linear operator $K : \mathcal{E} \to \mathcal{F}$ such that $K\phi(x) = \int_0^1 \phi(x) k(x,y) dx = \psi(y)$ where $k(x,y)$ is a given function. Then for any $\psi \in \mathcal{F}$, we have

$$\begin{aligned} \langle K\phi, \psi \rangle_{\mathcal{F}} &= \int_0^1 \int_0^1 \phi(x) k(x,y) dx \psi(y) dy \\ &= \int_0^1 \int_0^1 k(x,y) \phi(x) \psi(y) dx dy \\ &= \langle \phi, K^* \psi \rangle_{\mathcal{E}} \end{aligned}$$

Then we have $K^*\psi(y) = \int_0^1 k(x,y) \psi(y) dy$. We say K is **self-adjoint** if $k(x,y) = k(y,x)$, that is $K^* = K$.

We look at another example in the context of probability.

Example 1.5. Let $X = Y \times Z$, where X, Y, Z are random variables. We define a linear operator K from $L_Y^2 \to L_Z^2$ such that $K\phi = \mathbb{E}(\phi(Y) \mid Z = z)$. Then the adjoint operator K^* is given by $K^* h(y) = \int h(z) f(z \mid y) dz$. Because by definition,

$$\begin{aligned} \langle K\phi, \psi \rangle &= \langle \mathbb{E}(\phi(Y) | Z), \psi(Z) \rangle \\ &= \mathbb{E}[\mathbb{E}(\phi(Y) | Z) \psi(Z)] \\ &= \mathbb{E}[\phi(Y) \psi(Z)] \quad \text{by independence?} \\ &= \mathbb{E}[\phi(Y) \mathbb{E}(\psi(Z) | Y)] \\ &= \langle \phi, \mathbb{E}(\psi(Z) | Y) \rangle \end{aligned}$$

1.2.3 Compact operators and SVD

Definition 1.17 (Singular value decompostion)**.** Let $\mathcal{E}$ and $\mathcal{F}$ be two Hilbert spaces and $K : \mathcal{E} \to \mathcal{F}$ be a linear compact operator. The adjoint operator K^* is also compact. Then there exists a set of

1. A set of singular values $\lambda_j \geq 0$
2. Two orthonormal basis $\phi_j \in \mathcal{E}$ and $\psi_j \in \mathcal{F}$ such that $\forall \phi \in \mathcal{E}$, we have $\phi = \sum_j \lambda_j \langle \phi, \phi_j \rangle \phi_j + \phi_0$ for $\phi_0 \in \mathcal{N}(K)$. Similarly for $\psi \in \mathcal{F}$.

The following properties hold from SVD:

- $K^*K\phi_j = \lambda_j^2\phi_j$ and $KK^*\psi_j = \lambda_j^2\psi_j$
- $K\phi_i = \lambda_j\psi_j$ and $K^*\psi_j = \lambda_j\phi_i$
- Main implication:

$$\begin{aligned} K\phi &= \sum_j \langle \phi, \phi_j \rangle k\phi_j + K\phi_0 \\ &= \sum_j \langle \phi, \phi_j \rangle \lambda_j\psi_j + 0 \\ &= \sum_j \lambda_j \langle \phi, \phi_j \rangle \psi_j \end{aligned}$$

which leads to the following decomposition:

Definition 1.18 (Hilbert Schimidt and Nuclear operator)**.** A compact operator K is called a Hilbert Schimidt operator if the sum of the squares of the singular values is finite, that is $\sum_j \lambda_j^2 < \infty$. It is a nuclear operator if the sum of the singular values is finite, that is $\sum_j \lambda_j < \infty$.

A direct consequence is that

$$K^\beta\phi = \sum_j \lambda_j^\beta \langle \phi, \phi_j \rangle \psi_j$$

Remark. We have the following inclusion property of the operator

Nuclear $\subset$ Hilbert Schimidt $\subset$ Compact $\subset$ Continuous and Bounded $\subset$ Linear

Theorem 1.2. *Let K be a compact operator from the Hilbert space $\mathcal{E}$ to $\mathcal{F}$. Denote any orthonormal basis of $\mathcal{E}$ by $\{\varepsilon_j\}_{j\in J}$. Then the following properties hold:*

- $\sum \lambda_j^2 = \sum_j \langle K\varepsilon_j, K\varepsilon_j \rangle = \sum_j \langle \varepsilon_j, K^*K\varepsilon_j \rangle$
- $\sum \lambda_j = \sum_j \langle K\varepsilon_j, \varepsilon_j \rangle$

The theorem is obviously true if $\{\varepsilon_j\}$ is the basis from the SVD of K.

Example 1.6. For example, take $K\varphi := \int K(s,t)\varphi(s)ds$, the HS class is given by

$$\begin{aligned}\sum_{j=1}\langle K\varepsilon_j, K\varepsilon_j\rangle &= \sum_j \int K\varepsilon_j(x)K\varepsilon_j(x)dx \\ &= \sum_j \int \left(\int K(y,x)\varepsilon_j(y)dy\right)^2 dx \\ &\overbrace{=}^{?} \sum_j \int\int K(y,x)^2 dydx < \infty\end{aligned}$$

Example 1.7. Let us take

$$\begin{aligned}K\varphi &= \mathbb{E}\left[\varphi(Y)|Z\right] \\ &= \int \varphi(y)f(y|z)dy \\ &= \int \varphi(y)\frac{f(y,z)}{f(z)} \in L_Z^2\end{aligned}$$

K is an operator from $L_Y^2 \to L_Z^2$ where Y and Z are random variables with joint density function $f(y,z)$. Then, by definition the adjoint operator K^* is given by $K^*\psi(y) = \int \psi(z)f(z|y)dz$ The Hilbert-Schmidt class is given by

$$\begin{aligned}\int\left(\int \varphi_i(y)\frac{f(y,z)}{f(z)}dy\right)^2 dz &= \int\int \varphi_i(y)\frac{f(y,z)}{f(z)}\varphi_i(y)\frac{f(y,z)}{f(y)}dydz \\ &= \int\int \frac{f(y,z)}{f(z)}\frac{f(y,z)}{f(y)}dydz\end{aligned}$$

On a side note,

$$\frac{f(y,z)}{f(y)f(z)} = \sum_{j=1}^{\infty}\lambda_j\varphi_j(y)\psi_j(z)$$

(why?) and $\sum_{j=1}\lambda_j^2$ is a measure of the dependence between Y and Z.
Convergence in Hilbert-Schmidt class (?)

1.3 Linear Equations

For a linear equation $K\varphi = r$, we define two types of linear operators

1. Type 1: K is compact
2. Type 2: $K = I - T$ where T is compact

(Why two types?)

1.3.1 Genralized Inverse

The problem

$$\min_{\varphi \in \mathcal{E}} \|K\varphi - r\|^2$$

has a solution φ_0 if and only if $r \in \mathcal{R}(K) + \mathcal{R}K(r)^{\perp}$. Question: Is the column and null space (kernel) perpendicular to each other? In finite dimension, $\mathcal{F} = \mathcal{R}(K) + \mathcal{R}K(r)^{\perp}$ while infinite dimension, $\mathcal{F} = \overline{\mathcal{R}(K)} + \mathcal{R}K(r)^{\perp}$. If φ_0 exists, then $\varphi_0 + \varphi$ for any $\varphi \in \mathcal{N}(K)$ is also a solution. We want to find the minimum norm generalized inverse. We perform a SVD of the operator K and get $\lambda_i, \{\varphi_i\}, \{\psi_i\}$. Thus, r can be expressed in terms of the orthonormal basis defined by $\{\psi_i\}$ as

$$r = \sum_i \lambda_i \langle r, \psi_i \rangle \psi_i.$$

Similarly for $\varphi = \sum_i \lambda_i \langle \varphi, \varphi_i \rangle \varphi_i$. The linear equation becomes

$$K\varphi = \sum_i \lambda_i \langle \varphi, \varphi_i \rangle K\varphi_i = \sum_i \lambda_i \langle \varphi, \varphi_i \rangle \psi_i = \sum_i \lambda_i \langle r, \psi_i \rangle \psi_i = r$$

which implies that

$$\lambda_i \langle \varphi, \varphi_i \rangle = \langle r, \psi_i \rangle$$
$$\langle \varphi, \varphi_i \rangle = \frac{\langle r, \psi_i \rangle}{\lambda_i}$$

for $\lambda_i \neq 0$.

Definition 1.19 (Generalized inverse)**.** If $\sum_i \frac{\langle r, \psi_i \rangle}{\lambda_i}^2 < \infty$, then the solution to the linear equation is given by

$$\varphi^+ = \sum_i \frac{\langle r, \psi_i \rangle}{\lambda_i} \varphi_i$$

Remark. We can see that φ^+ is not continuous w.r.t. r as λ_i can go to zero.

1.3.2 Regularized solution

When λ_i goes to zero, we have a problem. We can regularize the solution by add a regularization term.

Definition 1.20 (Tikhonov regularization)**.** The new problem is

$$\min \|K\varphi - r\|^2 + \alpha^2 \|\varphi\|^2, \alpha > 0$$

The regularized solution is given by

$$\varphi_\alpha = \sum_i \frac{\lambda_i}{\lambda_i^2 + \alpha^2} \langle r, \psi_i \rangle \varphi_i$$

Proof. We make use of gâteau derivative. $\square$

Remark. This is similar to what we do in OLS regression where $\hat{\beta} = (X^\top X)^{-1} X^\top y$ yet $X^\top X$ is not invertible. We deal with the issue using **Ridge regression** by $\hat{\beta} = (X^\top X + \alpha I)^{-1} X^\top y$.

We now prove the existence of the solution:

1. $\alpha I + K^*K$ is a compact operator from $\mathcal{E}$ to $\mathcal{E}$ and it satisfies $(\alpha I + K^*K)\varphi_j = (\alpha + \lambda_j^2)\varphi_j$.
2. $K^*r = \sum_i \lambda_i \langle r, \psi_i \rangle \varphi_i$.

As a result,

$$\varphi_\alpha = (\alpha I + K^*K)^{-1} K^* r = \sum_i \frac{\lambda_i}{\lambda_i^2 + \alpha^2} \langle r, \psi_i \rangle \varphi_i \tag{1}$$

The behavior of φ_α as $\alpha \to 0$

1. If K is injective and $r \in \mathcal{R}(K)$ then $\|\varphi_\alpha - \varphi_0\| \to 0$. Define bias as $\varphi_\alpha - \varphi$, we have

$$\begin{aligned} b_\alpha &= (\alpha I + K^*K)^{-1} K^*Kr - \varphi \\ &= (\alpha I + K^*K)^{-1}(K^*K - \alpha I + K^*K)\varphi \\ &= -\alpha(\alpha I + K^*K)^{-1}\varphi \end{aligned}$$

   Therefore as in the derivation of (1), $\|b_\alpha\| = \alpha^2 \sum \frac{1}{\lambda_i^2 + \alpha^2} \langle \varphi, \varphi_i \rangle^2 \to 0$

2. If K is not injective, then $\varphi = \varphi_1 + \varphi_0$ where $\varphi_1 \in \mathcal{N}(K)$ then $\|\varphi_\alpha - \varphi_1\| \to 0$, not the true φ.

1.3.3 Sauce condition

We introduce a sauce condition to control the rate of convergence of φ_α. The condition that $\varphi \in \mathcal{R}(K^*K)^{\frac{\beta}{2}}$ is equivalent to there exists a $\delta \in (K^*K)^{\frac{\beta}{2}}\mathcal{E}$ such that $\varphi = K^*K\delta$. If φ satisfies the sauce condition, then by SVD of K we have

$$\varphi = \sum_i \lambda_i^\beta \langle \delta, \varphi_i \rangle \varphi_i$$

(why?) In general we have $\varphi = \sum_i \langle \varphi, \varphi_i \rangle \varphi_i$. Thus,

$$\begin{aligned} \langle \varphi, \varphi_i \rangle &= \lambda_i^\beta \langle \delta, \varphi_i \rangle \\ \Rightarrow \sum_i \frac{\langle \varphi, \varphi_i \rangle^2}{\lambda_j^{2\beta}} &= \langle \delta, \varphi_j \rangle^2 < \infty \end{aligned}$$

Under this condition, we can show that $\|b_\alpha\|^2 = O(\alpha^\beta)$ for $\beta < 2$, that is $O(\alpha^{\beta \wedge 2})$.

Proof. We have

$$\begin{aligned} \|b_\alpha\|^2 &= \alpha^2 \sum_i \frac{\lambda_i^{2\beta} \langle \varphi, \varphi_i \rangle^2}{\lambda_i^{2\beta}(\lambda_i^2 + \alpha)^2} \quad \text{let } \lambda_i^2 = \alpha \\ &\leq \sum_i \frac{\alpha^\beta \langle \varphi, \varphi_i \rangle^2}{\lambda_i^{2\beta} 2} \quad \text{let } \lambda_i^2 = \alpha \\ &= \frac{\alpha^\beta}{2} \sum \frac{\langle \varphi, \varphi_i \rangle^2}{\lambda_i^{2\beta}} \\ &= O(\alpha^\beta) \end{aligned}$$

□

If $\beta = 2$, then $\varphi \in \mathcal{R}(K^*K) \Leftrightarrow \varphi = K^*K\delta \Leftrightarrow \varphi \in \mathcal{D}(K^*K)^{-1}$. $(K^*K)^{-1}$ is a differentiation operator. φ is differentiable w.r.t. $(K^*K)^{-1}$.

Example 1.8. Let $K\varphi(t) = \int_s^t \varphi(s)ds$ and $K^*\psi(s) = \int_s^1 \psi(s)ds$. Then $K^*K\varphi(s) = \int_s^1 \int_0^t \varphi(u)dudt = \delta(s)$ with $\delta(1) = 0$. Also, $-\int_0^s \varphi(u)du = \delta'(s)$ with $\delta'(0) = 0$. We get

$$\varphi = -\delta'', \in \mathcal{R}(K^*K)$$

What is the purpose of showing all these?

Example 1.9. On a Hilbert space $\mathcal{E}$ we define a differential operator L. (For $a > 0$, L^{-a} is an integral operator.)

- For $K : \mathcal{E} \to \mathcal{F}$, we have $K \sim L^{-a} \subseteq \|L^{-a}\phi\| \subseteq \|K\varphi\| \subseteq \tau \|L^{-a}\varphi\|$. The degree of illposeness of K w.r.t. L is a.
- For φ, $L^b\varphi$ is defined for $b > 0$. φ is differentiable w.r.t. L. b is the order of smoothness of φ

What are we doing here?

1.3.4 Algorithm

First we introduce spectral cutoff. Given the equation $r = K\varphi$. By decomposition, $\varphi = \sum_j \frac{1}{\lambda_j} \langle r, \psi_j \rangle \varphi_j$ and $\varphi_\varepsilon = \sum_{j/\lambda_j \geq \varepsilon} \frac{1}{\lambda_j} \langle r, \psi_j \rangle \varphi_j$. Therefore,

$$\varphi_\varepsilon - \varphi = \sum_{j/\lambda_j < \varepsilon} \frac{1}{\lambda_j} \langle r, \psi_j \rangle \varphi_j - \sum_j \frac{1}{\lambda_j} \langle r, \psi_j \rangle \varphi_j = \sum_{j/\lambda_j < \varepsilon} \langle \varphi, \varphi_j \rangle \varphi_j$$

If $\varphi, \psi \in \mathcal{R}(K^*K)^{\frac{\beta}{2}}$, then $\|\varphi_\varepsilon - \varphi\|^2 = O(\alpha^\beta)$.

Proof.

$$\begin{aligned}
\|\varphi_\varepsilon - \varphi\|^2 &= \sum_{j/\lambda_j < \varepsilon} \frac{\langle r, \psi_j \rangle^2}{\lambda_j^{2\beta} \lambda_j^{2\beta}} \\
&\leq \varepsilon^{2\beta} \sum_j \frac{\langle \varphi, \varphi_i \rangle^2}{\lambda_i^{2\beta}} \\
&= O(\varepsilon^{2\beta})
\end{aligned}$$

□

Now we introduce **Tikhmov Iterated Algorithm**.

1. Choose α_0.
2. $\varphi_{\alpha_1} = \arg\min \|K\varphi - r\|^2 + \alpha_0 \|\varphi - \varphi_{\alpha_0}\|$. The solution is $\varphi_{\alpha_1} = (K^*K + \alpha_0 I)^{-1} K^* r + \alpha_0 (K^*K + \alpha_0 I)^{-1} \varphi_{\alpha_0}$.
3. Replace φ_{α_0} by φ_{α_1} and α_0 by α_1 in the next iteration.

Next, we introduce **Landweber Regularization Algorithm**. Save for later...read the book! If $\varphi \in \mathcal{R}(K^*K)^{\frac{\beta}{2}}$, the regularization converges in order of $\|\varphi_k - \varphi\|^2 = O(\frac{1}{k^\beta})$.

2 Statistical analysis of linear equation

In this section, we will first introduce general results then applications such as functional regression and instrumental variable approach.

Set up The model is $r = K\varphi$. We don't know K, r so we need to estimate them. Then solve for φ.

2.1 Rate of convergence

2.1.1 When K is given

We can write the estimate of $\hat{r} = K\varphi + u = r + u$.Then, $K^{-1}\hat{r} - K^{-1}r = \hat{\varphi} - \varphi$. The non continuity of K^{-1} causes problem in the convergence of $\hat{\varphi}$. Therefore, we need regularization. Recall the solution in the case of Tikhmov regularization, we have

$$\hat{\varphi_\alpha} = (\alpha I + K^*K)^{-1}K^*\hat{r}$$

Therefore,

$$\begin{aligned}
\hat{\varphi_\alpha} - \varphi &= (\alpha I + K^*K)^{-1}K^*\hat{r} - \varphi \\
&= \underbrace{(\alpha I + K^*K)^{-1}K^*(\hat{r} - r)}_{\text{variance term}} + \underbrace{(\alpha I + K^*K)^{-1}K^*r - \varphi}_{\text{bias term}}
\end{aligned}$$

2.1.2 When K is estimated

2.2 Applications

2.2.1 Regression

For $y_i \in \mathbb{R}$ and $y_i = \langle z_i, \beta\rangle + u_i$