Chapter 2

Hilbert Spaces

This chapter collects some basic Hilbert space theory related to the analysis of time series. Section 2.1 defines inner-product spaces and Hilbert spaces, gives examples and states some basic properties. Section 2.2 is on the projection theorem in Hilbert spaces and its applications in statistics, most notably we derive the *predition equations*.

2.1 Preliminaries

We define an inner-product and an inner-product space first. Denote by $\mathbb{R}$ and $\mathbb{C}$ the set of real and complex numbers, respectively. For a complex number z we denote by $\bar{z}$ its conjugate complex.

Definition 2.1.1 (Inner-Product Space) *A real or complex vector space $\mathcal{H}$ is called inner-product space if there is a mapping $\langle\cdot,\cdot\rangle\colon \mathcal{H}\times\mathcal{H}\to\mathbb{C}$ such that*

(a) $\langle x,y\rangle = \overline{\langle y,x\rangle}$ *for all* $x,y\in\mathcal{H}$;

(b) $\langle x+y,z\rangle = \langle x,z\rangle + \langle y,z\rangle$ *for all* $x,y,z\in\mathcal{H}$;

(c) $\langle \alpha x,y\rangle = \alpha\langle x,y\rangle$ *for all* $x,y\in\mathcal{H}$ *and* $\alpha\in\mathbb{C}$;

(d) $\langle x,x\rangle \geq 0$ *and* $\langle x,x\rangle = 0$ *if and only if* $x=0$.

Example 2.1.1 Let $\boldsymbol{x}=(x_1,\ldots,x_k)'$ and $\boldsymbol{y}=(y_1,\ldots,y_k)'$ be elements of the real vector space $\mathbb{R}^k$. Then, we can define the usual inner product as

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle = \sum_{j=1}^{k} x_j y_j.$$

If $\boldsymbol{z}=(z_1,\ldots,z_k)'$ and $\boldsymbol{w}=(w_1,\ldots,w_k)'$ are elements of the complex vector space $\mathbb{C}^k$, then we can define in a similar way their inner product by

$$\langle \boldsymbol{z},\boldsymbol{w}\rangle = \sum_{j=1}^{k} z_j \bar{w}_j,$$

In both cases, parts (a)–(d) are verified directly.

Each inner product $\langle\cdot,\cdot\rangle$ induces a norm $\|\cdot\|$ by letting $\|x\| = \sqrt{\langle x, x\rangle}$ for $x \in \mathcal{H}$. The norm may be interpreted as the length of x. In particular, there is the following relationship between inner products and norms (Cauchy-Schwarz inequality):

$$|\langle x, y\rangle| \leq \|x\|\|y\| \qquad \text{for all } x, y \in \mathcal{H},$$

with equality if and only if $x = y\langle x, y\rangle/\langle y, y\rangle$. The preceding allows to define the angle

$$\theta = \cos^{-1}\left(\frac{\langle x, y\rangle}{\|x\|\|y\|}\right)$$

between two general Hilbert space elements x and y (Geometric interpretation in $\mathbb{R}^k$!) and furthermore a notion of orthogonality. We say that $x, y \in \mathcal{H}$ are orthogonal if $\langle x, y\rangle = 0$ which is, in the case $x, y \neq 0$, equivalent to $\theta = \frac{\pi}{2}$.

Remark 2.1.1 Further properties of the norm include

(a) the triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathcal{H}$;

(b) $\|\alpha x\| = |\alpha|\|x\|$ for all $x \in \mathcal{H}$ and $\alpha \in \mathbb{C}$;

(c) $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$;

(d) the parallelogram law: $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$ for all $x, y \in \mathcal{H}$.

The above properties follow directly from the definition of the inner product. Note that the parallelogram law does not hold in general but is a consequence of our norms being defined from an inner product.

Definition 2.1.2 (Convergence in Norm) *A sequence $(x_n\colon n \in \mathbb{N})$ in $\mathcal{H}$ is said to converge in norm to $x \in \mathcal{H}$ if $\|x_n - x\| \to 0$ as $n \to \infty$.*

The following theorem shows that both inner product and norm are continuous mappings.

Theorem 2.1.1 *If $(x_n\colon n \in \mathbb{N})$ and $(y_n\colon n \in \mathbb{N})$ are sequences in an inner-product space $\mathcal{H}$ satisfying $\|x_n - x\| \to 0$ and $\|y_n - y\| \to 0$ for some $x, y \in \mathcal{H}$, then*

(a) $\|x_n\| \to \|x\|$ as $n \to \infty$;

(b) $\langle x_n, y_n\rangle \to \langle x, y\rangle$ as $n \to \infty$.

Proof. (a) Note that the inequalities $\|x\| = \|x - x_n + x_n\| \leq \|x - x_n\| + \|x_n\|$ and $\|x_n\| = \|x_n - x + x\| \leq \|x - x_n\| + \|x\|$ imply that

$$\big|\|x_n\| - \|x\|\big| \leq \|x_n - x\| \to 0 \qquad (n \to \infty),$$

which proves the assertion.

(b) Applying triangle and Cauchy-Schwarz inequalities, it follows that

$$\begin{aligned}\big|\langle x_n, y_n\rangle - \langle x, y\rangle\big| &= \big|\langle x_n, y_n - y\rangle + \langle x_n - x, y\rangle\big| \\ &\leq \big|\langle x_n, y_n - y\rangle\big| + \big|\langle x_n - x, y\rangle\big| \\ &\leq \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\|,\end{aligned}$$

where the right-hand side converges to zero because $\|x_n\|$ is bounded since $\|x_n\| \to \|x\|$ by (a). This completes the proof. $\square$

Definition 2.1.3 (Cauchy Sequence) *A sequence* $(x_n : n \in \mathbb{N})$ *in* $\mathcal{H}$ *is said to be Cauchy if* $\|x_m - x_n\| \to 0$ *as* $m, n \to \infty$.

Definition 2.1.4 (Hilbert Space) *An inner-product space* $\mathcal{H}$ *is a Hilbert space if each Cauchy sequence in* $\mathcal{H}$ *converges to a limit in* $\mathcal{H}$. *We say that* $\mathcal{H}$ *is complete.*

Example 2.1.2 The real vector space $\mathbb{R}^k$ equipped with the inner product defined in Example 2.1.1 is a Hilbert space. To see this, let $(\boldsymbol{x}_n : n \in \mathbb{N})$ be a Cauchy sequence in $\mathbb{R}^k$. Then,

$$\|\boldsymbol{x}_m - \boldsymbol{x}_n\|^2 = \sum_{j=1}^{k} |x_{mj} - x_{nj}|^2 \to 0 \qquad \text{as } m, n \to \infty.$$

For any j, the coordinates $(x_{nj} : n \in \mathbb{N})$ can be viewed as Cauchy sequences in $\mathbb{R}$, which is complete. Thus we can find $x_j \in \mathbb{R}$ such that $x_{nj} \to x_j$ as $n \to \infty$. This provides us with the limit candidate $\boldsymbol{x} = (x_1, \ldots, x_k)'$ for which it is indeed easily shown that $\|\boldsymbol{x}_n - \boldsymbol{x}\| \to 0$. Hence $\mathbb{R}^k$ is complete. Similar arguments apply to $\mathbb{C}^k$.

Example 2.1.3 Let $(\Omega, \mathcal{A}, P)$ be a probability space and set

$$\mathcal{L}^2 = \left\{ X : \Omega \to \mathbb{R} \text{ such that } E[X^2] = \int_\Omega X^2 dP < \infty \right\}.$$

Then $\mathcal{L}^2$ is a vector space over $\mathbb{R}$, since for $X, Y \in \mathcal{L}^2$ and $a \in \mathbb{R}$ we have that $E[(X + Y)^2] \leq 2E[X^2] + 2E[Y^2] < \infty$ and $E[(aX)^2] = a^2 E[X^2] < \infty$. Further, $0 \in \mathcal{L}^2$. Define $\langle X, Y \rangle = E[XY]$ for $X, Y \in \mathcal{L}^2$. This function satisfies all conditions of Definition 2.1.1 except (d) because $\langle X, X \rangle = 0$ implies only that $P(X = 0) = 1$ but not that $X = 0$. The way out is to define an equivalence relation on $\mathcal{L}^2$ which identifies two variables X and Y with each other if $P(X = Y) = 1$. The resulting equivalence classes define the space $L^2 = L^2(\Omega, \mathcal{A}, P)$ (as the quotient space of $\mathcal{L}^2$ with respect to the introduced equivalence relation). It can be shown that L^2 is complete and thus a Hilbert space.

Example 2.1.4 The complex counterpart of the L^2 space of the previous example is defined analogously using $\langle X, Y \rangle = E[X\overline{Y}]$. More to the point, if μ is an arbitrary finite non-zero measure on $(\Omega, \mathcal{A})$, then we can define the class

$$\mathcal{D} = \left\{ f : \Omega \to \mathbb{C} \text{ such that } \int_\Omega |f|^2 d\mu < \infty \right\}$$

Defining $\langle f, g \rangle = \int_\Omega f\overline{g} d\mu$ and identifying f and g with each other if $\int_\Omega |f - g|^2 d\mu = 0$ gives us (as quotient space) the Hilbert space $L^2(\Omega, \mathcal{A}, \mu)$.

2.2 Projections

Example 2.2.1 Let $\boldsymbol{y} = (\frac{1}{4}, \frac{1}{4}, 1)'$, $\boldsymbol{x}_1 = (1, 0, \frac{1}{4})'$ and $\boldsymbol{x}_3 = (0, 1, \frac{1}{4})'$ be three vectors in $\mathbb{R}^3$. Suppose that, given $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, you are trying to determine the best linear approximation to $\boldsymbol{y}$, that is you are looking for $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\hat{\boldsymbol{y}} = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2$ is closest to

$\boldsymbol{y}$. Since $\mathbb{R}^3$ is naturally equipped with the Euclidean norm induced by the inner product of Example 2.1.1, this problem is readily translated into minimizing the objective function

$$S = \|\boldsymbol{y} - \alpha_1 \boldsymbol{x}_1 - \alpha_2 \boldsymbol{x}_2\|^2 = (\tfrac{1}{4} - \alpha_1)^2 + (\tfrac{1}{4} - \alpha_2)^2 + (1 - \tfrac{1}{4}\alpha_1 - \tfrac{1}{4}\alpha_2)^2$$

with respect to α_1 and α_2. This can be achieved using calculus or the following geometric approach, noting that $\hat{\boldsymbol{y}}$ satisfies that $\boldsymbol{y} - \hat{\boldsymbol{y}} \perp \boldsymbol{x}_1, \boldsymbol{x}_2$. The latter means that

$$\langle \boldsymbol{y} - \alpha_1 \boldsymbol{x}_1 - \alpha_2 \boldsymbol{x}_2, \boldsymbol{x}_j \rangle = 0, \qquad j = 1, 2,$$

or, more explicitly,

$$\begin{aligned}
\alpha_1 \langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle + \alpha_2 \langle \boldsymbol{x}_2, \boldsymbol{x}_1 \rangle &= \langle \boldsymbol{y}, \boldsymbol{x}_1 \rangle, \\
\alpha_1 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle + \alpha_2 \langle \boldsymbol{x}_2, \boldsymbol{x}_2 \rangle &= \langle \boldsymbol{y}, \boldsymbol{x}_2 \rangle,
\end{aligned}$$

which is a linear equation system. For the current case, this is easily solved for $\alpha_1 = \alpha_2 = \frac{4}{9}$ and thus $\hat{\boldsymbol{y}} = (\frac{4}{9}, \frac{4}{9}, \frac{2}{9})'$.

Example 2.2.2 Suppose we have observed $X_1, X_2 \in L^2$ and would like to use the observed values to linearly predict $Y \in L^2$ via $\hat{Y} = \alpha_1 X_1 + \alpha_2 X_2$. Mimicking the steps taken in the previous example leads to the problem of minimizing the objective function

$$S = \|Y - \alpha_1 X_1 - \alpha_2 X_2\|^2 = E\big[|Y - \alpha_1 X_1 - \alpha_2 X_2|^2\big]$$

with respect to $\alpha_1, \alpha_2 \in \mathbb{R}$. Note that X_1 and X_2 span the space

$$\mathcal{M} = \big\{X \in L^2 \colon X = a_1 X_1 + a_2 X_2,\ a_1, a_2 \in \mathbb{R}\big\},$$

so that the task becomes finding an $\hat{Y} \in \mathcal{M}$ such that $\|Y - \hat{Y}\|^2$ is minimal. The analogy to Example 2.2.1 suggests that the solution satisfies the orthogonality relation $Y - \hat{Y} \perp \mathcal{M}$, that is, $\langle Y - \alpha_1 X_1 - \alpha_2 X_2, X \rangle = 0$ for all $X \in \mathcal{M}$. By linearity of $\mathcal{M}$ this reduces to

$$\langle Y - \alpha_1 X_1 - \alpha_2 X_2, X_j \rangle = 0, \qquad j = 1, 2,$$

and thus leads to the linear equation system

$$\begin{aligned}
\alpha_1 E[X_1^2] + \alpha_2 E[X_1 X_2] &= E[Y X_1], \\
\alpha_1 E[X_1 X_2] + \alpha_2 E[X_2^2] &= E[Y X_2],
\end{aligned}$$

which is structurally the same as in Example 2.2.1.

In the following, we will show that the intuitive approach developed is indeed valid. To do so, we introduce some more notation.

Definition 2.2.1 *(a) A linear subspace of a Hilbert space $\mathcal{H}$ is closed if every converging sequence in $\mathcal{M}$ has its limit x in $\mathcal{M}$.*

(b) For any subset $\mathcal{M}$ of a Hilbert space $\mathcal{H}$, we say that

$$\mathcal{M}^\perp = \big\{x \in \mathcal{H} \colon \langle x, y \rangle = 0 \textit{ for all } y \in \mathcal{M}\big\}$$

is its orthogonal complement.

Lemma 2.2.1 *If $\mathcal{M}$ is a subset of the Hilbert space $\mathcal{H}$, then its orthogonal complement $\mathcal{M}^\perp$ is a closed subspace of $\mathcal{H}$.*

Proof. Clearly $0 \in \mathcal{M}^\perp$. Also, if $x_1, x_2 \in \mathcal{M}^\perp$, then so is $\alpha_1 x_1 + \alpha_2 x_2$ because

$$\langle \alpha_1 x_1 + \alpha_2 x_2, y\rangle = \alpha_1\langle x_1, y\rangle + \alpha_2\langle x_2, y\rangle = 0$$

for all $y \in \mathcal{M}$. Consequently $\mathcal{M}^\perp$ is a subspace. Let $(x_n\colon n \in \mathbb{N})$ be a sequence in $\mathcal{M}^\perp$ that converges to an element $x \in \mathcal{H}$. The continuity of the inner product (Theorem 2.1.1(b)) implies that then $\langle x, y\rangle = 0$ for all $y \in \mathcal{M}$ and thus $x \in \mathcal{M}^\perp$. This proves the lemma. □

Theorem 2.2.1 *Let $\mathcal{M}$ be a closed subspace of a Hilbert space $\mathcal{H}$ and let $x \in \mathcal{H}$. Then,*
(a) There is a unique $\hat{x} \in \mathcal{M}$ such that

$$\|x - \hat{x}\| = \inf_{y\in\mathcal{M}} \|x - y\| = L;$$

(b) $\hat{x} \in \mathcal{M}$ and $\|x - \hat{x}\| = L$ if and only if $\hat{x} \in \mathcal{M}$ and $x - \hat{x} \in \mathcal{M}^\perp$.
We call $\hat{x}$ the (orthogonal) projection of x onto $\mathcal{M}$.

Proof. (a) Let $K = \inf_{y\in\mathcal{M}} \|x - y\|^2$. By definition of the infimum, there is a sequence $(y_n\colon n \in \mathbb{N})$ in $\mathcal{M}$ such that $\|y_n - x\| \to K$ as $n \to \infty$. Using the parallelogram law, we obtain

$$\begin{aligned}
0 &\le \|y_m - y_n\|^2 \\
&= \|(y_m - x) - (y_n - x)\|^2 \\
&= -\|y_m + y_n - 2x\|^2 + 2\|y_m - x\|^2 + 2\|y_n - x\|^2 \\
&= -4\left\|\tfrac{1}{2}(y_m + y_n) - x\right\|^2 + 2\|y_m - x\|^2 + 2\|y_n - x\|^2.
\end{aligned}$$

The second and third term on the right-hand side each converge to $2K$ by construction. Noting that $\frac{1}{2}(y_m+y_n) \in \mathcal{M}$, it follows that the first term is bounded from above by $-4K$, so that the right-hand side is non-positive in the limit. This shows that $(y_n\colon n \in \mathbb{N})$ is a Cauchy sequence and thus convergent. Since $\mathcal{M}$ is assumed closed, there is an element $\hat{x} \in \mathcal{M}$ such that $\|y_n - \hat{x}\| \to 0$. The continuity of the norm implies further that $\|x - \hat{x}\|^2 = \lim_n \|x - y_n\|^2 = K = L^2$, which proves existence. To show uniqueness, suppose there is an element $\tilde{x} \in \mathcal{M}$ that satisfies $\|x - \tilde{x}\|^2 = \|x - \hat{x}\|^2 = K$. Then, another application of the parallelogram law yields

$$\begin{aligned}
0 &\le \|\hat{x} - \tilde{x}\|^2 \\
&= -4\left\|\tfrac{1}{2}(\hat{x} - \tilde{x}) - x\right\|^2 + 2\|\hat{x} - x\|^2 + 2\|\tilde{x} - x\|^2 \\
&\le -4K + 2K + 2K = 0,
\end{aligned}$$

which leads to $\tilde{x} = \hat{x}$.

(b) Suppose first that $\hat{x} \in \mathcal{M}$ and $x - \hat{x} \in \mathcal{M}^\perp$. Then, for all $y \in \mathcal{M}$,

$$\begin{aligned}\|x-y\|^2 &= \langle x-y, x-y\rangle \\ &= \langle (x-\hat{x}) + (\hat{x}-y), (x-\hat{x}) + (\hat{x}-y)\rangle \\ &= \|x-\hat{x}\|^2 + \|\hat{x}-y\|^2 \\ &\geq \|x-\hat{x}\|^2\end{aligned}$$

with equality holding if and only if $y = \hat{x}$. Assume conversely that $\hat{x} \in \mathcal{M}$ but that $x - \hat{x} \notin \mathcal{M}^\perp$. Then $\|\hat{x} - x\| > L$. This can be seen as follows. Since $x - \hat{x} \notin \mathcal{M}^\perp$, there exists $y \in \mathcal{M}$ such that $a = \langle x - \hat{x}, y\rangle \neq 0$. Construct now the element

$$\tilde{x} = \hat{x} + \frac{ay}{\|y\|^2}.$$

We will show that then $\tilde{x}$ is closer to x than $\hat{x}$. Indeed,

$$\begin{aligned}\|x-\tilde{x}\|^2 &= \langle x-\tilde{x}, x-\tilde{x}\rangle \\ &= \langle (x-\hat{x}) + (\hat{x}-\tilde{x}), (x-\hat{x}) + (\hat{x}-\tilde{x})\rangle \\ &= \|x-\hat{x}\|^2 + \|\hat{x}-\tilde{x}\|^2 + 2\mathrm{Re}\langle x-\hat{x}, \hat{x}-\tilde{x}\rangle \\ &= \|x-\hat{x}\|^2 - \frac{|a|^2}{\|y\|^2} \\ &< \|x-\hat{x}\|^2,\end{aligned}$$

where the fourth equality sign follows from $\|\hat{x} - \tilde{x}\|^2 = |a|^2\|y\|^{-2}$ and $2\mathrm{Re}\langle x - \hat{x}, \hat{x} - \tilde{x}\rangle = -2|a|^2\|y\|^{-2}$. This completes the proof. □

Corollary 2.2.1 *If $\mathcal{M}$ is a closed subspace of a Hilbert space $\mathcal{H}$ and if* Id *is the identity mapping on $\mathcal{H}$, there is a unique mapping $P_{\mathcal{M}} \colon \mathcal{H} \to \mathcal{M}$ such that* $\mathrm{Id} - P_{\mathcal{M}} \colon \mathcal{H} \to \mathcal{M}^\perp$. *We call $P_{\mathcal{M}}$ the projection mapping of $\mathcal{H}$ onto $\mathcal{M}$.*

Proof. Let $x \in \mathcal{H}$. By the projection theorem, there is a unique $\hat{x} \in \mathcal{M}$ such that $x - \hat{x} \in \mathcal{M}^\perp$. Therefore, we can define $P_{\mathcal{M}}x = \hat{x}$. □

Remark 2.2.1 *The projection mapping $P_{\mathcal{M}}$ has the following properties:*

(a) $P_{\mathcal{M}}(\alpha x + \beta y) = \alpha P_{\mathcal{M}}x + \beta P_{\mathcal{M}}y$ *for all* $x, y \in \mathcal{H}$ *and* $\alpha, \beta \in \mathbb{C}$;

(b) $\|x\|^2 = \|P_{\mathcal{M}}x\|^2 + \|(\mathrm{Id} - \mathrm{P}_{\mathcal{M}})\mathrm{x}\|^2$ *for all* $x \in \mathcal{H}$;

(c) $x = P_{\mathcal{M}}x + (\mathrm{Id} - \mathrm{P}_{\mathcal{M}})\mathrm{x}$, *and this decomposition is unique;*

(d) $P_{\mathcal{M}}x_n \to P_{\mathcal{M}}x$ *if* $\|x_n - x\| \to 0$;

(e) $x \in \mathcal{M}$ *if and only if* $P_{\mathcal{M}}x = x$, *and* $x \in \mathcal{M}^\perp$ *if and only if* $P_{\mathcal{M}}x = 0$;

(f) $\mathcal{M}_1 \subset \mathcal{M}_2$ *if and only if* $P_{\mathcal{M}_1}P_{\mathcal{M}_2}x = P_{\mathcal{M}_1}x$ *for all* $x \in \mathcal{H}$.

The projection theorem leads to the following prediction equations. Let $\mathcal{H}$ be a Hilbert space. If $x \in \mathcal{H}$ and a closed subspace $\mathcal{M}$ of $\mathcal{H}$ are given, then the closest element of $\mathcal{M}$ to x is $\hat{x}$ such that

$$\langle x - \hat{x}, y \rangle = 0 \qquad \text{for all } y \in \mathcal{M}. \tag{2.2.1}$$

We say that $\hat{x} = P_{\mathcal{M}} x$ is the best predictor of x in $\mathcal{M}$. The prediction equations (2.2.1) are the same as the ones encountered in Examples 2.2.1 and 2.2.2. But more is possible.

Example 2.2.3 Let $(X_t \colon t \in \mathbb{Z})$ be a stationary stochastic process with zero mean and autocovariance function γ. Suppose you have observed $X_1, \ldots, X_n$ and you are interested in finding the best linear approximation

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \phi_{nj} X_{n+1-j}$$

to X_{n+1}. This problem is now easily phrased in terms of projections in the Hilbert space L^2. We aim to minimize the squared norm

$$\|X_{n+1} - \hat{X}_{n+1}\|^2 = E\big[|X_{n+1} - \hat{X}_{n+1}|^2\big],$$

where $\hat{X}_{n+1}$ is a member of the closed subspace $\mathcal{M} = \{\sum_{j=1}^{n} \alpha_j X_{n+1-j} \colon \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}$. From the projection theorem we know that the solution is $\hat{X}_{n+1} = P_{\mathcal{M}} X_{n+1}$ which satisfies the prediction equations

$$\langle X_{n+1} - \hat{X}_{n+1}, Y \rangle = \Big\langle X_{n+1} - \sum_{j=1}^{n} \phi_{nj} X_{n+1-j}, Y \Big\rangle = 0 \qquad \text{for all } Y \in \mathcal{M}.$$

By linearity this reduces to solving the system of linear equations

$$\Big\langle X_{n+1} - \sum_{j=1}^{n} \phi_{nj} X_{n+1-j}, X_k \Big\rangle = 0 \qquad k = n, \ldots, 1.$$

Using the specific inner product in L^2 and defining $\boldsymbol{\phi}_n = (\phi_{n1}, \ldots \phi_{nn})'$, $\boldsymbol{\gamma}_n = (\gamma_1, \ldots, \gamma_n)'$ and $\Gamma_n = (\gamma(j-k))_{j,k=1,\ldots,n}$, we can rewrite the latter equations in matrix-vector form as $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$. The projection theorem guarantees the existence of at least one solution $\boldsymbol{\phi}_n$. If Γ_n is singular, there will be infinitely many solutions each yielding the same (unique) predictor $\hat{X}_{n+1}$.

Example 2.2.4 (Linear Regression) Suppose the linear relationship

$$Y = \beta_1 X^{(1)} + \ldots + \beta_p X^{(p)} + Z,$$

and assume that we have taken observations on $(Y_j, X_j^1, \ldots, X_j^{(p)}$ for $j = 1, \ldots, n$, $n \geq p$. Letting $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, this problem is typically solved by minimizing the residual sum of squares

$$S(\boldsymbol{\beta}) = \sum_{j=1}^{p} \Big(Y_j - \sum_{\ell=1}^{p} \beta_\ell X_j^{(\ell)}\Big)^2 = \Big\|\boldsymbol{Y} - \sum_{\ell=1}^{p} \beta_\ell \boldsymbol{X}^{(\ell)}\Big\|^2,$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ and $\boldsymbol{X}^{(\ell)} = (X_1^{(\ell)}, \ldots, X_n^{(\ell)})'$, $\ell = 1, \ldots, p$. Phrasing the problem in our Hilbert space setting, we can use $\mathcal{H} = L^2(\Omega, \mathcal{A}, P)$ and $\mathcal{M} = \overline{\text{sp}}(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(p)})$. The projection theorem then guarantees the existence of

$$\hat{\boldsymbol{Y}} = P_{\mathcal{M}}\boldsymbol{Y} = \sum_{\ell=1}^{p} \hat{\beta}_\ell \boldsymbol{X}^{(\ell)} = X\hat{\boldsymbol{\beta}},$$

where $X = (\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(p)})$ is a $p \times n$ matrix and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$. The prediction equations for this example become

$$\langle X\hat{\boldsymbol{\beta}}, \boldsymbol{X}^{(\ell)} \rangle = \langle \boldsymbol{Y}, \boldsymbol{X}^{(\ell)} \rangle, \qquad \ell = 1, \ldots, p$$

which can be written more compactly as $X'X\hat{\boldsymbol{\beta}} = X'\boldsymbol{Y}$. In the case that $(X'X)^{-1}$ exists, we obtain the unique least squares solution

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{Y}.$$

Otherwise $\hat{\boldsymbol{Y}}$ is still unique but with infinitely many solutions $\hat{\boldsymbol{\beta}}$.

2.3 Complements

In all that follows we only study a particular type of Hilbert space. We introduce some new terminology first.

Definition 2.3.1 *Let $T \neq \emptyset$ be an index set.*

(a) A set $(e_t \colon t \in T)$ in an inner-product space $\mathcal{H}$ is orthonormal if, for all $s, t \in T$, $\langle e_s, e_t \rangle = \delta_{\{s=t\}}$, where δ is Kronecker's delta.

(b) A subset $(e_t \colon t \in T)$ of a Hilbert space $\mathcal{H}$ is an orthonormal basis of $\mathcal{H}$ if it is an orthonormal set and if $\overline{\text{sp}}(e_t \colon t \in T) = \mathcal{H}$.

(c) A Hilbert space is separable if it has a finite or countable orthonormal basis.

Remark 2.3.1 A separable Hilbert space with orthonormal basis $(e_j \colon j \in \mathbb{Z})$ has the following properties.

(a) If $\mathcal{S} = \cup_{j=1}^{\infty} \overline{\text{sp}}(e_1, \ldots, e_j)$, then $\overline{\mathcal{S}} = \mathcal{H}$;

(b) If $x \in \mathcal{H}$, then $x = \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j$, that is $\|x - \sum_{j=1}^{m} \langle x, e_j \rangle e_j\| \to 0$ as $m \to \infty$;

(c) If $x \in \mathcal{H}$, then $\|x\|^2 = \sum_{j=1}^{\infty} |\langle x, e_j \rangle|^2$;

(d) If $x, y \in \mathcal{H}$, then $\langle x, y \rangle = \sum_{j=1}^{\infty} \langle x, e_j \rangle \langle e_j, y \rangle$. This is known as Parseval's identity.

Example 2.3.1 Let $L^2[-\pi, \pi] = L^2([-\pi, \pi], \mathcal{B}, U)$ be the complex Hilbert space on the interval $[-\pi, \pi]$ with $\mathcal{B}$ being the Borel σ-algebra relative to $[-\pi, \pi]$ and U the uniform measure on $[-\pi, \pi]$ given by its Lebesgue density $U(du) = (2\pi)^{-1}du$. The inner product here is defined as

$$\langle f, g \rangle = E[f\bar{g}] = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(u)\bar{g}(u)du = \int_{[pi,\pi]} f\bar{g}dU.$$

If we define $(e_n\colon n \in \mathbb{Z})$ by setting

$$e_n(u) = \exp(inu),$$

then it is an orthonormal basis in $L^2[-\pi, \pi]$. Orthonormality is seen from

$$\begin{aligned}
\langle e_m, e_n \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)u} du \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \big[\cos(m-n)u + i \sin(m-n)u \big] du \\
&= \begin{cases} 1, & m = n. \\ 0, & m \neq n. \end{cases}
\end{aligned}$$

It can also be shown that $(e_n\colon n \in \mathbb{Z})$ is a basis.

Definition 2.3.2 *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces. A mapping $A\colon \mathcal{H}_1 \to \mathcal{H}_2$ is called an isomorphism if it is one-to-one and onto and if*

$$A(ax + by) = aAx + bAy \qquad \text{for all } x, y \in \mathcal{H}_1 \text{ and } a, b \in \mathbb{C}.$$

A is called an isometric isomorphism if in addition $\langle Ax, Ay \rangle = \langle x, y \rangle$ for all $x, y \in \mathcal{H}_1$.

Remark 2.3.2 An isometric isomorphism $A\colon \mathcal{H}_1 \to \mathcal{H}_2$ has the following properties.

(a) If $(e_t\colon t \in T)$ is an orthonormal basis in $\mathcal{H}_1$, then $(Ae_t\colon t \in T)$ is an orthonormal basis in $\mathcal{H}_2$;

(b) $\|Ax\| = \|x\|$ for all $x \in \mathcal{H}_1$;

(c) $\|x_n - x\| \to 0$ if and only if $\|Ax_n - Ax\| \to 0$;

(d) $AP_{\overline{\mathrm{sp}}(x_t\colon t \in T)}(x) = P_{\overline{\mathrm{sp}}(Ax_t\colon t \in T)}(Ax)$.

Example 2.3.2 We will use the last of the above properties and the isometric isomorphism $AX_t = \exp(it\,\cdot)$ to do prediction in the frequency domain for a stationary process $(X_t\colon t \in \mathbb{Z})$.