1.3 Math 1. Hilbert spaces

In this section I want to recall ideas and definitions that the reader should already know, and introduce her/him to infinite-dimensional spaces.

1.3.1 Convergence in $\mathbb{R}$

Start with the set of real numbers $\mathbb{R}$. A sequence r_n converges to r if given $\epsilon > 0$ there exists an integer n_ϵ such that if $n > n_\epsilon$ then $|r - r_n| < \epsilon$. Since the *triangular inequality* holds in $\mathbb{R}$,

$$|x - y| \leq |x - z| + |z - y|, \tag{1.8}$$

then if the sequence r_n converges to r and r', then $r = r'$ (prove, very easy).

1.3.2 The Cauchy condition is necessary

A sequence fulfills the *Cauchy condition* if given $\epsilon > 0$ there exists n_ϵ such that if $n, m > n_\epsilon$ then $|r_n - r_m| < \epsilon$. Convergent sequences fulfill the Cauchy condition. Easy to prove using (1.8). Thus convergence to a limit implies a condition on the sequence r_n in which the limit is not mentioned: If the sequence approaches r then the terms of the sequence must become very close to one another.

1.3.3 The Cauchy condition is not sufficient in $\mathbb{Q}$

The necessity of the Cauchy condition holds also if instead of the set $\mathbb{R}$ we take the set of rational numbers $\mathbb{Q}$. Obvious. However, there exist sequences in $\mathbb{Q}$ that fulfill the Cauchy condition but do not converge in $\mathbb{Q}$. Take the sequence $r_1 = 1$,

$$r_n = \begin{cases} r_{n-1} \text{ if } [r_{n-1} + 2^{-(n-1)}]^2 > 2 \\ r_{n-1} + 2^{-(n-1)} \text{ otherwise.} \end{cases}$$

Show that this sequence fulfills the Cauchy condition (we say “the sequence is Cauchy” or “is a Cauchy sequence”) and that it does not converge in $\mathbb{Q}$ ($\mathbb{Q}$ does not contain a square root of 2).

1.3.4 Completeness of $\mathbb{R}$

However, Cauchy sequences of real numbers do converge in $\mathbb{R}$. The proof of this theorem can be found in innumerable textbooks of Analysis. You start with $\mathbb{Q}$, define $\mathbb{R}$ as the set of Dedekind cuts in $\mathbb{Q}$, and prove the result, that is, that filling the gaps in the set of rational numbers has produced another set, the real numbers, in which there are no gaps. The theorem is equivalent to the statement that any bounded subset of $\mathbb{R}$ has a least upper bound.

1.3.5 Norms and distances in $\mathbb{R}^n$

Now consider the vector space $\mathbb{R}^n$. We can define a length or *norm* in $\mathbb{R}^n$ as follows. Given $\mathbf{v} = (v_1,\ v_2,\ \ldots,\ v_n)$,

$$|\mathbf{v}| = \sqrt{\sum_{k=1}^{n} v_k^2},$$

and a distance $\mathrm{d}(\mathbf{v}, \mathbf{w}) = |\mathbf{v} - \mathbf{w}|$. These norm and distance are called Euclidean.

Other definitions of distance are associated with the norms

$$|\mathbf{v}|_1 = \sum_{k=1}^{n} |v_k|, \qquad |\mathbf{v}|_2 = \max_k |v_k|.$$

Note that the three norms coincide for $n = 1$. The distance d and the distances corresponding to the other two norms fulfill the four axioms of a *metric space*:
(D1) The distance of $\mathbf{v}$ from $\mathbf{v}$ is zero.
(D2) If $\mathbf{v}$ and $\mathbf{w}$ are different then their distance is positive.
(D3) The distance is symmetric.
(D4) The triangular inequality holds: $\mathrm{d}(\mathbf{v}, \mathbf{w}) \leq \mathrm{d}(\mathbf{v}, \mathbf{z}) + \mathrm{d}(\mathbf{z}, \mathbf{w})$.
The proof of (D4) is very easy for the second and third distances. The proof for the Euclidean distance is given below.

Given a norm we define convergence: $\mathbf{v}_n \to \mathbf{v}$ if $\mathrm{d}(\mathbf{v}, \mathbf{v}_n) \to 0$. Although the above three norms define different distances they are equivalent in the following sense: Sequences which are converging with any one of the three distances converge with the other two. The triangular inequality ensures that if a sequence converges to $\mathbf{v}$ and $\mathbf{v}'$ then $\mathbf{v} = \mathbf{v}'$.

Observation 1.2 I use boldface characters for vectors only when I cannot avoid them. Here, as well as in Section 1.3.8, they are necessary to distinguish the coordinates v_k of the vector $\mathbf{v}$ from the vectors of the sequence $\mathbf{v}_k$, $k \in \mathbb{N}$. However, in Section 1.3.10, whose subject is general Hilbert spaces, I use regular characters.

1.3.6 Completeness of $\mathbb{R}^n$

Let us stick to the Euclidean norm. Note that if $\mathbf{v}_n \in \mathbb{R}^n$, $\mathbf{v}_n \to \mathbf{v}$ if and only if each of the coordinates of $\mathbf{v}_n$ converges to the corresponding coordinate of $\mathbf{v}$.

Analogously, the sequence $\mathbf{v}_n$ is Cauchy if and only if all the coordinate sequences are Cauchy. Thus completeness of $\mathbb{R}^n$, with the Euclidean norm, is an elementary consequence of the completeness of $\mathbb{R}$.

The reason why completeness is important is that in many a situation we can prove that a sequence is Cauchy but are not able to construct the limit. Completeness then ensures that a limit exists.

1.3.7 Inner product in $\mathbb{R}^n$

In the space $\mathbb{R}^n$ we define the *inner product* as follows:

$$\mathbf{v} \cdot \mathbf{w} = \sum_{k=1}^{n} v_k w_k . \tag{1.9}$$

The following properties hold:
(P1) $\mathbf{v} \cdot \mathbf{v} \geq 0$. Moreover, $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
(P2) $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$.
(P3) $(\alpha \mathbf{v}) \cdot \mathbf{w} = \alpha(\mathbf{v} \cdot \mathbf{w})$.
(P4) $(\mathbf{v} + \mathbf{w}) \cdot \mathbf{z} = \mathbf{v} \cdot \mathbf{z} + \mathbf{w} \cdot \mathbf{z}$.
The inner product and the Euclidean norm are linked by $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$. We have, using (P4),

$$|\mathbf{v} + \mathbf{w}|^2 = |\mathbf{v}|^2 + |\mathbf{w}|^2 + 2\mathbf{v} \cdot \mathbf{w}.$$

Thus if $\mathbf{v} \cdot \mathbf{w} = 0$,

$$|\mathbf{v} + \mathbf{w}|^2 = |\mathbf{v}|^2 + |\mathbf{w}|^2.$$

This equality looks like the Pitagorean Theorem: If we define *orthogonality* between $\mathbf{v}$ and $\mathbf{w}$ by $\mathbf{v} \cdot \mathbf{w} = 0$, then the diagonal of the rectangle constructed on $\mathbf{v}$ and $\mathbf{w}$ has square length equal the sum of the square lengths of $\mathbf{v}$ and $\mathbf{w}$.

To test whether the above definition of orthogonality corresponds to our elementary notion, let us consider the following problem. Given the vectors

$$\mathbf{v},\ \mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_s,$$

find the vector $\mathbf{w}$ which (i) is a linear combination of the vectors $\mathbf{v}_j$, and (ii) has minimum distance from $\mathbf{v}$. This is equivalent to finding s real numbers a_j such that

$$\left| \mathbf{v} - \sum_{j=1}^{s} a_j \mathbf{v}_j \right|^2 = |\mathbf{v}|^2 - \sum_{j=1}^{s} (\mathbf{v} \cdot \mathbf{v}_j) a_j + \sum_{j=1}^{s} \sum_{k=1}^{s} (\mathbf{v}_j \cdot \mathbf{v}_k) a_j a_k$$

is minimum. Of course we can assume that the vectors $\mathbf{v}_j$ form an independent s-tuple.

Note that we are minimizing the sum of the squares of the coordinates of the vector $\mathbf{v} - \sum a_j \mathbf{v}_j$; this is why the procedure ending up with the solution is called *least squares*. Taking the derivatives with respect to a_j,

$$\sum_{k=1}^{s} (\mathbf{v}_j \cdot \mathbf{v}_k) a_k = \mathbf{v} \cdot \mathbf{v}_j, \tag{1.10}$$

for $j = 1, \ldots, s$. In matrix form,

$$C\mathbf{a} = \mathbf{b},$$

where C has $\mathbf{v}_j \cdot \mathbf{v}_k$ as its (j, k) entry, $\mathbf{a}$ and $\mathbf{b}$ are the vectors with components a_j and $\mathbf{v} \cdot \mathbf{v}_j$ respectively.

Now define the *orthogonal projection* of $\mathbf{v}$ on the subspace spanned by $\mathbf{v}_1$, $\mathbf{v}_2$, $\ldots$, $\mathbf{v}_s$, as a vector $\mathbf{w} = a_1\mathbf{v}_1 + \cdots + a_s\mathbf{v}_s$, such that $\mathbf{v} - \mathbf{w}$ is orthogonal to all the vectors $\mathbf{v}_j$. The orthogonality conditions are

$$(\mathbf{v} - \mathbf{w}) \cdot \mathbf{v}_j = [\mathbf{v} - (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_s\mathbf{v}_s)] \cdot \mathbf{v}_j = 0,$$

for $j = 1, \ldots, s$. It is easily seen that this system of equations is (1.10) rewritten in a slightly different way.

Thus the orthogonal projection coincides with the minimum distance vector, which corresponds to geometric intuition based on two or three dimensional spaces.

Observation 1.3 Orthogonal projections and least squares occur all the time in linear regressions. Over the sample period $1, 2, \ldots, T$, the samples

$$\begin{aligned} \mathbf{y} &= (y_1, y_2, \ldots, y_T) \\ \mathbf{x}_j &= (x_{j1}, x_{j,2}, \ldots, x_{jT}), \quad \text{for } j = 1, 2, \ldots, m, \end{aligned}$$

are given, and we seek coefficients $a_0, a_1, \ldots, a_m$, such that

$$|\mathbf{y} - (a_0 + a_1\mathbf{x}_1 + \cdots + a_m\mathbf{x}_m)|^2 = \sum_{t=1}^{T} [\mathbf{y}_t - (a_0 + a_1\mathbf{x}_{1t} + \cdots + a_m\mathbf{x}_{mt})]^2,$$

which is the square norm of a vector of $\mathbb{R}^T$, is minimum. This is equivalent to seeking the coefficients a_j such that $\mathbf{y} - (a_0 + a_1\mathbf{x}_1 + \cdots + a_m\mathbf{x}_m)$ is orthogonal to $\mathbf{x}_j$ for $j = 1, 2, \ldots, m$, and to the vector $(1, 1, \ldots, 1)$. See, for example, [6, pp. 60-62].

Using the orthogonal projection we can prove the Cauchy-Schwartz inequality:

$$|\mathbf{x}\cdot\mathbf{y}| \le |\mathbf{x}|\,|\mathbf{y}|. \tag{1.11}$$

Firstly note that if $\mathbf{y} = \mathbf{0}$ then, by property (P3) of the inner product, $\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot(0\mathbf{y}) = 0\,(\mathbf{x}\cdot\mathbf{y}) = 0$, so that the inequality holds. Assume that $\mathbf{y} \neq 0$ and consider the orthogonal projection of $\mathbf{x}$ on $\mathbf{y}$:

$$\mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{y}}{\mathbf{y}\cdot\mathbf{y}}\,\mathbf{y} + \mathbf{R} \tag{1.12}$$

(check this). Since $\mathbf{R} \perp \mathbf{y}$,

$$|\mathbf{x}|^2 = \frac{(\mathbf{x}\cdot\mathbf{y})^2}{(\mathbf{y}\cdot\mathbf{y})^2}\,|\mathbf{y}|^2 + |\mathbf{R}|^2 = \frac{(\mathbf{x}\cdot\mathbf{y})^2}{|\mathbf{y}|^2} + |\mathbf{R}|^2.$$

This implies

$$\mathbf{x}\cdot\mathbf{x} \le \frac{(\mathbf{x}\cdot\mathbf{y})^2}{\mathbf{y}\cdot\mathbf{y}}.$$

The conclusion follows. The reader may think that this proof may be given in a simpler way under definition (1.9) of the inner product. For example, resorting to property (P3) to prove that $\mathbf{x}\cdot\mathbf{0} = 0$ is a little funny. However, the proof given here is valid in general, i.e. for any space and any definition of inner product, provided it fulfills properties (P1), (P2), (P3) and (P4).

An easy consequence of the Cauchy-Schwartz inequality is the triangular inequality:

$$\begin{aligned} |\mathbf{x}-\mathbf{y}|^2 &= |\mathbf{x}-\mathbf{z}|^2 + |\mathbf{z}-\mathbf{y}|^2 + 2(\mathbf{x}-\mathbf{z})\cdot(\mathbf{z}-\mathbf{y}) \\ &\le |\mathbf{x}-\mathbf{z}|^2 + |\mathbf{z}-\mathbf{y}|^2 + 2|\mathbf{x}-\mathbf{z}|\,|\mathbf{z}-\mathbf{y}| = (|\mathbf{x}-\mathbf{z}| + |\mathbf{z}-\mathbf{y}|)^2\,. \end{aligned}$$

The conclusion follows.

Taking the inner product of both sides of (1.12) by $\mathbf{y}$ we obtain

$$\mathbf{x}\cdot\mathbf{y} = \left(\frac{\mathbf{x}\cdot\mathbf{y}}{\mathbf{y}\cdot\mathbf{y}}\,\mathbf{y}\right)\cdot\mathbf{y},$$

so that the inner product of $\mathbf{x}$ and $\mathbf{y}$ is equal to the product of $\mathbf{y}$ and the projection of $\mathbf{x}$ on $\mathbf{y}$. Lastly, by (1.11),

$$-1 < \frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|\,|\mathbf{y}|} \le 1.$$

Thus, defining the angle between $\mathbf{x}$ and $\mathbf{y}$ as

$$\theta_{\mathbf{x},\mathbf{y}} = \arccos\left(\frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|\,|\mathbf{y}|}\right), \quad 0 \le \theta_{\mathbf{x},\mathbf{y}} \le \pi,$$

we have $\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}| \cos \theta_{\mathbf{x},\mathbf{y}}$, which is the elementary definition of inner product in $\mathbb{R}^2$ or $\mathbb{R}^3$. We can immediately check that orthogonality of $\mathbf{x}$ and $\mathbf{y}$ corresponds to $\theta_{\mathbf{x},\mathbf{y}} = \pi/2$, and that if $\mathbf{x} = \alpha\mathbf{y}$ then $\theta_{\mathbf{x},\mathbf{y}}$ is equal to 0 or π, according to the sign of α.

1.3.8 The infinite-dimensional space $l^2(-\infty, \infty)$

All this is very familiar, or should be. Now let us try to generalize what we have just recalled to an infinite-dimensional vector space. Consider the set of all the bilateral sequences $\{r_k,\ k \in \mathbb{Z}\}$, where r_k is a real number, such that

$$\sum_{k=-\infty}^{\infty} r_t^2 < \infty.$$

The set of all these square summable sequences is denoted by $l^2(-\infty, \infty)$. The sum of two vectors $\mathbf{v}$ and $\mathbf{w}$ being defined coordinatewise, we must check that $\mathbf{v}+\mathbf{w}$ still belongs to $l^2(-\infty, \infty)$. For, given a sequence $\mathbf{v}$ and a positive integer s, define $\mathbf{v}^{[s]}$ as the s-truncation of $\mathbf{v}$, i.e.

$$\mathbf{v}_k^{[s]} = \begin{cases} v_k & \text{if } |k| \le s \\ 0 & \text{if } |k| > s. \end{cases}$$

Setting $\mathbf{x} = \mathbf{v} + \mathbf{w}$, using (1.11)

$$|\mathbf{x}^{[s]}|^2 = |\mathbf{v}^{[s]}|^2 + |\mathbf{w}^{[s]}|^2 + 2\mathbf{v}^{[s]} \cdot \mathbf{w}^{[s]} \le |\mathbf{v}^{[s]}|^2 + |\mathbf{w}^{[s]}|^2 + 2|\mathbf{v}^{[s]}|\,|\mathbf{w}^{[s]}|$$

(we have applied the Cauchy-Schwartz inequality to $\mathbf{v}^{[s]} \cdot \mathbf{w}^{[s]}$). By assumption $|\mathbf{v}^{[s]}|$ and $|\mathbf{w}^{[s]}|$ converge to finite limits, so that the left hand side converges to a finite limit. Thus $l^2(-\infty, \infty)$ is a vector space.

Exercise 1.5 Prove that the sequence $r_k = \alpha^{|k|}$, with $|\alpha| < 1$, belongs to the space $l^2(-\infty, \infty)$. Prove also that if there exist $A > 0$ and α, with $0 \le \alpha < 1$, such that $|r_k| \le A\alpha^{|k|}$, then the sequence r_k belongs to $l^2(-\infty, \infty)$.

Observation 1.4 The definition of $l^2(-\infty, \infty)$ requires that the sum of the squares converges, not that $\sum r_k$ or $\sum |r_k|$ converges. You should remember that the series

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots$$

diverges, but that

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots$$

converges. Thus the sequence $r_k = \dfrac{1}{|k| + 1}$ belongs to $l^2(-\infty, \infty)$.

Obviously $l^2(-\infty,\infty)$ has no finite dimension, i.e. no finite number of vectors can generate the whole space by linear combinations. For, observe that the set of vectors $\mathbf{v}$ such that $v_k = 0$ for $k > n$ is a copy of $\mathbb{R}^n$, so that an n-tuple of independent vectors exists in $l^2(-\infty,\infty)$ for all n. We can define an inner product as

$$\mathbf{v}\cdot\mathbf{w} = \sum_{k=-\infty}^{\infty} v_k w_k = \lim_{s\to\infty} \mathbf{v}^{[s]}\cdot\mathbf{w}^{[s]} \tag{1.13}$$

(convergence of the right hand side is, again, a consequence of the Cauchy-Schwartz inequality applied to $\mathbf{v}^{[s]}\cdot\mathbf{w}^{[s]}$). The corresponding norm and distance are defined as

$$\|\mathbf{v}\|^2 = \sum_{k=-\infty}^{\infty} v_k^2, \qquad \|\mathbf{v}-\mathbf{w}\|^2 = \sum_{k=-\infty}^{\infty} (v_k - w_k)^2.$$

Now observe that we can reproduce almost all the constructions and statements of Section (1.3.7), namely:

(I) The definition of orthogonality. The construction of the vector of the subspace generated by a *finite* number of vectors $\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_s$, which has minimum distance from a given vector $\mathbf{v}$.

(II) The orthogonal projection of $\mathbf{v}$ on the subspace generated by $\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_s$. The proof that the minimum distance vector and the orthogonal projection coincide.

(III) The Cauchy-Schwartz and the triangular inequality.

(IV) The triangular inequality implies the uniqueness of the limit of a convergent sequence (see Section 1.3.5) and the necessity of the Cauchy condition (see Section 1.3.2).

For, note that (I), (II) and (III) depend only on the properties (P1), (P2), (P3) and (P4) of the inner product (see Section 1.3.7), and that such properties hold for the inner product (1.13).

Moreover, it is clear that if the sequence $\mathbf{v}_n$ converges to $\mathbf{v}$ then all coordinate sequences of $\mathbf{v}_n$ converge to the corresponding coordinates of $\mathbf{v}$; and that if $\mathbf{v}_n$ is Cauchy then all its coordinates are Cauchy. However, the converse is not true.

Example 1.10 Let $\mathbf{v}_n = \{v_{nk},\ k \in \mathbb{Z}\}$, where

$$v_{nk} = \begin{cases} \alpha_n \text{ if } |k| \le n \\ 0 \ \text{ if } |k| > n. \end{cases}$$

Suppose that $\alpha_n \to 0$, then all coordinates tend to zero as $n \to \infty$. However,

$$\|\mathbf{v}_n\|^2 = (2n+1)\alpha_n^2,$$

which does not converge to zero if for example $\alpha_n = 1/\sqrt{n}$. Thus the simple rule stated in Section 1.3.6, that convergence in $\mathbb{R}^n$ is nothing else than coordinatewise convergence, does not extend to the infinite-dimensional space $l^2(-\infty, \infty)$.

An immediate consequence of this example is that completeness of $l^2(-\infty, \infty)$ cannot be obtained in a simple way from completeness of $\mathbb{R}$. So we need a proof that if $\mathbf{v}_n$ is a Cauchy sequence in $l^2(-\infty, \infty)$, there exists $\mathbf{v} \in l^2(-\infty, \infty)$ such that $\mathbf{v}_n \to \mathbf{v}$, i.e.

$$\lim_{n\to\infty} \sum_{k=-\infty}^{\infty} (v_k - v_{nk})^2 = 0.$$

Given $\mathbf{x} \in l^2(-\infty, \infty)$, set $\mathbf{x}^{(s)} = \mathbf{x} - \mathbf{x}^{[s]}$, the tail of $\mathbf{x}$. Note that $\mathbf{x}^{(s)} \perp \mathbf{x}^{[s]}$. The proof of completeness of $l^2(-\infty, \infty)$ that follows—a very interesting exercise—is based on the idea that tails are small when s is large, and that the tails of the elements of a Cauchy sequence are uniformly small.

Step 1. Since $\mathbf{v}_n$ is Cauchy then each coordinate sequence is Cauchy and therefore converges. Let

$$v_k = \lim_{n\to\infty} v_{nk},$$

for $k \in \mathbb{Z}$. We must prove that $\mathbf{v} = \{v_k,\ k \in \mathbb{Z}\}$ is a member of $l^2(-\infty, \infty)$ and that $\mathbf{v}_n \to \mathbf{v}$.

Step 2. The sequence $\mathbf{v}_n$ is bounded. For, using the triangular inequality,

$$\|\mathbf{v}_n\| \le \|\mathbf{v}_n - \mathbf{v}_m\| + \|\mathbf{v}_m\|.$$

Given $\epsilon > 0$, for all $n > n_\epsilon$ and a fixed $m > n_\epsilon$,

$$\|\mathbf{v}_n\| \le \epsilon + \|\mathbf{v}_m\|.$$

Thus $\mathbf{v}_n$ is bounded by the greater between $\max_{n=1,\ldots,n_\epsilon} \|\mathbf{v}_n\|$ and $\epsilon + \|\mathbf{v}_m\|$.

Step 3. Of course $\mathbf{v}$ belongs to $l^2(-\infty, \infty)$ if and only if $\mathbf{v}^{[s]}$ is bounded; in that case $\|\mathbf{v}\|^2 = \lim_{n\to\infty} \|\mathbf{v}^{[s]}\|^2$. Suppose that $\mathbf{v}^{[s]}$ is not bounded. Given $\epsilon > 0$ and s, there exists $n_{\epsilon,s}$ such that for $n > n_{\epsilon,s}$,

$$\|\mathbf{v}^{[s]} - \mathbf{v}_n^{[s]}\| < \epsilon.$$

On the other hand,

$$\|\mathbf{v}^{[s]}\| \le \|\mathbf{v}_n^{[s]}\| + \|\mathbf{v}^{[s]} - \mathbf{v}_n^{[s]}\|,$$

so that

$$\|\mathbf{v}_n^{[s]}\| \ge \|\mathbf{v}^{[s]}\| - \epsilon.$$

But then there exists a sequence $\mathbf{v}_{n_s}^{[s]}$, with $\lim_{s\to\infty} n_s = \infty$, such that $\|\mathbf{v}_{n_s}^{[s]}\| > \|\mathbf{v}^{[s]}\| - \epsilon$, so that $\|\mathbf{v}_{n_s}^{[s]}\| \to \infty$ and therefore $\|\mathbf{v}_{n_s}\| \to \infty$, contrary to the result of Step 2. Thus we only have to prove that $\mathbf{v}_n \to \mathbf{v}$.

Step 4. (The tails of the vectors $\mathbf{v}_n$ are uniformly small.) Given ϵ, there exist $m(\epsilon)$ and $s(\epsilon)$ (a slight change of notation is useful here) such that for $n > m(\epsilon)$ and $s > s(\epsilon)$, $\|\mathbf{v}_n^{(s)}\| < \epsilon$. For, let $m(\epsilon)$ be such that for $n, m > m(\epsilon)$,

$$\|\mathbf{v}_n - \mathbf{v}_m\|^2 = \|\mathbf{v}_n^{[s]} - \mathbf{v}_m^{[s]}\|^2 + \|\mathbf{v}_n^{(s)} - \mathbf{v}_m^{(s)}\|^2 < \epsilon^2/4,$$

so that

$$\|\mathbf{v}_n^{(s)} - \mathbf{v}_m^{(s)}\|^2 < \epsilon^2/4.$$

Now pick up any $h > m(\epsilon)$ and let $s(\epsilon)$ be such that, for $s > s(\epsilon)$, $\|\mathbf{v}_h^{(s)}\| < \epsilon/2$. Then, for $n > m(\epsilon)$ and $s > s(\epsilon)$,

$$\|\mathbf{v}_n^{(s)}\| \leq \|\mathbf{v}_h^{(s)}\| + \|\mathbf{v}_n^{(s)} - \mathbf{v}_h^{(s)}\| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Step 5. Given $\epsilon > 0$ and s there exist $M(\epsilon, s)$ such that if $n > M(\epsilon, s)$ then $\|\mathbf{v}^{(s)}\| < \sqrt{\epsilon/5}$ and $\|\mathbf{v}^{[s]} - \mathbf{v}_n^{[s]}\|^2 < \epsilon/5$. Given ϵ, set

$$N(\epsilon) = \max[m(\sqrt{\epsilon/5}), M(\epsilon, s(\sqrt{\epsilon/5}))].$$

Now take $n > N(\epsilon)$. The following inequality

$$\begin{aligned} \|\mathbf{v} - \mathbf{v}_n\|^2 &= \|\mathbf{v}^{[s]} - \mathbf{v}_n^{[s]}\|^2 + \|\mathbf{v}^{(s)} - \mathbf{v}_n^{(s)}\|^2 \\ &\leq \|\mathbf{v}^{[s]} - \mathbf{v}_n^{[s]}\|^2 + \|\mathbf{v}^{(s)}\|^2 + \|\mathbf{v}_n^{(s)}\|^2 + 2\|\mathbf{v}^{(s)}\|\,\|\mathbf{v}_n^{(s)}\| \end{aligned}$$

holds for any s, and therefore for $s > s(\sqrt{\epsilon/5})$. Thus, for $n > N(\epsilon)$ we have $\|\mathbf{v} - \mathbf{v}_n\|^2 < \epsilon$.

Observation 1.5 Note that Step 2 is a general Lemma: A Cauchy sequence in a metric space is bounded, irrespective of whether the space is complete or not.

Observation 1.6 Obviously all the definitions and results of this sections apply to the space of "one-sided" sequences $l^2(1, \infty)$.

1.3.9 The space L^2

Consider a measure space $(\Omega, \mathcal{F}, m)$, i.e a set, a σ-field and a measure. We denote by $L^2(\Omega, \mathcal{F}, m)$ the set of all the functions $f : \Omega \to \mathbb{R}$ that are measurable with respect to $\mathcal{F}$ and such that

$$\int_\Omega f(\omega)^2 dm(\omega) < \infty.$$

Definition 1.4 When $\Omega = [a\ b] \subseteq \mathbb{R}$, $\mathcal{F}$ is the Borel σ-field and m is the Lebesgue measure, $m([c\ d]) = d - c$, we write $L^2([a\ b])$.

Exercise 1.6 Let $\Omega = [0\ 1]$. Consider the functions $g_\alpha(r) = r^{-\alpha}$ and give the condition on α ensuring that g_α belongs to $L^2([0\ 1])$.

Exercise 1.7 The set $L^1(\Omega, \mathcal{F}, m)$ is defined as the set of all functions f such that

$$\int_\Omega |f(\omega)| dm(\omega) < \infty.$$

If $m(\Omega) < \infty$ then $L^1(\Omega, \mathcal{F}, m) \supseteq L^2(\Omega, \mathcal{F}, m)$ (see Exercise 1.3). Find an example in which $m(\Omega) = \infty$ and this inclusion does not hold.

We have to show that if f and g belong to $L^2(\Omega, \mathcal{F}, m)$, then $f + g$ belong to $L^2(\Omega, \mathcal{F}, m)$ as well. We can use the following inequality:

$$|a + b|^p \leq 2^p(|a|^p + |b|^p),$$

holding for any complex numbers a and b and any real number $p \geq 0$. For, assuming that $|b| \geq |a|$,

$$|a + b|^p \leq (|a| + |b|)^p \leq 2^p \left(\frac{|a| + |b|}{2}\right)^p \leq 2^p|b|^p \leq 2^p(|a|^p + |b|^p).$$

If $p > 1$ then $[(a+b)/2]^p \leq (a^p + b^p)/2$, so that $|a+b|^p \leq 2^{p-1}(|a|^p + |b|^p)$. Thus $(f(\omega) + g(\omega))^2 \leq 2(f(\omega)^2 + g(\omega)^2)$, which implies the result we need.

Now note that $fg = (f + g)^2 - f^2 - g^2$, so that the integral of fg is finite, and define the inner product as

$$f \cdot g = \int_\Omega f(\omega)g(\omega)dm(\omega), \tag{1.14}$$

the corresponding norm and distance being respectively

$$\|f\| = \sqrt{\int_\Omega f(\omega)^2 dm(\omega)}, \quad \|f - g\| = \sqrt{\int_\Omega (f(\omega) - g(\omega))^2 dm(\omega)}$$

Convergence with respect to the norm $\|\cdot\|$ is called *mean-square* convergence.

Observation 1.7 When $(\Omega, \mathcal{F}, P)$ is a probability space, x_n converges to x in mean square if and only if both the variance and the mean of $x - x_n$ converge to zero. For, observe that $\text{var}(x - x_n) + [E(x - x_n)]^2 = \|x - x_n\|^2$.

Observation 1.8 Definitions 1.9 and 1.13 are particular cases of 1.14. For 1.9, set $\Omega = \{1, 2, \ldots, n\}$, $m(\{k\}) = 1$, for $k = 1, \ldots, n$, and note that the vector $(x_1, x_2, \ldots, x_n)$ is nothing other than the function that maps k to x_k. Specify Ω and m for 1.13.

Properties (P2), P3) and (P4) of Section 1.3.7 hold for the inner product (1.14). Property (P1) also holds, provided that we identify functions that differ only on a zero measure subset of Ω, i.e. functions that are equal *almost everywhere* (*almost surely*, that is with probability one, when m is a probability measure). We will use sometimes the abbreviations "a.e" and "a.s". Having this identification in mind we will use expressions like "$f = 0$", with the meaning "$f(\omega) = 0$ almost everywhere in Ω". Note that the distance between f and g, that is $\|f - g\|$, is zero if and only if $f(\omega) = g(\omega)$ almost everywhere in Ω.

Observation 1.9 A standard result in Lebesgue-integral theory is that a *non-negative* function g is zero almost everywhere in Ω if and only if $\int_\Omega g(\omega)dm(\omega) = 0$ (see [9], p. 104, Theorem B; [15], p. 332, Exercise 1). This motivates the statements just above. For,

$$f \cdot f = \int_\Omega |f(\omega)|^2 dm(\omega).$$

Thus $f \cdot f = 0$ if and only if $f = 0$ almost everywhere in Ω.

Just as in Section 1.3.7, we obtain the orthogonal projection of f on the subspace spanned by $f_1, f_2, \ldots, f_s$, the Cauchy-Schwartz inequality, the triangular inequality, uniqueness of the limit of convergent sequences, necessity of the Cauchy condition.

Exercise 1.8 The L^2 spaces are typically infinite dimensional. Consider $L^2([0\ 1])$. Recall that the *indicator function* of a set A, denoted by χ_A, is the function whose value is 1 on A and 0 elsewhere. Let $B_{n,k} = \left[\frac{k-1}{n}\ \frac{k}{n}\right]$, $k = 1, \ldots, n$. Prove that the n functions

$$\chi_{B_{1,n}}, \chi_{B_{2,n}}, \ldots, \chi_{B_{n,n}}$$

are independent. Thus there are n independent functions for any n.

The proof that $L^2(\Omega, \mathcal{F}, m)$ is complete, though not difficult, would require the introduction of intermediate results that are outside the scope of these lectures, so I will skip it. However, let me remark an important difference with respect to the proof of completeness for $l^2(-\infty, \infty)$. Mean-square convergence of f_n to f does not imply convergence of $f_n(\omega)$ to $f(\omega)$ for all $\omega \in \Omega$, or convergence for ω a.e. in Ω. It does not even imply convergence for at least one $\omega \in \Omega$. The following exercise provides a famous example.

Exercise 1.9 Consider the space of Exercise 1.8 and define the sequence f_n as follows. Let

$$A_1 = [0 \;\; \tfrac{1}{2}], \quad A_2 = (\tfrac{1}{2} \;\; 1],$$

$$A_3 = [0 \;\; \tfrac{1}{4}], \quad A_4 = (\tfrac{1}{4} \;\; \tfrac{1}{2}], \quad A_5 = (\tfrac{1}{2} \;\; \tfrac{3}{4}], \quad A_6 = (\tfrac{3}{4} \;\; 1]$$

$$A_7 = [0 \;\; \tfrac{1}{8}], \quad A_8 = (\tfrac{1}{8} \;\; \tfrac{1}{4}], \qquad \text{etc.}$$

and let $f_k = \chi_{A_k}$. *Prove that f_k converges in mean square to 0, but that $f_k(r)$ does not converge for any $r \in [0 \;\; 1]$.* However, prove that you can find a subsequence of f_k which converges a.e. to 0 in $[0 \;\; 1]$ (you find a hint for a solution to this problem in Exercise 1.10).

This has the consequence that $f_n(\omega)$ does not inherit the Cauchy property from f_n, so that the line of reasoning applied to prove the completeness of $l^2(-\infty, \infty)$, firstly constructing the limit by exploiting coordinatewise convergence, then proving convergence in the norm of $l^2(-\infty, \infty)$, cannot be pursued here. What can be proved is that under the Cauchy condition there exists a subsequence converging a.e. in Ω and that the limit is a function f belonging to $L^2(\Omega, \mathcal{F}, P)$. Then mean-square convergence of f_n to f must be demonstrated. Proofs of completeness for $L^2(\Omega, \mathcal{F}, m)$ can be found in [6, pp. 68-69], [3, pp. 297-298], [13, pp. 123-126].

Exercise 1.10 Consider the sequence of functions in Exercise 1.9. Construct a subsequence by selecting one function in each row and prove that it converges almost everywhere.

Observation 1.10 In Exercise 1.1 we claim that the series $\sum_{j=1}^{k} \alpha^j Z_{t-j}$ converges surely and in mean square. Both statements can be proved directly in that particular case. More precisely, we firstly prove that the partial sums converge surely and call Y_t^* the limit. Then

$$\left\| Y_t^* - \sum_{j=1}^{k} \alpha^j Z_{t-j} \right\|^2 = \left\| \sum_{j=k+1}^{\infty} \alpha^j Z_{t-j} \right\|^2 = \sigma_Z^2 \sum_{j=k+1} \alpha^{2j} = \sigma_Z^2 \frac{\alpha^{2(k+1)}}{1-\alpha^2},$$

so that the partial sum converges to Y_t^* in mean square. Note that, by virtue of the completeness theorem for L^2 spaces, convergence in mean square can also be obtained as a consequence of the Cauchy condition holding for the partial sums sequence (we have proved this on p. 5). This however, firstly, would only provide existence of the limit, not its construction, and, secondly would not imply that the partial sums converge surely to the mean square limit.

Observation 1.11 Completeness of $l^2(-\infty, \infty)$, directly proved in Section 1.3.8, is also a consequence of the general theorem stating completeness of $L^2(\Omega, \mathcal{F}, m)$, see Observation 1.8.

1.3.10 Hilbert spaces and the projection theorem

Now we take an axiomatic approach. Suppose that H is a vector space with an inner product fulfilling properties (P1), (P2), (P3) and (P4), Section 1.3.7. Define the norm as $\|v\| = \sqrt{v \cdot v}$, the distance as $\|v - w\|$. Orthogonal projection on finite-dimensional subspaces, Cauchy-Schwartz inequality, triangular inequality, necessity of the Cauchy condition, are all obtained as in Section 1.3.7.

If H is complete, i.e. if the Cauchy condition is also sufficient for convergence we say that H is a *Hilbert space.*

Example 1.11 The spaces $\mathbb{R}^n$, $L^2(\Omega, \mathcal{F}, P)$, $l^2(-\infty, \infty)$, with their respective inner product, are Hilbert spaces.

Exercise 1.11 Consider the subspace $\ell^2 \subset l^2(-\infty, \infty)$ whose elements are all the sequences $\mathbf{v}$ such that $\mathbf{v}^{(n)} = 0$ for some n. This is obviously an infinite-dimensional vector space. Prove that it is not complete and therefore is not a Hilbert space.

Exercise 1.12 Consider the subspace $C \subset L^2([0\ 1])$ whose elements are all continuous functions. This is obviously an infinite-dimensional vector space. Prove that it is not complete and therefore is not a Hilbert space.

Exercise 1.13 Let H be a Hilbert space and let S be a *vector subspace* of H, that is a subset such that if v and w are in H, then $\alpha v + \beta w$ is in H. Obviously S is a vector space with an inner product, so that we may ask the question whether it is a Hilbert space or not. The answer is that S is a Hilbert space if and only if it is a *closed subset* of H (i.e. if $x_n \in S$ and $x_n \to x$ implies $x \in S$). For example, ℓ^2, Exercise 1.11, and C, Exercise 1.12, as subspaces of $l^2(-\infty, \infty)$ and $L^2([0\ 1]$ respectively, are not closed. The subset of $L^2([0\ 1])$ containing all the functions whose value is zero in $[\frac{1}{2}\ 1]$, call it A, is closed. The subset of $l^2(-\infty, \infty)$ containing all sequences $\{r_k,\ k \in \mathbb{Z}\})$ such that $r_k = 0$ for k odd, call it B, is closed.

Now let H be a Hilbert space, S a closed subspace of H, and v any vector of H. As we have argued above, if

$$S = \{y,\ y = a_1 v_1 + a_2 v_2 + \cdots + a_h v_h,\ a_j \in \mathbb{R}\},$$

i.e. if S is a finite-dimensional subspace, then we know how to obtain the orthogonal projection of x on S, and that such projection is the vector of S that minimizes the distance from x. For, we just have to solve the system of equations (1.10), which is reproduced here:

$$\sum_{k=1}^{s} (v_j \cdot v_k) a_k = v \cdot v_j.$$

Now we want to extend the construction of the orthogonal projection to the case in which S is infinite-dimensional. We will make use of the following general statement.

Proposition 1.1 The inner product and the norm are continuous, i.e. if $v_n \to v$ and $w_n \to w$ then $w_n \cdot v_n \to w \cdot v$. In particular $\|v_n\| \to \|v\|$.

PROOF. By the Cauchy-Schwartz inequality

$$|w_n \cdot v_n - w \cdot v| \le |w_n \cdot v_n - w_n \cdot v| + |w_n \cdot v - w \cdot v| \le \|w_n\|\,\|v_n - v\| + \|v\|\,\|w_n - w\|.$$

Since w_n is bounded (see Observation 1.5), the left hand side tends to zero.

Proposition 1.2 (Existence and uniqueness of the orthogonal projection) Let H be a Hilbert space, S a closed subspace of H and $v \in H$.
(a) There exists an orthogonal projection of v on S, i.e. a vector w such that (1) $w \in S$ and (2) $v - w \perp S$ ($v - w$ is orthogonal to all vectors of S). If w' fulfills (1) and (2) then $w' = w$ (the orthogonal projection is unique). The orthogonal projection of v on S is denoted by $\mathrm{Proj}(v|S)$.
(b) The square distance $|v - \mathrm{Proj}(v|S)|^2$ minimizes the function $|v - y|^2$, for $y \in S$. Conversely, if z minimizes $|v - y|^2$ for $y \in S$, then $z = \mathrm{Proj}(v|S)$.

PROOF. We proceeds by steps.
Step 1. If w and w' are orthogonal projections of v on S, then $w = w'$. For, we have $v = w + R$ and $v = w' + R'$, with $R \perp S$ and $R' \perp S$. We have

$$(w - w') + (R - R').$$

On the other hand, $w - w' \in S$, while $(R - R') \perp S$ (prove), so that $(R - R') \perp (w - w')$. Therefore

$$\|w - w'\|^2 + \|R - R'\|^2 = 0,$$

so that $w = w'$. Note that we have not yet proved that the orthogonal projection exists.

Step 2. If w is an orthogonal projection of v on S, then

$$\|v - w\|^2 = \min_{y \in S} \|v - y\|^2 \tag{1.15}$$

and, conversely, if w fulfills (1.15) then it is an orthogonal projection of v on S. Suppose that w is an orthogonal projection of v on S. For any $y \in S$ we have $w - y \in S$ and therefore $(w - y) \perp (v - w)$, so that

$$\|v - y\|^2 = \|v - w\|^2 + \|w - y\|^2.$$

Thus $\|v - y\|^2 \geq \|v - w\|^2$. Conversely, let w fulfill (1.15). For $y \in S$, project $v - w$ on y (this is a one-dimensional projection):

$$v - w = \alpha y + \rho, \quad \alpha = \frac{(v - w) \cdot y}{y \cdot y}.$$

If $\alpha \neq 0$ then $\|\rho\|^2 < \|v - w\|^2$, so that $\|v - w - \alpha y\|^2 < \|v - w\|^2$, while $w + \alpha y \in S$. Thus $\alpha = 0$ for any $y \in S$, that is $(v - w) \cdot y = 0$.

Step 3. Step 1 and 2 are valid irrespective of whether S is closed or not. We have not proved that an orthogonal projection exists but only that *if* it exists then it is unique and solves (1.15). Now, under the assumption that S is closed we prove existence. Let

$$b = \inf_{y \in S} \|v - y\|^2.$$

Note that such greatest lower bound exists and is finite irrespective of whether S is closed or not (finiteness is a consequence of $\|v - y\|^2 \geq 0$ for all y). For any $n \in \mathbb{N}$, there exists $y_n \in S$ such that $b \leq \|v - y_n\|^2 < b + 1/n$. Now consider the projection

$$v = \alpha(y_n - y_m) + \rho_{nm} = Y_{nm} + \rho_{nm}.$$

We have:

$$\|y_n - y_m\| \leq \|y_n - Y_{nm}\| + \|y_m - Y_{nm}\|. \tag{1.16}$$

On the other hand,

$$\|y_n - Y_{nm}\|^2 + \|v - Y_{nm}\|^2 = \|v - y_n\|^2.$$

Since $Y_{nm} \in S$, this implies that

$$\|y_n - Y_{nm}\|^2 \leq -b + \|v - y_n\|^2.$$

Using this inequality, the analogous for $\|y_m - Y_{nm}\|^2$, and (1.16), we obtain that given ϵ, there exists $n_\epsilon > 0$ such that for $n, m > n_\epsilon$, $\|y_n - y_m\|^2 < \epsilon$, that is the sequence y_n is Cauchy.

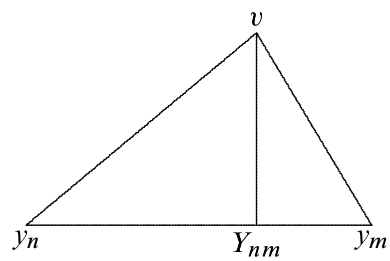

FIGURE 1.1

Figure 1.1 illustrates what we have just proved, the only finesse of the demonstration. The length of both sides $\overline{v\, y_n}$ and $\overline{v\, y_m}$ become almost equal to b as n and m become large. On the other hand the length of $\overline{v\, Y_{nm}}$ cannot become smaller that b, because $Y_{nm} \in S$. Thus y_n and y_m must approach one another.

Since y_n is Cauchy there exists w such that $y_n \to w$. Since S is closed, $w \in S$. Moreover, the definition of y_n and continuity of the norm implies that $\|v - w\| = b$. Thus $\|v - y\|$ reaches a minimum when $y = w$. By the result of Step 2, w is the orthogonal projection and the proof is complete.

Observation 1.12 If $v \in S$ then $b = 0$ and $v = \text{Proj}(v|S)$.

Exercise 1.14 Consider the subspaces $A \subset L^2([0\ 1])$ and $B \subset l^2(-\infty, \infty)$, defined in Exercise 1.13. Both are closed. Given $f \in L^2([0\ 1])$ and $a \in l^2(-\infty, \infty)$, determine their projections on A and B respectively.

Exercise 1.15 Consider the subspace $\ell^2 \in l^2(-\infty, \infty)$ defined in Exercise 1.11. Let $a \in l^2(-\infty, \infty)$. Show that

$$\inf_{y \in \ell^2} \|a - y\|^2 = 0,$$

and that the projection of a on ℓ^2 does not exist unless $a \in \ell^2$.

Very often S will be the minimum closed subspace containing a sequence of vectors $\{v_j,\ j \in \mathbb{N}\}$, or $\{v_j,\ j \in \mathbb{Z}\}$.

Definition 1.5 Given the subset $G \subseteq H$, the closed span of G, denoted by $\overline{\text{sp}}(G)$, is the intersection of all closed subspaces of H containing G (prove that such intersection is itself a closed subspace). When $S = \overline{\text{sp}}(G)$ we say that S is *generated* by G.The closed span can also be obtained "from inside" by taking, firstly, all linear combinations of elements of G, call it $\text{sp}(G)$, and then all the limits of convergent sequences of elements of $\text{sp}(G)$. Show that $\text{sp}(\ell^2) = \ell^2$ and $\overline{\text{sp}}(\ell^2) = l^2(-\infty, \infty)$, where ℓ^2 is defined in Exercise 1.11.

Observation 1.13 If the subset G is finite then obviously $\text{sp}(G) = \overline{\text{sp}}(G)$.

Exercise 1.16 Prove the assertion in Definition 1.5, that adding to $\text{sp}(G)$ the limits of convergent sequences in $\text{sp}(G)$, call $\overline{\text{sp}}(G)$ the resulting set, one has $\overline{\text{sp}}(G) = \overline{\text{sp}}(G)$. What must be proved is that $\overline{\text{sp}}(G)$ is closed, i.e. that if $\{g_k,\ k \in \mathbb{N}\}$ is a converging sequence of elements in $\overline{\text{sp}}(G)$, then the limit belongs to $\overline{\text{sp}}(G)$. Hint: by definition, the elements g_k can be approximated by elements of $\text{sp}(G)$.

Proposition 1.3 Now let $S = \overline{\text{sp}}(v_j,\ j \in \mathbb{N})$. Then

$$\text{Proj}(v|S) = \lim_{n\to\infty} \text{Proj}(v|v_1,\ v_2,\ \ldots,\ v_n). \tag{1.17}$$

A proof of Proposition 1.3 is fairly obvious. The reader can find it by proving that the distance between v and the partial projection tends to the distance between v and S. The proof that the partial projections are a Cauchy sequence goes exactly as in the proof of Proposition 1.2, Step 3.

Note that in general

$$\text{Proj}(v|v_1,\ v_2,\ \ldots,\ v_n) = a_{n1}v_1 + a_{n2}v_2 + \cdots + a_{nn}v_n,$$

with the coefficients depending on n (if we add a "regressor" the coefficients of the "previous regressors" change). However, if the vectors v_j are *mutually orthogonal*, i.e. if $v_i \cdot v_j = 0$ for $i \neq j$, then $a_{ns} = \frac{v \cdot v_s}{v_s \cdot v_s}$, and is therefore independent of n (this is very easy to prove, and is formally identical to the statement that if the regressors are orthogonal, so that the variance-covariance matrix is diagonal, then the coefficients of the regression can be computed one at a time). Thus, if $v_i \perp v_j$, for $i \neq j$, then

$$\text{Proj}(v|S) = \lim_{n\to\infty} \sum_{j=1}^{n} a_j v_j = \sum_{j=1}^{\infty} a_j v_j = \sum_{j=1}^{\infty} \frac{v \cdot v_j}{v_j \cdot v_j} v_j. \tag{1.18}$$

If the sequence of the v's is bilateral, i.e. $v_j,\ j \in \mathbb{Z}$ (but this is only a matter of representation), then

$$\text{Proj}(v|S) = \lim_{n\to\infty} \sum_{j=-n}^{n} a_j v_j = \sum_{j=-\infty}^{\infty} a_j v_j = \sum_{j=-\infty}^{\infty} \frac{v \cdot v_j}{v_j \cdot v_j} v_j. \tag{1.19}$$

The following proposition is very important and easy to prove.

Proposition 1.4 Let H be a Hilbert space and let $\{v_k,\ k \in \mathbb{Z}\}$ be an *orthonormal sequence*, i.e. suppose that $v_k \perp v_j$ for $k \neq j$ and that $\|v_k\| = 1$ for all k. Let $M = \overline{\mathrm{sp}}(\{v_k,\ k \in \mathbb{Z}\})$. Then:
(1) If $v \in M$,

$$v = \sum_{k=-\infty}^{\infty} (v \cdot v_k) v_k, \quad \|v\|^2 = \sum_{k=-\infty}^{\infty} (v \cdot v_k)^2. \tag{1.20}$$

The first equality (1.20) and the coefficients $v \cdot v_k$ are called the Fourier expansion and the Fourier coefficients of v respectively. If v and w have the same Fourier expansion then $v = w$.
(2) If v and w belong to M,

$$v \cdot w = \sum_{k=-\infty}^{\infty} (v \cdot v_k)(w \cdot v_k). \tag{1.21}$$

(3) If $\{c_k,\ k \in \mathbb{Z}\}$ is a sequence of complex numbers such that $\sum_{k=-\infty}^{\infty} |c_k|^2 < \infty$, then $\sum_{k=-m}^{m} c_k v_k$ converges. Calling z the limit, $c_k = z \cdot v_k$, so that

$$z = \sum_{k=-\infty}^{\infty} c_k v_k$$

coincides with the expansion (1.20) for z.

PROOF. Since $v \in M$, $v = \mathrm{Proj}(v|M)$ (see Observation 1.12), so that the first equality in (1.20) is a consequence of (1.19). The second equality follows from orthonormality of the vectors v_k and the continuity of the norm (Proposition 1.1). If v and w have the same Fourier expansions they are limits of the same sequence of partial sums and are therefore equal. Continuity of the inner product (Proposition 1.1) implies that

$$v \cdot w = \lim_{m \to \infty} \left(\sum_{k=-m}^{m} (v \cdot v_k) v_k \right) \left(\sum_{k=-m}^{m} (w \cdot v_k) v_k \right) = \lim_{m \to \infty} \sum_{k=-m}^{m} (v \cdot v_k)(w \cdot v_k).$$

Statement (2) follows. For (3), observe that the Cauchy condition for the sequence $\sum_{k=-m}^{m} c_k v_k$ is identical to the Cauchy condition for the sequence $\sum_{k=-m}^{m} c_k^2$, which is convergent by assumption. Moreover, using again continuity of the inner product,

$$z \cdot v_k = \lim_{m \to \infty} \sum_{h=-m}^{m} [(c_h v_h) \cdot v_k] = c_k. \tag{1.22}$$

Observation 1.14 If $H = L^2([0\ 1])$ then (1.22) takes the form

$$\int_0^1 \left[\sum_{h=-\infty}^{\infty} c_h v_h(x) \right] v_k(x) dx = \sum_{h=-\infty}^{\infty} \int_0^1 c_h v_h(x) v_k(x) dx = c_k.$$

More in general, given the sequence $f_n \in L^2([0\ 1])$, for $n \in \mathbb{Z}$, suppose that in mean-square $\lim_{m\to\infty} \sum_{h=-m}^{m} f_k = f$, i.e.

$$f = \sum_{h=-\infty}^{\infty} f_h.$$

Then, continuity of the inner product implies

$$\int_0^1 \left[\sum_{h=-\infty}^{\infty} f_h(x) \right] g(x) dx = \sum_{h=-\infty}^{\infty} \int_0^1 f_h(x) g(x) dx,$$

for any $g \in L^2([0\ 1])$. In particular, taking $g = 1$ (which belongs to $L^2([0\ 1])$ because the measure of [0 1] is finite),

$$\int_0^1 \sum_{h=-\infty}^{\infty} f_h(x) dx = \sum_{h=-\infty}^{\infty} \int_0^1 f_h(x) dx.$$

The reader should keep in mind that this term-by-term integration is possible because the series $\sum f_h$ converges in mean square, and because the measure of Ω is finite. Term-by-term integration is also possible if $\sum_{h=-m}^{m} f_h$ converges almost everywhere and $|\sum_{h=-m}^{m} f_h| \le M$ for all m (this is a particular case of the Lebesgue bounded convergence theorem), whereas convergence a.e. alone does not ensure term-by-term integration, as the following example shows:

$$g_h(x) = \begin{cases} 0 \text{ if } 0 \le x \le 1 - 1/h \text{ or } x = 1 \\ s_h \text{ if } 1 - 1/h < x < 0 \end{cases}$$

where s_h is a monotonically increasing sequence. Then set $f_h = 0$ for $h \le 0$, $f_1 = g_1$ and $f_h = g_h - g_{h-1}$ for $h > 1$, so that $\sum_{h=-m}^{m} f_h = g_m$. We have:

$$\int_0^1 \sum_{h=-m}^{m} f_h(x) dx = \int_0^1 g_m(x) dx = s_m/m,$$

whose limit can be zero, a positive number or $+\infty$, although g_m converges to zero everywhere. Note the symmetry between this example and Example 1.9. In that case convergence of an integral does not imply even the weaker pointwise convergence, here convergence everywhere does not ensure convergence of an integral.

Exercise 1.17 Assume that the sequence f_n and f belong to $L^2([0\ 1])$ and that f_n converges to f uniformly. Prove that f_n converges to f in mean square. Would this result hold in $L^2(\mathbb{R})$?

Exercise 1.18 Assume that the vectors of the family $\{v_j,\ j \in \mathbb{N}\}$ are mutually orthogonal and that there exists a positive real ρ such that $||v_j|| \geq \rho$ for all j. Given $v \in H$, define the coefficients a_j as in (1.18), i.e. $a_j = v \cdot v_j / ||v_j||^2$. Prove that

$$\lim_{j \to \infty} a_j = 0 \tag{1.23}$$

(use the argument used to prove the first part of Proposition 1.4 to show that $\sum a_j^2 < \infty$). Obviously for a bilateral sequence we have

$$\lim_{|j| \to \infty} a_j = 0. \tag{1.24}$$

In particular, the Fourier coefficients converge to zero as $|j| \to \infty$.

Part of Proposition 1.4 can be restated as follows:

Proposition 1.5 Under the assumptions of Proposition 1.4, the map $\Phi : M \to l^2(-\infty, \infty)$ defined as

$$v \to \{v \cdot v_k,\ k \in \mathbb{Z}\},$$

i.e. mapping each vector of M to the sequence of its Fourier coefficients, is an *isomorphism*, namely (1) $\Phi(v + w) = H(v) + H(w)$, (2) $\Phi(av) = a\Phi(v)$, (3) if $v \neq w$ then $\Phi(v) \neq \Phi(w)$, (4) for each $a \in l^2(-\infty, \infty)$ there exists $v \in M$ such that $H\Phi(v) = a$, (5) $\Phi(v) \cdot \Phi(w) = v \cdot w$. Obviously (5) implies that (6) $\|\Phi(v)\| = \|v\|$. Lastly, (6) implies that if $v_n \to v$ then $\Phi(v_n) \to \Phi(v)$, i.e. that (7) Φ is continuous.

Summary. Weakly stationary stochastic processes are best analyzed (as we will see) within the vector space $L^2(\Omega, \mathcal{F}, P)$. The latter is endowed with an inner product, a norm and a distance. It shares many properties of the Euclidean space $\mathbb{R}^n$, but is in general infinite dimensional. We have given examples in which intuition based on $\mathbb{R}^n$ fails to generalize to infinite-dimensional spaces. In particular, completeness, which is an elementary property of $\mathbb{R}^n$, holds for $L^2(\Omega, \mathcal{F}, P)$ (we have not given this proof), but not for all its vector subspaces (it holds only for closed subspaces). We have proved the orthogonal projection theorem for Hilbert spaces, which include spaces $L^2(\Omega, \mathcal{F}, P)$, stating that there exists a (unique) orthogonal projection of a vector v on a closed subspace S. Lastly, if S is generated by a sequence of orthogonal vectors, the orthogonal projection takes the form of a series.