by the Cauchy-Schwarz inequality, and the right-hand side of (13) converges to 0. Q.E.D.

COROLLARY 7.6.2. *If $x^{(n)} \to x$, then $\|x^{(n)}\| \to \|x\|$.*

When we take (x, y) to be $\mathscr{E}xy$ in the Hilbert space generated by $\{y_t\}$ Theorem 7.6.4 justifies the limits of variances and covariances in Sections 7.4 and 7.5.

7.6.2. Projections and Linear Prediction

In terms of the stochastic process $\{y_t\}$ with mean $\mathscr{E}y_t = 0$ and covariance function $\mathscr{E}y_t y_s = \sigma(t-s)$, we may be interested in predicting y_t from knowledge of $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ (a part of the past) or from $y_{t-1}, y_{t-2}, \ldots$ (the entire past). If $\hat{y}_t$ is the prediction, we shall take $\mathscr{E}(y_t - \hat{y}_t)^2$ as the criterion to be made small. This mean square error of prediction is minimized if $\hat{y}_t$ is taken as the conditional expectation of y_t given $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ or $y_{t-1}, y_{t-2}, \ldots$ as the case may be. This conditional expectation will, of course, depend on the distribution of $y_t, y_{t-1}, \ldots, y_{t-p}$ or the probability measure of $y_t, y_{t-1}, \ldots$. If the distribution is normal or the process is Gaussian, the conditional expectation will be a linear function of $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ or $y_{t-1}, y_{t-2}, \ldots$, but in other instances it may not be a linear function.

We shall consider prediction of y_t by a linear function, $\hat{y}_t = \sum_{s=1}^{p} c_s y_{t-s}$ in the case of prediction from $y_{t-1}, \ldots, y_{t-p}$ or by a limit $\hat{y}_t$ of such linear functions. Then $\mathscr{E}(y_t - \hat{y}_t)^2$ depends only on $\{\sigma(h)\}$. It will be convenient to put this problem into the geometry of Hilbert space.

A *linear manifold* is a nonempty subset of the Hilbert space such that for every real α and β $\alpha x + \beta y$ is in the subset if x and y are. The linear manifold is *closed* if it contains the limit of every Cauchy sequence in it. A finite or infinite set of points in the Hilbert space may generate a closed linear manifold, which then consists of all finite linear combinations of these points and their limits; the original set of points is said to span the closed linear manifold. A closed linear manifold is a Hilbert space, and will be called a *subspace*.

Since $\mathscr{E}(y_t - \hat{y}_t)^2$ is $\|y_t - \hat{y}_t\|^2$, finding the best linear prediction of y_t is the problem of finding the point in the linear manifold spanned by the y_s's $(s < t)$ which is closest (in the distance defined by the norm) to the point y_t. For a T-dimensional Euclidean space with norm $\boldsymbol{x}'\boldsymbol{x}$, this problem was solved in Section 2.2. The vector in the linear manifold spanned by the p columns of $\boldsymbol{z}$ closest to $\boldsymbol{y}$ was a uniquely defined vector $\boldsymbol{z}$ such that $\boldsymbol{y} - \boldsymbol{z}$ was orthogonal to the p columns of $\boldsymbol{z}$ (and hence to every vector in the linear manifold). We shall now show that this solution holds in a general Hilbert space.

In terms of the inner product (x, y) two elements x and y are orthogonal if $(x, y) = 0$. We may write $x \perp y$.

THEOREM 7.6.5. *(Projection Theorem) If $\mathcal{M}$ is a subspace of a Hilbert space $\mathcal{H}$ and y is an element in $\mathcal{H}$, there exists a unique element x in $\mathcal{M}$ such that $\|y - x\| = \min_{z\in\mathcal{M}} \|y - z\|$; if y is not in $\mathcal{M}$, $y - x$ is orthogonal to every $z \in \mathcal{M}$.*

PROOF. Let $\min_{z\in\mathcal{M}} \|y - z\| = d$ and let $\{x^{(n)}\}$ be a sequence such that $\|y - x^{(n)}\| \to d$. (Such a sequence exists by the definition of the minimum.) The existence of the element x will follow from the demonstration that $\{x^{(n)}\}$ is a Cauchy sequence. We have

$$\text{(14)} \quad \|(y - x^{(n)}) - (y - x^{(m)})\|^2 + \|(y - x^{(n)}) + (y - x^{(m)})\|^2 = 2\|y - x^{(n)}\|^2 + 2\|y - x^{(m)}\|^2.$$

The left-hand side is

$$\text{(15)} \qquad \|x^{(m)} - x^{(n)}\|^2 + 4\|y - \tfrac{1}{2}(x^{(n)} + x^{(m)})\|^2;$$

since $\frac{1}{2}(x^{(n)} + x^{(m)}) \in \mathcal{M}$, the second term in (15) is at least $4d^2$. Thus (14) implies

$$\text{(16)} \qquad \|x^{(m)} - x^{(n)}\|^2 \leq 2\|y - x^{(n)}\|^2 + 2\|y - x^{(m)}\|^2 - 4d^2,$$

and the right-hand side of (16) converges to 0. Thus $\{x^{(n)}\}$ is a Cauchy sequence and a limit x exists; by Corollary 7.6.2 $\|y - x\| = d$.

For an arbitrary element $z \in \mathcal{M}$

$$\text{(17)} \qquad d^2 \leq \|y - x - \alpha z\|^2 = \|y - x\|^2 - 2(y - x, \alpha z) + \|\alpha z\|^2 = d^2 - 2\alpha(y - x, z) + \alpha^2\|z\|^2.$$

If $y \notin \mathcal{M}$ and hence $y - x \neq 0$, the inequality can hold for all α only if $(y - x, z) = 0$; that is, if $y - x$ is orthogonal to z.

If there were another element x' such that $\|y - x'\|^2 = d^2$, then $(y - x', z) = 0$ for every $z \in \mathcal{M}$; for $z = x - x'$ we would have $0 = (y - x', x - x') - (y - x, x - x') = (x - x', x - x') = \|x - x'\|^2$, which implies $x = x'$. Q.E.D.

The point x is called the *projection* of y on the subspace $\mathcal{M}$ and $y - x$ shall be termed the *residual* of y from $\mathcal{M}$.

If $\mathcal{M}$ is a subspace of a Hilbert space $\mathcal{H}$, then the set of all elements orthogonal to $\mathcal{M}$ is also a subspace, say $\mathcal{M}^\perp$. (It is closed by continuity of the inner product and completeness of $\mathcal{H}$.) Every element y in $\mathcal{H}$ can be written uniquely as $x + z$, where $x \in \mathcal{M}$ and $z \in \mathcal{M}^\perp$. We write this as $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$.

In terms of Hilbert space linear prediction is simply a projection. The Hilbert space $\mathscr{H}$ is the closed linear manifold spanned by $y_t, y_{t-1}, \ldots,$ and the inner product is $\mathscr{E}xy$. Let $\mathscr{M}_{t-1,p}$ be the subspace spanned by $y_{t-1}, \ldots, y_{t-p}$. The best prediction of y_t by a linear combination $c_1y_{t-1} + \cdots + c_py_{t-p}$ is the projection of y_t on $\mathscr{M}_{t-1,p}$, say $\hat{y}_{t,p}$; it minimizes

$$\mathscr{E}\left(y_t - \sum_{j=1}^{p} c_j y_{t-j}\right)^2 = \sigma(0) - 2\sum_{j=1}^{p} c_j\sigma(j) + \sum_{i,j=1}^{p} c_ic_j\sigma(j-i). \tag{18}$$

The problem is analogous to that treated in Chapter 2. The solution is given by the normal equations

$$\begin{bmatrix} \sigma(0) & \sigma(1) & \cdots & \sigma(p-1) \\ \sigma(1) & \sigma(0) & \cdots & \sigma(p-2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma(p-1) & \sigma(p-2) & \cdots & \sigma(0) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_p \end{bmatrix} = \begin{bmatrix} \sigma(1) \\ \sigma(2) \\ \cdot \\ \cdot \\ \cdot \\ \sigma(p) \end{bmatrix}. \tag{19}$$

Let $\mathscr{M}_{t-1}$ be the closed linear manifold spanned by $y_{t-1}, y_{t-2}, \ldots$. The best linear prediction of y_t by $y_{t-1}, y_{t-2}, \ldots$ is the projection of y_t on $\mathscr{M}_{t-1}$, say $\hat{y}_t$.

THEOREM 7.6.6.

$$\lim_{p\to\infty} \mathscr{E}(\hat{y}_t - \hat{y}_{t,p})^2 = 0. \tag{20}$$

PROOF. Let $\mathscr{E}(y_t - \hat{y}_{t,p})^2 = d_p$, and let $\mathscr{E}(y_t - \hat{y}_t)^2 = d$. Since $\mathscr{M}_{t-1,p} \subset \mathscr{M}_{t-1,p+1}$ and $\mathscr{M}_{t-1} = \bigcup_{p=1}^{\infty} \mathscr{M}_{t-1,p}$, there is a sequence $\{x^{(p)}\}$ with $x^{(p)} \in \mathscr{M}_{t-1,p}$ such that $\mathscr{E}(x^{(p)} - \hat{y}_t)^2 \to 0$ as $p \to \infty$. Since $\mathscr{E}(x^{(p)} - y_t)^2 \geq d_p \geq d$ and

$$\mathscr{E}(x^{(p)} - y_t)^2 \leq \mathscr{E}(x^{(p)} - \hat{y}_t)^2 + \mathscr{E}(\hat{y}_t - y_t)^2, \tag{21}$$

which tends to d, we have

$$\lim_{p\to\infty} \mathscr{E}(x^{(p)} - y_t)^2 = \lim_{p\to\infty} d_p = d. \tag{22}$$

We can also write

$$\mathscr{E}(x^{(p)} - y_t)^2 = \mathscr{E}(x^{(p)} - \hat{y}_{t,p})^2 + \mathscr{E}(\hat{y}_{t,p} - y_t)^2 \tag{23}$$

because $y_t - \hat{y}_{t,p}$ is orthogonal to every vector in $\mathscr{M}_{t,p}$. This implies $\mathscr{E}(x^{(p)} - \hat{y}_{t,p})^2 \to 0$ inasmuch as the other two terms in (23) approach d. The theorem follows from

$$\mathscr{E}(\hat{y}_t - \hat{y}_{t,p})^2 \leq \mathscr{E}(\hat{y}_t - x^{(p)})^2 + \mathscr{E}(x^{(p)} - \hat{y}_{t,p})^2. \tag{24}$$

This theorem shows that the prediction based on the finite past approximates the prediction based on the infinite past. In practice the statistician would like a large sample to estimate the covariance structure as well as for prediction on the basis of this estimated covariance structure.

The variance of the residual may be any nonnegative number not exceeding $\sigma(0)$. If $\mathscr{E}(y_t - \hat{y}_t)^2 = \sigma^2 = 0$, the process is called *deterministic*. This means that y_t can be predicted perfectly by an element of $\mathscr{M}_{t-1}$. In most cases this element is an infinite linear combination $\sum_{s=1}^{\infty} c_s y_{t-s}$. If $\mathscr{E}(y_t - \hat{y}_t)^2 = \sigma^2 > 0$, the process is called *regular*.

Since $y_t - \hat{y}_t$ is orthogonal to every point in $\mathscr{M}_{t-1}$, we can write

$$y_t = \hat{y}_t + u_t, \tag{25}$$

where $u_t = y_t - \hat{y}_t$ is uncorrelated with $y_{t-1}, y_{t-2}, \ldots$. The random variable u_t may be called the *innovation* or disturbance.

7.6.3. The Wold Decomposition

The so-called Wold decomposition [Wold (1954)] clarifies the structure of a stationary process.

THEOREM 7.6.7. *If $\{y_t\}$ is a regular stationary stochastic process with $\mathscr{E}y_t = 0$, it can be written as*

$$y_t = \sum_{s=0}^{\infty} \gamma_s u_{t-s} + v_t, \tag{26}$$

where $\sum_{s=0}^{\infty} \gamma_s^2 < \infty$, $\gamma_0 = 1$, $\mathscr{E}u_s = \mathscr{E}v_s = 0$, $\mathscr{E}u_s^2 = \sigma^2$, $\mathscr{E}u_s u_t = 0$, $s \neq t$, and $\mathscr{E}u_s v_t = 0$, $u_s \in \mathscr{M}_s$ and $v_t \in \mathscr{M}_{-\infty} = \bigcap_{s=0}^{\infty} \mathscr{M}_{t-s}$. The sequences $\{\gamma_s\}$ and $\{u_s\}$ are unique.

PROOF. Let $u_s = y_s - \hat{y}_s$, $s = t, t-1, \ldots$, and $\sigma^2 = \mathscr{E}u_s^2$. Since u_s is orthogonal to $\mathscr{M}_{s-1}$, $\mathscr{E}u_s u_r = 0$, $r < s$. Let $\gamma_s = \mathscr{E}y_t u_{t-s}/\sigma^2$, $s = 1, 2, \ldots$. Then

$$0 \leq \mathscr{E}\left(y_t - \sum_{s=0}^{m} \gamma_s u_{t-s}\right)^2 = \mathscr{E}y_t^2 - \sigma^2 \sum_{s=0}^{m} \gamma_s^2, \tag{27}$$

and hence $\sum_{s=0}^{\infty} \gamma_s^2 < \infty$. Therefore $\sum_{s=0}^{\infty} \gamma_s u_{t-s}$ converges in the mean and is in the subspace spanned by $u_t, u_{t-1}, \ldots$. Define v_t by

$$v_t = y_t - \sum_{s=0}^{\infty} \gamma_s u_{t-s}. \tag{28}$$

Then $\mathscr{E}v_t u_r = \mathscr{E}y_t u_r - \sigma^2 \gamma_{t-r} = 0$ for $r \leq t$; $\mathscr{E}v_t u_r = 0$ for $r > t$ because u_r is orthogonal to all points in $\mathscr{M}_t$ and $v_t \in \mathscr{M}_t$. Since v_t is orthogonal to u_t, it is in

$\mathcal{M}_{t-1}$. By induction it is in $\mathcal{M}_s$ for $s \leq t$ and hence in $\mathcal{M}_{-\infty} = \bigcap_{s=0}^{\infty} \mathcal{M}_{t-s}$. The resolution is unique because $u_s \in \mathcal{M}_s$ and u_s is orthogonal to $\mathcal{M}_{s-1}$ and $\gamma_s = \mathscr{E} y_t u_{t-s}/\sigma^2$. Q.E.D.

THEOREM 7.6.8. *The process $\{v_t\}$ defined in Theorem 7.6.7 is deterministic, and the process $\{w_t\}$, where $w_t = \sum_{s=0}^{\infty} \gamma_s u_{t-s}$, is regular.*

PROOF. To prove $\{v_t\}$ is deterministic we shall show that the closed linear manifold spanned by $v_t, v_{t-1}, \ldots$, say $\mathcal{M}_{tv}$, is $\mathcal{M}_{-\infty}$; that is, $\mathcal{M}_{tv} = \mathcal{M}_{t-1,v} = \cdots = \mathcal{M}_{-\infty}$; then v_t is a (possibly infinite) linear combination of $v_{t-1}, v_{t-2}, \ldots$.

LEMMA 7.6.1.

$$\mathcal{M}_{-\infty} = \mathcal{M}_u^{\perp}. \tag{29}$$

PROOF. Here $\mathcal{M}_u^{\perp}$ is the subspace orthogonal to $\mathcal{M}_u$, the subspace spanned by $\{u_t\}$. If $x \in \mathcal{M}_{-\infty} = \bigcap_{t=-\infty}^{\infty} \mathcal{M}_t$, then $x \in \mathcal{M}_t$ and hence is orthogonal to u_{t+1} for every t; $x \in \mathcal{M}_u^{\perp}$. Conversely, suppose $x \in \mathcal{M}_u^{\perp}$. Since $\mathscr{H} = \bigcup_{t=-\infty}^{\infty} \mathcal{M}_t$, $x \in \mathcal{M}_t$ for some t. Because $x \perp u_t$, $x \in \mathcal{M}_{t-1}$, and by induction $x \in \mathcal{M}_s$ for $s \leq t$. Moreover, $x \in \mathcal{M}_s$ for $s > t$ because $\mathcal{M}_t \subseteq \mathcal{M}_s$. This proves the lemma.

LEMMA 7.6.2.

$$\mathcal{M}_{tu} = \mathcal{M}_{tw}. \tag{30}$$

PROOF. Since $w_t = \sum_{s=0}^{\infty} \gamma_s u_{t-s}$, the subspace $\mathcal{M}_{tw}$ spanned by $w_t, w_{t-1}, \ldots$ is contained in the subspace $\mathcal{M}_{tu}$ spanned by $u_t, u_{t-1}, \ldots$. Conversely, $u_t \in \mathcal{M}_t = \mathcal{M}_{tw} \oplus \mathcal{M}_{tv}$ and $u_t \perp \mathcal{M}_{tv}$ and therefore $u_t \in \mathcal{M}_{tw}$. This proves the lemma.

We now complete the proof of Theorem 7.6.8. Since $v_s \in \mathcal{M}_{-\infty}$ for every s, $\mathcal{M}_{tv} \subseteq \mathcal{M}_{-\infty}$. Since $\mathcal{M}_{-\infty} = \bigcap_{t=-\infty}^{\infty} \mathcal{M}_t$, $x \in \mathcal{M}_{-\infty}$ implies $x \in \mathcal{M}_t$; since $x \perp \mathcal{M}_{tu} = \mathcal{M}_{tw}$, $x \in \mathcal{M}_{tv}$. This shows $\mathcal{M}_{tv} = \mathcal{M}_{-\infty}$ and $\{v_t\}$ is deterministic. Since $w_t = u_t + \sum_{s=1}^{\infty} \gamma_s u_{t-s}$ and $u_t \perp \sum_{s=1}^{\infty} \gamma_s u_{t-s} \in \mathcal{M}_{t-1,u}$, $\mathscr{E}(w_t - \hat{w}_t)^2 = \sigma^2 > 0$. Thus $\{w_t\}$ is regular. Q.E.D.

The process $\{v_t\}$ is often called purely deterministic. The process $\{w_t\}$ is called *purely indeterministic* because it does not have a deterministic component. Since $y_t = u_t + \sum_{s=1}^{\infty} \gamma_s u_{t-s} + v_t$ and $u_t \perp (\sum_{s=1}^{\infty} \gamma_s u_{t-s} + v_t) \in \mathcal{M}_{t-1}$, the projection on $\mathcal{M}_{t-1}$ is $\hat{y}_t = \sum_{s=1}^{\infty} \gamma_s u_{t-s} + v_t$; this is the best linear prediction of y_t and $\mathscr{E}(y_t - \hat{y}_t)^2 = \mathscr{E} u_t^2 = \sigma^2$. By Lemma 7.6.2 $\hat{y}_t \in \mathcal{M}_{t-1,w} \oplus \mathcal{M}_{t-1,v} = \mathcal{M}_{t-1}$ and hence $\hat{y}_t$ is implicitly a linear function of $y_{t-1}, y_{t-2}, \ldots$, or a limit of such.