3. Orthogonal projection. Conditional expectation in the wide sense

Let \((X_n)_{n \geq 1}\) be a sequence of random variables with
\[ EX_n^2 = \sigma_n^2 \quad \text{and} \quad EX_n \equiv 0. \]

If
\[ EX_kX_j = \begin{cases} \sigma_k^2, & k = j \\ 0, & \text{otherwise,} \end{cases} \]
\((X_n)_{n \geq 1}\) is the sequence with orthogonal elements. For \(\sigma_k^2 \equiv 1\), this sequence forms white noise.

If random variables with finite second moments are not orthogonal, they might be orthogonalize by a specific orthogonalization procedure. We start with an example.

**Example 1.** Let \(Y\) and \(Z\) be random variables with \(EY = 0, EZ = 0\) and \(EY^2 < \infty, EZ^2 < \infty\). We introduce a projection of \(X\) on a pair \((1, Z)\) by taking a linear combination of 1 and \(Z\):
\[ \hat{Y} = c_1 + c_2Z. \]
If constants \(c_1, c_2\) can be chosen such that \(E(Y - \hat{Y})Z = 0\) the projection \(\hat{Y}\) is named the orthogonal projection. Under \(EZ^2 > 0\), a direct verification shows that the choice \(c_1 = 0\) and \(c_2 = \frac{EYZ}{EZ^2}\) gives the orthogonal projection
\[ \hat{Y} = \frac{EYZ}{EZ^2} Z. \] (3.1)

Consider now the next setting. Let \(Y, Z_1, \ldots, Z_n\) be random variables with finite second moments: \(EY^2 < \infty\) and \(EZ_k^2 < \infty, k = 1, \ldots, n\) and introduce a linear space generated by \(1, Z_1, \ldots, Z_n\):
\[ \mathcal{M} = \left\{ \eta : \eta = \text{linear combination of } 1, Z_1, \ldots, Z_n \right\}; \]
if \(1, Z_1, \ldots, Z_n\) is replaced by \(1, Z_1, \ldots, Z_n, \ldots\), \(\mathcal{M}\) contains not only linear combination of \(1, Z_1, \ldots, Z_n, \ldots\) but also their limits in \(L^2\)-norm.

The orthogonal projection \(\hat{Y}\) of \(Y\) on \(\mathcal{M}\) is an element of \(\mathcal{M}\) such that \(E\eta(Y - \hat{Y}) = 0\) for any \(\eta \in \mathcal{M}\); symbolically
\[ Y - \hat{Y} \perp \mathcal{M}. \]

3.1. Computation of the orthogonal projection. Set
- \(m = EY\);
- \(m_k = EZ_k, k = 1, \ldots, n\);
- \(\mathcal{Z}\) vector(column) with entries \(Z_1, \ldots, Z_n\)
- \(\mathcal{m}\) vector(column) with entries \(m_1, \ldots, m_n\)
- \(\text{var}(Y) = E(Y - m)^2\);
- \(\text{cov}(Z, Z) = E(Z - m)^2(Z - m)^T (T\) is the transposition symbol\);
- \(\text{cov}(Y, Z) = E(Y - m^y)(Z - m)^T.\)
Theorem 3.1. Under nonsingular matrix $\text{cov}(Z, Z)$, the orthogonal projection of $Y$ on $\mathcal{M}$ exists and is defined as follows:

$$\hat{Y} = m + \text{cov}(Y, Z)\text{cov}^{-1}(Z, Z)(Z - mZ).$$  \hspace{1cm} (3.2)

Moreover, the variance of the perpendicular $Y - \hat{Y}$ is:

$$E(Y - \hat{Y})^2 = \text{var}(Y) - \text{cov}(Y, Z)\text{cov}^{-1}(Z, Z)\text{cov}^T(Y, Z).$$  \hspace{1cm} (3.3)

Proof. With an arbitrary vector $C$, let us introduce a random variable

$$\xi = (Y - m) + C(Z - mZ).$$  \hspace{1cm} (3.4)

We intend to choose $C$ such that $\xi$ is orthogonal to $\mathcal{M}$. If it is possible, (3.4) provides the decomposition for $Y$ (here $\mathcal{M}^\perp$ is the orthogonal complement to $\mathcal{M}$):

$$Y = m - C(Z - mZ) + \underbrace{\xi}_{\in \mathcal{M}} + \underbrace{\xi}_{\in \mathcal{M}^\perp}.$$  \hspace{1cm} (3.5)

Hence, if $C$ is found, we have $\hat{Y} = m - C(Z - mZ)$. Now, owing $Z - mZ \in \mathcal{M}$ and $\xi \in \mathcal{M}^\perp$ with $E(\xi) = 0$ and $E\xi = 0$, we find $E\xi(Z - mZ) = 0$. The latter and (3.4) provide

$$\text{cov}(Y, Z) + C\text{cov}(Z, Z) = 0$$  \hspace{1cm} (3.6)

and, whereas $\text{cov}(Z, Z)$ is assumed to be nonsingular matrix, the vector $C^o = -\text{cov}(Y, Z)\text{cov}^{-1}(Z, Z)$ solves (3.6).

Hence, the first statement holds true.

Notice now that $\xi = Y - \hat{Y}$ and

$$E\xi^2 = \text{var}(Y) - 2\text{cov}(Y, Z)\text{cov}^{-1}(Z, Z)\text{cov}^T(Y, Z)$$

$$+ \text{cov}(Y, Z)\text{cov}^{-1}(Z, Z)\text{cov}^T(Y, Z)$$

$$= \text{var}(Y) - \text{cov}(Y, Z)\text{cov}^{-1}(Z, Z)\text{cov}^T(\alpha, Z).$$

\[ \square \]

Remark 3.2. If the matrix $\text{cov}(Z, Z)$ is singular, the statement of the theorem remains valid with $\text{cov}^{-1}(Z, Z)$ replaced by the Moore-Penrose pseudo-inverse matrix $\text{cov}^+(Z, Z) = T^T(T^TT)^{-2}T$, where

$$T^TT = \text{cov}(Z, Z)$$

with $T$ the rectangular matrix of the full rank so that $TT^T$ a quadratic nonsingular matrix.

3.2. The conditional expectation in the wide sense.

The orthogonal projection $\hat{Y}$ of $Y$ on $\mathcal{M}$ is referred as conditional expectation in the wide sense given $\mathcal{M}$, $\hat{Y} = \hat{E}(Y|M)$, and plays an important role in the optimal in the mean square sense linear estimation.

We establish below the main properties of $\hat{E}(Y|M)$.
1. \[ \begin{align*}
E\hat{E}(Y|M) &= EY, \\
\hat{E}(Y|M) &= Y, \quad \text{if } Y \in M \\
\hat{E}(Y|M) &= 0, \quad \text{if } Y \in M^\perp.
\end{align*} \]

Proof. The random variable \( Y - \hat{Y} \) is orthogonal to \( M \), particularly, orthogonal to 1, that is \( E(Y - \hat{Y}) = 0 \).
If \( \alpha \perp M \), then by the definition \( \hat{E}(\alpha|M) = 0 \).

\( \square \)

2. If \( c \) is a constant, then 
\[ \hat{E}(cY|M) = c\hat{E}(Z|M). \]

Proof. It holds true since the operator of orthogonal projection is linear.

\( \square \)

3. If \( c_1, c_2 \) are constants and \( Y_1, Y_2 \) are random variables, possessing second moments, then 
\[ \hat{E}(c_1Y_1 + c_2Y_2|M) = c_1\hat{E}(Y_1|M) + c_2\hat{E}(Y_2|M). \]

Proof. It holds true since the operator of orthogonal projection is linear.

\( \square \)

4. If \( M' \) is a linear subspace of \( M \): \( M' \subseteq M \), then 
\[ \hat{E}\left(\hat{E}(Y|M)|M'\right) = \hat{E}(Y|M'). \]

Proof. Since both \( \hat{E}\left(\hat{E}(Y|M)|M'\right) \) and \( \hat{E}(Y|M') \) are from \( M' \), it suffices to show that \( \phi = \hat{E}\left(\hat{E}(\alpha|M)|M'\right) - \hat{E}(\alpha|M') \) is orthogonal to \( M' \). To this purpose, the decomposition 
\[ Y = \hat{E}(Y|M') + (Y - \hat{E}(Y|M')) \]

is used. Then, whereas \( (Y - \hat{E}(Y|M')) \perp M' \), for any \( \psi \in M' \) we have 
\[ E\psi Y = E\psi \hat{E}(Y|M'). \]

On the other side, since \( M' \subseteq M \), similarly we get 
\[ E\psi \hat{E}\left(\hat{E}(Y|M)|M'\right) = E\psi \hat{E}(Y|M) = E\psi Y. \]

Thus, \( E\psi \phi = 0 \).

\( \square \)

5. Let \( M' \) and \( M'' \) are orthogonal linear spaces: \( M' \perp M'' \). Set \( M = M' \oplus M'' \). Then for \( Y \), with \( EY = 0 \),
\[ \hat{E}(Y|M) = \hat{E}(Y|M') + \hat{E}(Y|M''). \]

Proof. Notice that \( \phi = Y - \hat{E}(Y|M') - \hat{E}(Y|M'') \) is orthogonal to \( M' \) and \( M'' \), that is orthogonal to \( M' \oplus M'' := M \) and the result holds true.

\( \square \)
6. For every $\eta \in \mathcal{M}$

$$E(Y - \eta)^2 \geq E(Y - \hat{Y})^2 = EY^2 - E\hat{Y}^2.$$  

Proof. Write

$$E(Y - \eta)^2 = E\left(Y - \hat{Y} - (\eta - \hat{Y})\right)^2$$

$$= E(Y - \hat{Y})^2 + E(\eta - \hat{Y})^2 - 2E(Y - \hat{Y})(\eta - \hat{Y})$$

$$= E(Y - \hat{Y})^2 + E(\eta - \hat{Y})^2$$

$$\geq E(Y - \hat{Y})^2.$$

□

7. The Cauchy-Schwarz inequality

$$E\left(E(Y|\mathcal{M})\right)^2 \leq EY^2.$$  

Proof. Write

$$0 \leq E(Y - \hat{E}(Y|\mathcal{M}))^2 = EY^2 - E\left(E(Y|\mathcal{M})\right)^2.$$

□

3.3. Law of large numbers for sequence of orthogonal random variables.

Let $X_1, X_2, \ldots$ be a sequence of orthogonal random variables with $EX_n \equiv 0$ and $EX_n^2 \equiv \sigma_n^2$. Denote by

$$a_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

**Proposition 3.3.** Under $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, it holds $\lim_{n \to \infty} Ea_n^2 = 0$.

Proof. The result is implied by $Ea_n^2 = \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2$ and the Kronecker lemma, which being adapted to the case considered, states:

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty \implies \lim_{n \to \infty} \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2 = 0. \quad (3.7)$$

We give below a sketch of the proof for (3.7). Set $V_n = \sum_{k=1}^{n} \frac{\sigma_k^2}{k^2}$ and notice that

$$\sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} (V_k - V_{k-1})k^2.$$  

Now, summing by parts we find

$$\sum_{k=1}^{n} \sigma_k^2 = V_n n^2 - \sum_{k=1}^{n} V_{k-1}(k^2 - (k - 1)^2),$$

so that

$$\frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2 = \left(V_n - \frac{1}{n^2} \sum_{k=1}^{n} V_{k-1}(k^2 - (k - 1)^2)\right).$$
The use of a telescopic sum \( n^2 = \sum_{k=1}^{n} (k^2 - (k-1)^2) \) allows to transform the above equality to the following one

\[
\frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2 = \frac{1}{n^2} \sum_{k=1}^{n} (V_n - V_{k-1}) (k^2 - (k-1)^2)
\]

and evaluate the right hand side of that equality. For \( n > N \), write

\[
\frac{1}{n^2} \sum_{k=1}^{n} (V_n - V_{k-1}) (k^2 - (k-1)^2) = \frac{1}{n^2} \sum_{k=1}^{N-1} (V_n - V_{k-1}) (k^2 - (k-1)^3)
\]

\[+ \frac{1}{n^2} \sum_{k=N}^{n} (V_n - V_{k-1}) (k^2 - (k-1)^3)\]

and notice that the first summand is evaluated above by

\[
\frac{N^2}{n^2} V_\infty = \frac{N^2}{n^2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \to 0, \ n \to \infty
\]

while the second by \( V_\infty - V_{N-1} = \sum_{k=N-1}^{\infty} \frac{\sigma_k^2}{k^2} \to 0, \ N \to \infty. \] □

3.4. The martingale in the wide sense.

Let \( X_1, X_2, \ldots \) be a sequence of random variables, with \( \mathbb{E}X_k^2 < \infty, k \geq 1 \), and \( \mathcal{M} \) be a linear space generated by \( 1, X_1, X_2, \ldots \) and \( \mathcal{M}_n \) be a linear space generated by \( 1, X_1, X_2, \ldots, X_n \). For a random variable \( \xi \), with \( \mathbb{E}\xi^2 < \infty \), set \( Y_n = \widehat{\mathbb{E}}(\xi|\mathcal{M}_n) \), \( n \geq 1 \). By property 4. of the conditional expectation in the wide sense, for \( n > m \) provides

\[ \widehat{\mathbb{E}}(Y_n|\mathcal{M}_m) = Y_m. \]

The random sequence with this property, created by the real conditional expectation, is named martingale. In our setting, the conditional expectation in the wide sense is used and so a notion of martingale in the wide sense is appropriate.

The aim of this Section is to show that \( Y_n, n \to \infty \), converges in \( L^2 \)-norm to a limit \( \widehat{\mathbb{E}}(\xi|\mathcal{M}) \):

\[
\text{l.i.m.}_{n \to \infty} \widehat{\mathbb{E}}(\xi|\mathcal{M}_n) = \widehat{\mathbb{E}}(\xi|\mathcal{M}). \tag{3.8}
\]

Proof of (3.8). Denote \( y_1 = Y_1, Y_n = Y_n - Y_{n-1}, n \geq 2 \) and notice that for \( n > m \) the random variable \( y_n \) is orthogonal to \( \mathcal{M}_m^\perp \), whereas \( \mathbb{E}y_n = 0 \) and moreover \( \widehat{\mathbb{E}}(y_n|\mathcal{M}_m) = \widehat{\mathbb{E}}(Y_{n-1}|\mathcal{M}_m) = Y_{m-1} = Y_m - Y_n = 0 \). Hence, \( Y_n = \sum_{k=1}^{n} y_k \) is the sum of zero mean orthogonal random variables \( x_1, \ldots, x_n \) and thanks of that \( \mathbb{E}Y_n^2 = \sum_{k=1}^{n} \mathbb{E}y_k^2 \). Then, obviously, \( \mathbb{E}Y_n^2 \) increases in \( n \). On the other hand, since by property 7. for the
conditional expectation in the wide sense $EY_n^2 \leq E\xi^2$, we have that for any $n$

$$\sum_{k=1}^{n} EY_k^2 \leq E\xi^2,$$

that is $\sum_{k=1}^{\infty} EY_k^2 \leq E\xi^2$. Then, the sequence $Y_n, n \geq 1$ converges in $L^2$-norm to $Y_\infty = \sum_{k=1}^{\infty} y_k$, since by the Cauchy criteria (Theorem 2.2 for Lect. 2)

$$E(Y_n - Y_m)^2 = E\left(\sum_{k=m+1}^{n} y_k\right)^2 = \sum_{k=m+1}^{n} EY_k^2 \to 0, \ n, m \to \infty.$$

3.5. Markov process in the wide sense.

Assume

$$\hat{E}(X_n|M_{n-1}) = a_nX_{n-1}, \ n \geq 2, \quad (3.9)$$

where $a_n$ is a sequence of numbers. Set $\varepsilon_n = X_n - a_nX_{n-1}$ and show that $(\varepsilon_n)_{n \geq 2}$ is a sequence of zero mean orthogonal random variables:

$$E\varepsilon_n \equiv 0$$

$$E\varepsilon_n\varepsilon_m = \begin{cases} 0, & n \neq m, \\ EY_n^2 - a_n^2EX_{n-1}^2 := b_n^2, & \text{otherwise.} \quad (3.10) \end{cases}$$

In fact, whereas by property 1. of the conditional expectation $EX_n = \hat{E}(X_n|M_{n-1})$, the first part from (3.10) is valid. By the definition $\varepsilon_n$ is orthogonal to $M_{n-1}$ and so for $M_m$ with $m < n$. The last property from (3.10) is provided by $E(X_n - a_nX_{n-1})^2 = EY_n^2 - a_n^2EX_{n-1}^2$.

Thus, the sequence $(X_n)_{n \geq 1}$ is defined by a recursion

$$X_n = a_nX_{n-1} + \varepsilon_n$$

subject by $X_1$ as the initial value, where $(\varepsilon_n)_{n \geq 2}$ is the sequence of zero mean and orthogonal random variables with $E\varepsilon_n^2 = b_n^2$.

Random sequences of such the type are named Markov processes in the wide sense.