4. LINEAR FILTERING. WIENER FILTER

Assume \((X_n, Y_n)\) is a pair of random sequences in which the first component \(X_n\) is referred a signal and the second an observation. The paths of the signal are unobservable and, on the contrary, the paths of observation \(Y_n\) are observable (measurable).

**Filtering** is a procedure of estimating the signal value \(X_n\) by a past of the observation \(Y_n, Y_{n-1}, \ldots, Y_{n-m} := Y_{n-m}^n\) (traditionally, \(m = n\) or \(m = \infty\)). Denote \(\hat{X}_n\) a filtering estimate which is a function of observation, that is \(\hat{X}_n = \hat{X}_n(Y_{n-m}^n)\). Here and now, we restrict ourselves by a consideration of linear estimates (shortly - linear filters) and shall evaluate a quality of filtering by the mean square error \(P_n = E(X_n - \hat{X}_n)^2\) for every time parameter \(n\). Obviously, such a setting requires the existence of second moments for the signal and observation \(EX_n^2 < \infty, EY_n^2 < \infty\) for all \(n\) and gives the optimal estimate in the form of conditional expectation in the wide sense

\[
\hat{X}_n = \hat{E}(X_n | Y_n), \quad (4.1)
\]

where \(Y_n\) is a linear space generated by “1, \(Y_n, Y_{n-1}, \ldots\)”, (henceforth, \(\hat{X}_n\) is assumed to be defined by (4.1)).

A filtering problem for random processes consists in an estimating of an unobservable signal (random process) via paths of an observable one.

4.1. The Wiener filter.

Assume that “signa-observation pair” \((X_n, Y_n)\) is a stationary random sequence in the wide sense. To have \(\hat{X}_n\) also a stationary sequence, it is assumed that \(m = \infty\), that is the path \(Y_{n-\infty}^n\) measurable. Since \(\hat{X}_n \in Y_n\) it is defined as:

\[
\hat{X}_n = c_n + \sum_{k=-\infty}^{n} c_{n,k} Y_k, \quad (4.2)
\]

where the sum in (4.2) converges in \(L^2\)-norm and parameters \(c_n, c_{n,k}\)'s are chosen such that the error \(X_n - \hat{X}_n\) is orthogonal to \(Y_n\). Without loss of a generality we may assume that both the signal and observation are zero mean and so taking the expectation from both sides of (4.2) we find that \(c_n \equiv 0\). In spite of a general formula for \(\hat{E}(X_n | Y_n)\) is given in Theorem 3.1 (Lect. 3), its application to the case considered is problematic owing to infinite dimensional of \(Y_{n-\infty}^n\) (one could apply that formula only if we are ready to be satisfied by an approximation of \(\hat{X}_n\) being the orthogonal projection on the linear space \(Y_{[n-m,n]}\) generated by “1, \(Y_n, Y_{n-1}, \ldots, Y_{n-m}\)” for large \(m\)).

Set \(R_{yy}(k - \ell) = EY_kY_\ell\) and \(R_{xy}(k - \ell) = EX_kY_\ell\).
Theorem 4.1. The optimal coefficients, \( c_{n,k} \equiv c_{n-k} \), and solves an infinite system of linear algebraic equation (Wiener-Hopf equation)

\[
R_{xy}(n - \ell) = \sum_{k=-\infty}^{n} c_{n-k} R_{y,y}(k - \ell). \tag{4.3}
\]

Proof. Since the filtering error \( X_n - \hat{X}_n \) is orthogonal to \( Y_n \), for any \( Y_\ell, \ell \leq n \) we have \( E(X_n - \hat{X}_n | Y_n) Y_\ell = 0 \). Taking into consideration (4.2), with \( c_n \equiv 0 \), from the latter equality we derive an infinite system of linear equations

\[
R_{xy}(n - \ell) = \sum_{k=-\infty}^{n} c_{n,k} R_{y,y}(k - \ell). \tag{4.4}
\]

So it remains to prove only that \( c_{n,k} = c_{n-k} \). Notice that \( \ell = n \), (4.4) is transformed to

\[
R_{xy}(0) = \sum_{k=-\infty}^{\ell} c_{\ell,k} R_{y,y}(k - \ell) = \sum_{j=\infty}^{0} c_{\ell,\ell-j} R_{y,y}(j).
\]

Since the left side of this equality is independent of \( \ell \), the right one has to be independent of \( \ell \) as well, what provides \( c_{\ell,\ell-j} = c_{\ell-(\ell-j)} = c_j \).

Example. Clearly, a solution of the Wiener-Hopf equation would be difficult to find directly. So, we give here a simple example for which solution for Wiener-Hopf equation is found.

Let \((X_n, Y_n)\) be defined as follows: for any \( n > -\infty \)

\[
X_n = a X_{n-1} + \varepsilon_n, \quad Y_n = X_n + \delta_n,
\]

where \((\varepsilon_n), (\delta_n)\) are orthogonal white noises, that is any linear combinations of \((\varepsilon_n)\) and \((\delta_n)\) are uncorrelated, with \( E\varepsilon_n \equiv 0, E\varepsilon_n^2 \equiv 1 \) and \( E\delta_n = 0, E\delta_n^2 \equiv 1 \). An absolute value of the parameter \( a \) is smaller than 1, \( |a| < 1 \), what guarantees that is the recursion for \( X_n \) is stable and \( X_n \) is a stationary process.

Parallel to \( \hat{X}_n \), let us introduce

\[
\hat{X}_{n,n-1} = E(X_n | Y_{n-1}) \quad \text{and} \quad \hat{Y}_{n,n-1} = E(Y_n | Y_{n-1})
\]

the orthogonal projections of \( X_n \) and \( Y_n \) on \( Y_{n-1} \) the subspace of \( Y_n \) and denote also by

\[
\text{Var}_{n,n-1}(X) = E\left( X_n - \hat{X}_{n,n-1} \right)^2, \quad \text{Var}_{n,n-1}(X,Y) = E\left( X_n - \hat{X}_{n,n-1} \right)\left( Y_n - \hat{Y}_{n,n-1} \right).
\]

It is obvious that

\[
\hat{X}_{n,n-1} = E(X_n | Y_{n-1}) = E(a X_{n-1} + \varepsilon_n | Y_{n-1}) = a \hat{X}_{n-1}
\]
and similarly
\[ \hat{Y}_{n,n-1} = E(X_n + \delta_n | Y_{n-1}) = a \hat{X}_{n-1}. \]

These relations allow us to find explicitly that
\[
\begin{align*}
\Var_{n,n-1}(X) &= E(X_n - a \hat{Y}_{n-1})^2 \\
&= E(a(X_n - a \hat{X}_{n-1}) + \varepsilon_n)^2 \\
&= a^2 P_{n-1} + 1 \\
\Var_{n,n-1}(Y) &= E(Y_n - a \hat{Y}_{n-1})^2 \\
&= E(a(X_{n-1} - a \hat{X}_{n-1}) + \varepsilon_n + \delta_n)^2 \\
&= a^2 P_{n-1} + 2
\end{align*}
\]

and that
\[
\begin{align*}
\Cov_{n,n-1}(X, Y) &= E(X_n - a \hat{X}_{n-1})(Y_n - a \hat{X}_{n-1}) \\
&= E(a(X_n - a \hat{X}_{n-1}) + \varepsilon_n)(a(X_{n-1} - a \hat{X}_{n-1}) + \varepsilon_n + \delta_n) \\
&= a^2 P_{n-1} + 1. \quad (4.7)
\end{align*}
\]

We show now that
\[
\hat{X}_n = a \hat{X}_{n-1} + \frac{a^2 P_{n-1} + 1}{a^2 P_{n-1} + 2} \left( Y_n - a \hat{X}_{n-1} \right). \quad (4.8)
\]

The proof of (4.8) looks like the proof of Theorem 3.1 (Lect.3). Set
\[
\eta = X_n - \hat{X}_{n,n-1} + C(Y_n - \hat{Y}_{n,n-1}) \quad (4.9)
\]
and choose the constant \( C \) such that \( \eta \perp Y_n \), that is \( E\eta \xi = 0 \) for any \( \xi \in Y_n \). With \( \xi = Y_n - \hat{Y}_{n,n-1} \), we have
\[
0 = \Cov_{n,n-1}(X, Y) + C\Var_{n,n+1}(Y),
\]
i.e. \( C \) is uniquely defined: \( C = -\frac{a^2 P_{n-1} + 1}{a^2 P_{n-1} + 2} \) and (4.8) is provided by (4.9). Notice also that the mean square error \( P_n \equiv P \), since \( (X_n, \hat{X}_n) \) forms stationary process. Hence, the recursion, given in (4.8), is transformed to
\[
\hat{X}_n = \left( a - \frac{a^2 P + 1}{a^2 P + 2} \right) \hat{X}_{n-1} + \frac{a^2 P + 1}{a^2 P + 2} Y_n. \quad (4.10)
\]

We have to find now the number \( P \). Notice that \( \eta = X_n - \hat{X}_n \), so that \( P = E\eta^2 \) and from (4.9) it follows
\[
P = a^2 P + 1 - \frac{(a^2 P + 1)^2}{a^2 P + 2} = a^2 P + 1 < 1
\]
and, then (4.10) looks as:
\[
\hat{X}_n = a(1 - P) \hat{X}_{n-1} + PY_n. \quad (4.11)
\]
Emphasizing that \(|a(1 - P)| < 1\) we derive from (4.11) that
\[
\hat{X}_n = \sum_{k=-\infty}^{n} (a(1 - P))^{n-k} PY_k,
\]
so that \(c_j \equiv (a(1 - P))^j P\).

**Generalization of the example.** \((X_n, Y_n)\), treated as the signal and observation, is zero mean stationary in wide sense process being part of entries for \(Z_n\) the stationary Markov process in the wide sense:
\[
Z_n = A Z_{n-1} + \xi_n, \quad (4.12)
\]
where \(A\) is a matrix eigenvalues of which have negative real parts and \((\xi_n)\) is a vector-valued white noise type sequence; the latter means that for some nonnegative definite matrix \(B\)
\[
E \xi_k \xi^T_\ell = \begin{cases} B, & k = \ell \\ 0 & \text{otherwise.} \end{cases}
\]
For notational convenience, assume that \(X_n\) is the first entry of \(Z_n\) and \(Y_n\) the last one. Also introduce \(Z'_n\) the subvector of \(Z_n\) containing all entries excluding \(Y_n\). Let for a definiteness, \(Z_n\) be the vector(column) of the size \(p\). Since \(Z_n = \begin{pmatrix} Z'_n \\ Y_n \end{pmatrix}\), we rewrite (4.12) into the form
\[
\begin{pmatrix} Z'_n \\ Y_n \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} Z'_{n-1} \\ Y_{n-1} \end{pmatrix} + \begin{pmatrix} \xi'_n \\ \xi_n \end{pmatrix}, \quad (4.13)
\]
where \(\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = A\) and \(\begin{pmatrix} \xi'_n \\ \xi_n \end{pmatrix} = \xi_n\); \(\xi'_n\) contains \(q - 1\) first component of \(\xi_n\). Notice also that \(B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\), where \(B_{11} \equiv E \xi'_n (\xi'_n)^T, B_{12} = \xi'_n (\xi'_n)^T, B_{21} = B_{21}, B_{22} \equiv E (\xi'_n)^2\).

Denote \(\hat{Z}'_n = \hat{E}(Z'_n | Y_n)\). Under
\[
B_{22} > 0, \quad (4.14)
\]
\(\hat{Z}'_n\) is the stationary process defined recurrently:
\[
\hat{Z}'_n = A_{11} \hat{Z}'_{n-1} + A_{12} Y_{n-1} + \frac{B_{12} + A_{11} PA_{12}^T}{B_{22} + A_{21} PA_{21}^T} (Y_n - A_{21} \hat{Z}'_{n-1} - A_{22} Y_{n-1}), \quad (4.15)
\]
where \(P \equiv E (Z'_n - \hat{Z}'_n)(Z'_n - \hat{Z}'_n)^T\) solves the algebraic Riccati equation
\[
P = A_{11} PA_{11}^* + B_{11} - \frac{(B_{12} + A_{11} PA_{12}^T)(B_{12} + A_{11} PA_{12}^T)^*}{B_{22} + A_{21} PA_{21}^T}. \quad (4.16)
\]
A derivation of (4.15) and (4.16) follows from Theorem 5.1 (Lect. 5). The first coordinate of $\hat{Z}_n$, obviously, is $\hat{X}_n$. That estimate possesses a structure $\hat{X}_n = \sum_{k=-\infty}^{n} c_{n-k} Y_k$ with coefficients $c_{n-k}$ calculating in accordance with (4.15) and (4.16).

This result shows how intricate might be a method for solving of the Wiener-Hopf equation.

There is another way for solving of the Wiener-Hopf equation given in terms spectral densities. For instance, to apply the Fourier transform to both sides of (4.3). It is exposed in others courses.

Fixed length of observation. Let $m$ be fixed number and $\mathcal{Y}_{[n-m,n]}$ be a linear space generated by “1, $Y_n$, $Y_{n-1}$, . . . , $Y_{n-m}$”. Define

$$\hat{X}_{[n-m,n]} = \hat{E}(X_n | \mathcal{Y}_{[n-m,n]})$$

the filtering estimate under fixed length of observation. Denote

$$Y_{n-m} = \begin{pmatrix} Y_n \\ Y_{n-1} \\ \vdots \\ Y_{n-m} \end{pmatrix}$$

and

$$\begin{cases} \text{Cov}_{[n-m,n]}(Y, Y) = E(Y_n^T (Y_{n-m}^n)^T) \\ \text{Cov}_{[n-m,n]}(X, Y) = E(X_n (Y_{n-m}^n)^T) \end{cases}$$

Since $(X_n, Y_n)$ is the stationary sequence, for any number $k$ we have

$$\begin{cases} \text{Cov}_{[n-m,n]}(Y, Y) = \text{Cov}_{[k-m,k]}(Y, Y) \\ \text{Cov}_{[n-m,n]}(X, Y) = \text{Cov}_{[k-m,k]}(X, Y). \end{cases}$$

The latter property provides, assuming that $\text{Cov}_{[n-m,n]}(Y, Y)$ is a non-singular matrix, that the vector

$$H_m = \text{Cov}_{[n-m,n]}(X, Y) \text{Cov}_{[n-m,n]}^{-1}(Y, Y)$$

depends only on the number $m$ and so is fixed for any $n$.

Thus, by Theorem 3.1 (Lect. 3) we have (recall that $EX_n \equiv 0$ and $EY_{n-m}^n \equiv 0$)

$$\hat{X}_{[n-m,n]} = H_m Y_{n-m}^n.$$

An attractiveness of this type filtering estimate is twofold. First, for fixed $m$ the vector $H_m$ is independent of $n$ and so is fixed (computed only once. Second, taking into consideration that for fixed $n$ the family $\mathcal{M}_{[n-m,n]}$, $m \geq 1$ increases, i.e. $\mathcal{M}_{[n-m,n]} \subseteq \mathcal{M}_{[n-(m+1),n]} \subseteq \ldots \subseteq \mathcal{M}_n$, the sequence of random variables $\hat{X}_{[n-m,n]}$, $m \geq 1$ forms a martingale in the wide sense. Therefore and by $EX_n^2 < \infty$, we have (see Section 3.4 in Lect. 3)

$$\lim_{m \to \infty} E((\hat{X}_n - \hat{X}_{[n-m,n]})^2) = 0.$$