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 $Y_n, Y_{n-1}, \dots, Y_{n-m} := Y_{n-m}^n$ (traditionally, $m = n$ or $m = \infty$). Denote $\widehat{X}_n$ a filtering estimate which is a function of observation, that is $\widehat{X}_n = \widehat{X}_n(Y_{n-m}^n)$. 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 $P_n = \boldsymbol{E}(X_n - \widehat{X}_n)^2$ 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

$$\widehat{X}_n = \widehat{\boldsymbol{E}}(X_n | \mathcal{Y}_n), \tag{4.1}$$

where $\mathcal{Y}_n$ is a linear space generated by "$1, Y_n, Y_{n-1}, \dots,$" (henceforth, $\widehat{X}_n$ 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 $\widehat{X}_n$ also a stationary sequence, it is assumed that $m = \infty$, that is the path $Y_{-\infty}^n$ measurable. Since $\widehat{X}_n \in \mathcal{Y}_n$ it is defined as:

$$\widehat{X}_n = c_n + \sum_{k=-\infty}^{n} c_{n,k} Y_k, \tag{4.2}$$

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 $X_n - \widehat{X}_n$ is orthogonal to $\mathcal{Y}_n$. 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 $\widehat{\boldsymbol{E}}(X_n | \mathcal{Y}_n)$ is given in Theorem 3.1 (Lect. 3), its application to the case considered is problematic owing to infinite dimensional of Y_∞^n (one could apply that formula only if we are ready to be satisfied by an approximation of $\widehat{X}_n$ being the orthogonal projection on the linear space $\mathcal{Y}_{[n-m,n]}$ generated by "$1, Y_n, Y_{n-1}, \dots, Y_{n-m}$" for large m).

Set $R_{yy}(k-\ell) = \boldsymbol{E} Y_k Y_\ell$ and $R_{xy}(k-\ell) = \boldsymbol{E} X_k Y_\ell$.

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 $X_n - \widehat{X}_n$ is orthogonal to $\mathcal{Y}_n$, for any Y_ℓ, $\ell \le n$ we have $\boldsymbol{E}\big(X_n - \widehat{X}_n|\mathcal{Y}_n\big)Y_\ell = 0$. 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 ℓ, the right one has to be independent of ℓ 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 $n > -\infty$

$$\begin{aligned} X_n &= aX_{n-1} + \varepsilon_n \\ Y_n &= X_n + \delta_n, \end{aligned}$$

where (ε_n), (δ_n) are orthogonal white noises, that is any linear combinations of (ε_n) and (δ_n) are uncorrelated, with $\boldsymbol{E}\varepsilon_n \equiv 0$, $\boldsymbol{E}\varepsilon_n^2 \equiv 1$ and $\boldsymbol{E}\delta_n \equiv 0$, $\boldsymbol{E}\delta_n^2 \equiv 1$. 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 $\widehat{X}_n$, let us introduce

$$\widehat{X}_{n,n-1} = \boldsymbol{E}(X_n|\mathcal{Y}_{n-1}) \quad \text{and} \quad \widehat{Y}_{n,n-1} = \boldsymbol{E}(Y_n|\mathcal{Y}_{n-1})$$

the orthogonal projections of X_n and Y_n on $\mathcal{Y}_{n-1}$ the subspace of $\mathcal{Y}_n$ and denote also by

$$\begin{aligned} \mathsf{Var}_{n,n-1}(X) &= \boldsymbol{E}\big(X_n - \widehat{X}_{n,n-1}\big)^2 \\ \mathsf{Var}_{n,n-1}(X,Y) &= \boldsymbol{E}\big(X_n - \widehat{X}_{n,n-1}\big)\big(Y_n - \widehat{Y}_{n,n-1}\big) \end{aligned}$$

It is obvious that

$$\widehat{X}_{n,n-1} = \boldsymbol{E}(X_n|\mathcal{Y}_{n-1}) = \boldsymbol{E}(aX_{n-1} + \varepsilon_n|\mathcal{Y}_{n-1}) = a\widehat{X}_{n-1}$$

and similarly

$$\widehat{Y}_{n,n-1} = \boldsymbol{E}(X_n + \delta_n | \mathcal{Y}_{n-1}) = a\widehat{X}_{n-1}.$$

These relations allow us to find explicitly that

$$\begin{aligned}
\mathsf{Var}_{n,n-1}(X) &= \boldsymbol{E}\big(X_n - a\widehat{Y}_{n-1}\big)^2 \\
&= \boldsymbol{E}\big(a(X_{n-1} - a\widehat{X}_{n-1}) + \varepsilon_n\big)^2 \\
&= a^2 P_{n-1} + 1 \qquad (4.5) \\
\mathsf{Var}_{n,n-1}(Y) &= \boldsymbol{E}\big(Y_n - a\widehat{Y}_{n-1}\big)^2 \\
&= \boldsymbol{E}\big(a(X_{n-1} - a\widehat{X}_{n-1}) + \varepsilon_n + \delta_n\big)^2 \\
&= a^2 P_{n-1} + 2 \qquad (4.6)
\end{aligned}$$

and that

$$\begin{aligned}
\mathsf{Cov}_{n,n-1}(X,Y) &= \boldsymbol{E}\big(X_n - a\widehat{X}_{n-1}\big)\big(Y_n - a\widehat{X}_{n-1}\big) \\
&= \boldsymbol{E}\big(a(X_{n-1} - a\widehat{X}_{n-1}) + \varepsilon_n\big)\big(a(X_{n-1} - a\widehat{X}_{n-1}) + \varepsilon_n + \delta_n\big) \\
&= a^2 P_{n-1} + 1. \qquad (4.7)
\end{aligned}$$

We show now that

$$\widehat{X}_n = a\widehat{X}_{n-1} + \frac{a^2 P_{n-1} + 1}{a^2 P_{n-1} + 2}\Big(Y_n - a\widehat{X}_{n-1}\Big). \qquad (4.8)$$

The proof of (4.8) looks like the proof of Theorem 3.1 (Lect.3). Set

$$\eta = X_n - \widehat{X}_{n,n-1} + C(Y_n - \widehat{Y}_{n,n-1}) \qquad (4.9)$$

and choose the constant C such that $\eta \perp \mathcal{Y}_n$, that is $\boldsymbol{E}\eta\xi = 0$ for any $\xi \in \mathcal{Y}_n$. With $\xi = Y_n - \widehat{Y}_{n,n-1}$, we have

$$0 = \mathsf{Cov}_{n,n-1}(X,Y) + C\mathsf{Var}_{n,n+1}(Y),$$

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 $(X_n, \widehat{X}_n)$ forms stationary process. Hence, the recursion, given in (4.8), is transformed to

$$\widehat{X}_n = \Big(a - a\frac{a^2 P + 1}{a^2 P + 2}\Big)\widehat{X}_{n-1} + \frac{a^2 P + 1}{a^2 P + 2} Y_n. \qquad (4.10)$$

We have to find now the number P. Notice that $\eta = X_n - \widehat{X}_n$, so that $P = \boldsymbol{E}\eta^2$ and from (4.9) it follows

$$P = a^2 P + 1 - \frac{(a^2 P + 1)^2}{a^2 P + 2} = \frac{a^2 P + 1}{a^2 P + 2} < 1$$

and, then (4.10) looks as:

$$\widehat{X}_n = a(1 - P)\widehat{X}_{n-1} + P Y_n. \qquad (4.11)$$

Emphasizing that $|a(1-P)|<1$ we derive from (4.11) that

$$\widehat{X}_n = \sum_{k=-\infty}^{n} \Big(a(1-P)\Big)^{n-k} PY_k,$$

so that $c_j \equiv \Big(a(1-P)\Big)^j P$.

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:

$$Z_n = AZ_{n-1} + \xi_n, \tag{4.12}$$

where A is a matrix eigenvalues of which have negative real parts and (ξ_n) is a vector-valued white noise type sequence; the latter means that for some nonnegative definite matrix B

$$\boldsymbol{E}\xi_k\xi_\ell^T = \begin{cases} B, & k=\ell \\ 0 & \text{otherwise.} \end{cases}$$

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

$$\begin{pmatrix} Z'_n \\ Y_n \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} Z'_{n-1} \\ Y_{n-1} \end{pmatrix} + \begin{pmatrix} \xi'_n \\ \xi''_n \end{pmatrix}, \tag{4.13}$$

where $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ $(=A)$ and $\begin{pmatrix} \xi'_n \\ \xi''_n \end{pmatrix}$ $(=\xi_n)$; ξ'_n contains $q-1$ first component of ξ_n. Notice also that $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, where $B_{11} \equiv \boldsymbol{E}\xi'_n(\xi'_n)^T$, $B_{12} \equiv \xi'_n(\xi''_n)^T$, $B_{21} = B_{21}$, $B_{22} \equiv \boldsymbol{E}(\xi'_n)^2$.

Denote $\widehat{Z'}_n = \widehat{\boldsymbol{E}}(Z'_n|\mathcal{Y}_n)$. Under

$$B_{22} > 0, \tag{4.14}$$

$\widehat{Z'}_n$ is the stationary process defined recurrently:

$$\begin{aligned} \widehat{Z'}_n &= A_{11}\widehat{Z'}_{n-1} + A_{12}Y_{n-1} \\ &\quad + \frac{B_{12} + A_{11}PA_{12}^T}{B_{22} + A_{21}PA_{21}^T}\Big(Y_n - A_{21}\widehat{Z'}_{n-1} - A_{22}Y_{n-1}\Big) \end{aligned} \tag{4.15}$$

where $P \equiv \boldsymbol{E}\big(Z'_n - \widehat{Z'}_n\big)\big(Z'_n - \widehat{Z'}_n\big)^T$ solves the algebraic Riccati equation

$$P = A_{11}PA_{11}^* + B_{11} - \frac{\big(B_{12} + A_{11}PA_{12}^T\big)\big(B_{12} + A_{11}PA_{12}^T\big)^*}{B_{22} + A_{21}PA_{21}^T}. \tag{4.16}$$

A derivation of (4.15) and (4.16) follows from Theorem 5.1 (Lect. 5). The first coordinate of $\widehat{Z'}_n$, obviously, is $\widehat{X}_n$. That estimate possesses a structure $\widehat{X}_n = \sum\limits_{k=-\infty}^{n} c_{n-k}Y_k$ 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 "$1, Y_n, Y_{n-1}, \dots, Y_{n-m}$". Define

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

the filtering estimate under fixed length of observation. Denote

$$Y_{n-m}^n = \begin{pmatrix} Y_n \\ Y_{n-1} \\ \vdots \\ Y_{n-m} \end{pmatrix} \quad \text{and} \quad \begin{cases} \mathsf{Cov}_{[n-m,n]}(Y,Y) = \boldsymbol{E}\Big(Y_{n-m}^n\big(Y_{n-m}^n\big)^T\Big) \\ \mathsf{Cov}_{[n-m,n]}(X,Y) = \boldsymbol{E}\Big(X_n\big(Y_{n-m}^n\big)^T\Big) \end{cases}$$

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

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

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

$$H_m = \mathsf{Cov}_{[n-m,n]}(X,Y)\mathsf{Cov}^{-1}_{[n-m,n]}(Y,Y) \tag{4.17}$$

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

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

$$\widehat{X}_{[n-m,n]} = H_m Y_{n-m}^n. \tag{4.18}$$

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 $\mathcal{M}_{[n-m,n]}$, $m \ge 1$ increases, i.e. $\mathcal{M}_{[n-m,n]} \subseteq \mathcal{M}_{[n-(m+1),n]} \subseteq \dots \subseteq \mathcal{M}_n$, the sequence of random variables $\widehat{X}_{[n-m,n]}, m \ge 1$ forms a martingale in the wide sense. Therefore and by $\boldsymbol{E}X_n^2 < \infty$, we have (see Section 3.4 in Lect. 3)

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

Consequently, for m large enough $\widehat{X}_{n-m,n]}$ might be an appropriate approximation for $\widehat{X}_n$.