## 13. KALMAN FILTER AND GENERALIZATIONS. Discrete time case

As in the continuous time case, the Kalman filter for the discrete time case is adapted to a special stochastic linear model given below. Let $X_n, Y_n, n = 0, 1...,$ be a sequence of a random vectors defined by the recursion

$$\begin{aligned} X_n &= a_1 X_{n-1} + a_2 Y_{n-1} + b_1 \varepsilon_n^1 + b_2 \varepsilon_n^2 \\ Y_n &= A_1 X_{n-1} + A_2 Y_{n-1} + B_1 \varepsilon_n^1 + B_2 \varepsilon_n^2 \end{aligned} \tag{13.1}$$

subject to $X_0, Y_0$, where all matrices can be dependent on $n$: $a_1 = a_1(n)$,...etc.

$X_n$ is column-vector of size $k$;

$Y_n$ is column-vector of size $\ell$;

the sizes of all matrices, involving in (13.1), are

$$\begin{aligned} a_1 &= a_{1(k\times k)}, \quad A_1 = A_{1(\ell\times k)}, \\ a_2 &= a_{2k\times\ell}, \quad A_2 = A_{2(\ell\times\ell)}; \\ b_1 &= b_{1k\times k}, \quad B_1 = B_{1(\ell\times k)}, \\ b_2 &= b_{2k\times\ell}, \quad B_2 = B_{2(\ell\times\ell)}; \end{aligned}$$

$(\varepsilon_n^1)$ is a sequence of uncorrelated random vectors (columns) of size $k$ with $\boldsymbol{E}\varepsilon_n^1 \equiv 0$ and $\boldsymbol{E}\varepsilon_n^1(\varepsilon_n^1)^T = I$ ($I$ is the unit matrix of the size $k \times k$, "$T''$ designates transpose),

$(\varepsilon_n^2)$ is a sequence of uncorrelated random vectors (columns) of size $\ell$ with $\boldsymbol{E}\varepsilon_n^2 \equiv 0$ and $\boldsymbol{E}\varepsilon_n^2(\varepsilon_n^2)^T = I$ ($I$ is the unit matrix of size $\ell \times \ell$),

sequences $(\varepsilon_n^1)$ and $(\varepsilon_n^2)$ are assumed to be not correlated and both are not correlated with the random vector $X_0, Y_0$ parameters of which are known

$$\begin{aligned} &\boldsymbol{E}X_0, \quad \boldsymbol{E}Y_0, \\ &\mathrm{Var}(X_0, X_0) = \boldsymbol{E}(X_0 - \boldsymbol{E}X_0)(X_0 - \boldsymbol{E}X_0)^T, \\ &\mathrm{Var}(X_0, Y_0) = \boldsymbol{E}(X_0 - \boldsymbol{E}X_0)(Y_0 - \boldsymbol{E}Y_0)^T, \\ &\mathrm{Var}(Y_0, Y_0) = \boldsymbol{E}(Y_0 - \boldsymbol{E}Y_0)(Y_0 - \boldsymbol{E}Y_0)^T. \end{aligned}$$

For further convenience, introduce the matrices

$$\begin{aligned} b \circ b &= b_1 b_1^T + b_2 b_2^T \\ b \circ B &= b_1 B_1^T + b_2 B_2^T \\ B \circ B &= B_1 B_1^T + B_2 B_2^T. \end{aligned}$$

The vector $X_n$ is unobservable signal while the vector $Y_n$ is observation. As the linear and optimal in the mean square filtering estimate of $X_n$ (given observations $Y_0, ..., Y_n$), we take the orthogonal projection $\widehat{X}_n$ of the random vector $X_n$ on linear space generated by the random vectors $Y_0, ..., Y_n$. The definition of the above-mentioned projection is the following: $\widehat{X}_n$ can be represented as a linear combination of $Y_0, ..., Y_n$ and a non random vector $C$

$$\widehat{X}_n = C + \sum_{k=0}^{n} C_{nk} Y_k,$$

where matrices $C_{nk}$'s are chosen such that the matrix $P_n = E(X_n - \widehat{X}_n)(X_n - \widehat{X}_n)^T$ is minimal, that is for any other set $\widetilde{C}$, $\widetilde{C}_{nk}$'s the matrix $\widetilde{P}_n = \boldsymbol{E}(X_n - \widetilde{X}_n)(X_n - \widetilde{X}_n)^T$ is less than $P_n$ or, what is equivalent, the matrix $\widetilde{P}_n - P_n$ is non negative definite.

The Kalman filter defines the pair $\widehat{X}_n$ and $P_n$ by the following recursions

$$\begin{aligned} \widehat{X}_n &= a_1 \widehat{X}_{n-1} + a_2 Y_{n-1} \\ &\quad + [b \circ B + a_1 P_{n-1} A_1^T][B \circ B + A_1 P_{n-1} A_1^T]^{-1}(Y_n - A_1 \widehat{X}_{n-1} - A_2 Y_{n-1}), \\ P_n &= a_1 P_{n-1} a_1^T + b \circ b \\ &\quad - [b \circ B + a_1 P_{n-1} A_1^T][B \circ B + A_1 P_{n-1} A_1^T]^{-1}[b \circ B + a_1 P_{n-1} A_1^T]^T \end{aligned} \quad (13.2)$$

subject to

$$\begin{aligned} \widehat{X}_0 &= \boldsymbol{E} X_0 + \mathrm{Var}(X_0, Y_0)\mathrm{Var}^{-1}(Y_0, Y_0)(Y_0 - \boldsymbol{E} Y_0) \\ P_0 &= \mathrm{Var}(X_0, X_0) - \mathrm{Var}(X_0, Y_0)\mathrm{Var}^{-1}(Y_0, Y_0)\mathrm{Var}^T(X_0, Y_0). \end{aligned} \quad (13.3)$$

**Remark.** In a case of singular matrix $\mathrm{Var}(Y_0, Y_0)$ the inverse matrix $\mathrm{Var}^{-1}(Y_0, Y_0)$ can be replaced on the Moore-Penrouse pseudo inverse matrix $\mathrm{Var}^{+}(Y_0, Y_0)$.

**Proof of (13.3).** Assume the matrix $\mathrm{Var}(Y_0, Y_0)$ is non singular. For some matrix $G$, Define a random vector

$$\eta = (X_0 - \boldsymbol{E} X_0) + G(Y_0 - \boldsymbol{E} Y_0).$$

The matrix $G$ can be chosen such that the vector $\eta$ and $Y_0 - \boldsymbol{E} Y_0$ are orthogonal to each other. Since $\eta = 0$, $G$ has to taken such that

$$\boldsymbol{E} \eta (Y_0 - \boldsymbol{E} Y_0)^T = 0.$$

The last condition implies $0 = \mathrm{Var}(X_0, Y_0) + G\mathrm{Var}(Y_0, Y_0)$ and so

$$G = -\mathrm{Var}(X_0, Y_0)\mathrm{Var}^{-1}(Y_0, Y_0).$$

Therefore, with chosen matrix $G$ the vector $X_0$ can be decomposed into $X_0 = \widehat{X}_0 + \eta$, where the vector $\eta$ is orthogonal to any of vectors $C + C_1 Y_0$ ($C, C_1$ are arbitrary vector and matrix respectively).

Thus $\widehat{X}_0$ is the required orthogonal projection. $C + C_1 Y_0$.

To find the matrix $P_0$, note that $P_0 = \boldsymbol{E}\eta\eta^T$. Consequently

$$\begin{aligned} P_0 &= \mathrm{Var}(X_0, X_0) + \mathrm{Var}(X_0, Y_0)G^T + G\mathrm{Var}(Y_0, X_0) \\ &\quad + G\mathrm{Var}(Y_0, Y_0)G^T \\ &= \mathrm{Var}(X_0, X_0) - 2\mathrm{Var}(X_0, Y_0)\mathrm{Var}^{-1}(Y_0, Y_0)\mathrm{Var}^T(X_0, Y_0) \\ &\quad + \mathrm{Var}(X_0, Y_0)\mathrm{Var}^{-1}(Y_0, Y_0)\mathrm{Var}^T(X_0, Y_0) \\ &= \mathrm{Var}(X_0, X_0) - \mathrm{Var}(X_0, Y_0)\mathrm{Var}^{-1}(Y_0, Y_0)\mathrm{Var}^T(X_0, Y_0). \end{aligned}$$

**Proof of (13.2).** Denote by $\widehat{X}_{n|n-1}$ and $\widehat{Y}_{n|n-1}$ the orthogonal projections of $X_n$ and $Y_n$ on the linear space generated by $Y_0, ..., Y_{n-1}$. Since an operator of the orthogonal projection is linear, due to (13.1), we get

$$\begin{aligned} \widehat{X}_{n|n-1} &= a_1\widehat{X}_{n-1} + a_2 Y_{n-1} + b_1\widehat{\varepsilon^1}_{n|n-1} + b_2\widehat{\varepsilon^2}_{n|n-1} \\ \widehat{Y}_{n|n-1} &= A_1\widehat{X}_{n-1} + A_2 Y_{n-1} + B_1\widehat{\varepsilon^1}_{n|n-1} + B_2\widehat{\varepsilon^2}_{n|n-1}, \end{aligned} \tag{13.4}$$

where $\widehat{\varepsilon^1}_{n|n-1}$ and $\widehat{\varepsilon^2}_{n|n-1}$ are the orthogonal projections of $\varepsilon^1_n$ and $\varepsilon^2_n$ on the linear space generated by $Y_0, ..., Y_{n-1}$. By the assumptions, the random variables $\varepsilon^1_n$ and $\varepsilon^2_n$ are orthogonal to any linear combination of $Y_0, ..., Y_{n-1}$. Therefore their orthogonal projections coincide with their expectations, i.e. $\widehat{\varepsilon^i}_{n|n-1} = 0, i = 1, 2$ and, in turn,

$$\begin{aligned} \widehat{X}_{n|n-1} &= a_1\widehat{X}_{n-1} + a_2 Y_{n-1} \\ \widehat{Y}_{n|n-1} &= A_1\widehat{X}_{n-1} + A_2 Y_{n-1}. \end{aligned} \tag{13.5}$$

As previously, define the random vector

$$\eta_n = X_n - \widehat{X}_{n|n-1} + G(Y_n - \widehat{Y}_{n|n-1}),$$

where the matrix $G$ is chosen such that $\boldsymbol{E}\eta_n(Y_n - \widehat{Y}_{n|n-1}) = 0$. The last requirement is nothing but

$$G = -\boldsymbol{E}[X_n - \widehat{X}_{n|n-1}][Y_n - \widehat{Y}_{n|n-1}]^T \Big([Y_n - \widehat{Y}_{n|n-1}][Y_n - \widehat{Y}_{n|n-1}]^T\Big)^{-1}.$$

Using this formula, we find now other type of representation for $G$ expressed via $P_{n-1} = \boldsymbol{E}[X_{n-1} - \widehat{X}_{n|n-1}][X_{n-1} - \widehat{X}_{n|n-1}]^T$. From (13.1) and (13.5), it follows that

$$\begin{aligned} X_n - \widehat{X}_{n|n-1} &= a_1[X_{n-1} - \widehat{X}_{n-1}] + b_1\varepsilon^1_n + b_2\varepsilon^2_n \\ Y_n - \widehat{Y}_{n|n-1} &= A_1[X_{n-1} - \widehat{X}_{n-1}] + B_1\varepsilon^1_n + B_2\varepsilon^2_n. \end{aligned}$$

Hence,

$$\begin{aligned} \boldsymbol{E}[X_n - \widehat{X}_{n|n-1}][Y_n - \widehat{Y}_{n|n-1}]^T &= a_1P_{n-1}A_1^T + b \circ B \\ \boldsymbol{E}[Y_n - \widehat{Y}_{n|n-1}][Y_n - \widehat{Y}_{n|n-1}]^T &= A_1P_{n-1}A_1^T + B \circ B \end{aligned}$$

and consequently

$$G = -[a_1P_{n-1}A_1^T + b \circ B][A_1P_{n-1}A_1^T + B \circ B]^{-1}. \tag{13.6}$$

Summing now all previous results, we arrive at the following decomposition

$$X_n = \widehat{X}_{n|n-1} + [a_1P_{n-1}A_1^T + b \circ B][A_1P_{n-1}A_1^T + B \circ B]^{-1}[Y_n - \widehat{Y}_{n|n-1}] + \eta_n. \tag{13.7}$$

To get the announced result for $\widehat{X}_n$, it remains to recall (13.5) and to show that the orthogonal projection $\widehat{\eta}_n$ of $\eta_n$ on the linear space generated by $Y_0, ..., Y_n$ is equal zero: $\widehat{\eta}_n = 0$. To this end, it is sufficient to check that the random vector $\eta_n$ is orthogonal to any linear combination of $Y_0, ..., Y_n$, say, $Y_n + \sum_{k=0}^{n-1} C_{nk}Y_k + CY_0$. On the other hand, since $\widehat{Y}_{n|n-1}$ possesses a form $\sum_{k=0}^{n-1} C_{nk}Y_k + CY_0$, it is sufficient to check only the orthogonality of vectors $\eta_n$ and $Y_n - \widehat{Y}_{n|n-1}$. Since the matrix $G$ has been chosen to satisfy this requirement we get $\widehat{\eta}_n = 0$.

Therefore, (13.7) and (13.5) imply the first recursion in (13.2).

The second recursion in (13.2) is derived by using the decomposition

$$X_n - \widehat{X}_{n|n-1} = [a_1P_{n-1}A_1^T + b \circ B][A_1P_{n-1}A_1^T + B \circ B]^{-1}[Y_n - \widehat{Y}_{n|n-1}] + \eta_n$$

with orthogonal vectors $[a_1P_{n-1}A_1^T+b\circ B][A_1P_{n-1}A_1^T+B\circ B]^{-1}[Y_n-\widehat{Y}_{n|n-1}]$ and $\eta_n$. Due to this decomposition, we get

$$\begin{aligned}
P_n &= \boldsymbol{E}\eta_n\eta_n^T \\
&\quad +[a_1P_{n-1}A_1^T+b\circ B][A_1P_{n-1}A_1^T+B\circ B]^{-1} \\
&\quad \times\boldsymbol{E}[Y_n-\widehat{Y}_{n|n-1}][Y_n-\widehat{Y}_{n|n-1}]^T[A_1P_{n-1}A_1^T+B\circ B]^{-1}[a_1P_{n-1}A_1^T+b\circ B]^T \\
&= \boldsymbol{E}\eta_n\eta_n^T \\
&\quad +[a_1P_{n-1}A_1^T+b\circ B][A_1P_{n-1}A_1^T+B\circ B]^{-1}[a_1P_{n-1}A_1^T+b\circ B]^T.
\end{aligned}$$

On the other hand, the definition of the random vector $\eta_n$ and the matrix $G$ imply

$$\begin{aligned}
\boldsymbol{E}\eta_n\eta_n^T &= [a_1P_{n-1}a_1^T+b\circ b]+G[A_1P_{n-1}A_1^T+B\circ B]G^T \\
&\quad +[a_1P_{n-1}a_1^T+b\circ b][A_1P_{n-1}A_1^T+B\circ B]G^T \\
&= [a_1P_{n-1}a_1^T+b\circ b] \\
&\quad -[a_1P_{n-1}A_1^T+b\circ B][A_1P_{n-1}A_1^T+B\circ B]^{-1}[a_1P_{n-1}A_1^T+b\circ B]^T.
\end{aligned}$$

**Remark.** If all vectors $X_0, Y_0, \varepsilon_k^1, \varepsilon_k^2, k\geq 1$ are jointly Gaussian, then the sequence $X_n, Y_n, n=0,1,...$ is Gaussian as well and the orthogonal projection coincides with the conditional expectation:

$$\widehat{X}_n=\boldsymbol{E}\Big(X_n\Big|Y_k, 0\leq k\leq n\Big).$$

**Generalized Kalman filter.** Parallel to the model given in (13.1), consider non linear model

$$\begin{aligned}
X_n &= a_1X_{n-1}+a_2(Y_0^{n-1})+b_1\varepsilon_n^1+b_2\varepsilon_n^2 \\
Y_n &= A_1X_{n-1}+A_2(Y_0^{n-1})+B_1\varepsilon_n^1+B_2\varepsilon_n^2,
\end{aligned}$$

where the conditions from the remark above takes place and $a_2(Y_0^{n-1})$ and $A_2(Y_0^{n-1})$ are past-dependent nonlinear functionals of the observable values $Y_k, k=0,...,n-1$. Denote by

$$\begin{aligned}
\widehat{X}_n &= E\Big(X_n\Big|Y_k, 0\leq k\leq n\Big) \\
P_n &= E[X_n-\widehat{X}_n][X_n-\widehat{X}_n]^T.
\end{aligned}$$

The generalized Kalman filter defines recursions ((compare (13.2) and (13.3))

$$
\begin{aligned}
\widehat{X}_n &= a_1\widehat{X}_{n-1} + a_2(Y_0^{n-1}) \\
&\quad + [b\circ B + a_1P_{n-1}A_1^T][B\circ B + A_1P_{n-1}A_1^T]^{-1}(Y_n - A_1\widehat{X}_{n-1} + A_2(Y_0,^{n-1}), \\
P_n &= a_1P_{n-1}a_1^T + b\circ b \\
&\quad - [b\circ B + a_1P_{n-1}A_1^T][B\circ B + A_1P_{n-1}A_1^T]^{-1}[b\circ B + a_1P_{n-1}A_1^T]^T
\end{aligned}
$$

subject to

$$
\begin{aligned}
\widehat{X}_0 &= \boldsymbol{E}X_0 + \mathrm{Var}(X_0,Y_0)\mathrm{Var}^{-1}(Y_0,Y_0)(Y_0 - \boldsymbol{E}Y_0) \\
P_0 &= \mathrm{Var}(X_0,X_0) - \mathrm{Var}(X_0,Y_0)\mathrm{Var}^{-1}(Y_0,Y_0)\mathrm{Var}^T(X_0,Y_0).
\end{aligned}
$$

**Example.** Consider the scalar signal $X_n$ defined by linear recursion

$$X_n = aX_{n-1} + b\varepsilon_n^1,$$

where $a$ and $b$ are known constants and $\varepsilon_n^1, n \geq 1$ is i.i.d. sequence of random variables with $\boldsymbol{E}\varepsilon_1^1 = 0$ and $\boldsymbol{E}(\varepsilon_1^1)^2 = 1$. The observation

$$Y_n = X_n + \varepsilon_n^2,$$

where $\varepsilon_n^2, n \geq 0$ is i.i.d. sequence of random variables with $\boldsymbol{E}\varepsilon_1^2 = 0$ and $\boldsymbol{E}(\varepsilon_1^2)^2 = 1$ and it is independent of $X_0$ and $\varepsilon_n^1, n \geq 1$.

Formally, we could not apply the result obtained to the filtering $X_n$ given $Y_0, ..., Y_n$. So, it is shown below how this model can be adapted to the general theoretical result. It is clear, there exists an equivalent recursion for $X_n, Y_n$

$$
\begin{aligned}
X_n &= aX_{n-1} + b\varepsilon_n^1 \\
Y_n &= aX_{n-1} + b\varepsilon_n^1 + \varepsilon_n^2
\end{aligned}
$$

for which the general filtering recursions are applicable:

$$
\begin{aligned}
\widehat{X}_n &= a\widehat{X}_{n-1} + \frac{b^2 + a^2P_{n-1}}{b^2 + 1 + a^2P_{n-1}}(Y_n - a\widehat{X}_{n-1}) \\
P_n &= a^2P_{n-1} + b^2 - \frac{[b^2 + a^2P_{n-1}]^2}{b^2 + 1 + a^2P_{n-1}}.
\end{aligned}
$$

Let us find now $\widehat{X_0}$ and $P_0$. Obviously $Y_0 = X_0 + \varepsilon_0^2$. The independence assumption for $X_0$ and $\varepsilon_0^2$ implies

$$\begin{aligned} \widehat{X}_0 &= \boldsymbol{E}X_0 + \frac{\mathrm{Var}(X_0, X_0)}{1 + \mathrm{Var}(X_0, X_0)}(Y_0 - \boldsymbol{E}X_0) \\ P_0 &= \mathrm{Var}(X_0, X_0) - \frac{\mathrm{Var}^2(X_0, X_0)}{1 + \mathrm{Var}(X_0, X_0)}. \end{aligned}$$

**Innovation sequence.** Let $\widehat{X}_n$ and $P_n$ be defined by the Kalman filter (13.2). Introduce a sequence $\overline{\varepsilon}_n, n \geq 1$, putting

$$\overline{\varepsilon}_n = [B \circ B + A_1 P_{n-1} A_1^T]^{-1/2}(Y_n - A_1 \widehat{X}_{n-1} - A_2 Y_{n-1}). \tag{13.8}$$

The sequence $\overline{\varepsilon}_n, n \geq 1$ is named the innovation sequence by virtue of the following properties:

$$\begin{aligned} E\overline{\varepsilon}_n &\equiv 0 \\ E\overline{\varepsilon}_n\overline{\varepsilon}_m &= \begin{cases} I \ (I = I_{\ell\times\ell} \text{ is the unite matrix}), & n = m \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The first property and the second for $n = m$ are obvious since

$$Y_n - A_1\widehat{X}_{n-1} - A_2 Y_{n-1} = A_1[X_n - \widehat{X}_{n-1}] + B_1\varepsilon_n^1 + B_2\varepsilon_n^2.$$

For $n > m$, the second property follows due to the fact that $A_1[X_n - \widehat{X}_{n-1}] + B_1\varepsilon_n^1 + B_2\varepsilon_n^2$ is orthogonal to any linear combination of $Y_0, ..., Y_m$. For $n < m$, the proof is similar.