

5. RECURRENT ORTHOGONAL PROJECTIONS. KALMAN FILTER

5.1. **Main theorem.** It Lect. 4, the recurrent orthogonal projection was used twice for specific settings. Here, a general result will be given.

Let (X_n, Y_n) be a pair of vector-valued random sequences started from $n = 0$. For a definiteness, assume that size of vector X_n and Y_n is ℓ . Denote by \mathcal{Y}_n the linear space generated by “1, Y_0, \dots, Y_n ” and for $n \geq 1$ set

$$\begin{aligned}\widehat{X}_{n,n-1} &= \widehat{\mathbf{E}}(X_n | \mathcal{Y}_{n-1}) \\ \widehat{Y}_{n,n-1} &= \widehat{\mathbf{E}}(Y_n | \mathcal{Y}_{n-1}) \\ \text{Cov}_{n,n-1}(X, X) &= \mathbf{E} \left(X_n - \widehat{\mathbf{E}}(X_n | \mathcal{Y}_{n-1}) \right) \left(X_n - \widehat{\mathbf{E}}(X_n | \mathcal{Y}_{n-1}) \right)^T \\ \text{Cov}_{n,n-1}(X, Y) &= \mathbf{E} \left(X_n - \widehat{\mathbf{E}}(X_n | \mathcal{Y}_{n-1}) \right) \left(Y_n - \widehat{\mathbf{E}}(Y_n | \mathcal{Y}_{n-1}) \right)^T \\ \text{Cov}_{n,n-1}(Y, Y) &= \mathbf{E} \left(Y_n - \widehat{\mathbf{E}}(Y_n | \mathcal{Y}_{n-1}) \right) \left(Y_n - \widehat{\mathbf{E}}(Y_n | \mathcal{Y}_{n-1}) \right)^T.\end{aligned}\tag{5.1}$$

Also set

$$\begin{aligned}\widehat{X}_n &= \widehat{\mathbf{E}}(X_n | \mathcal{Y}_n) \\ P_n &= \mathbf{E} \left(X_n - \widehat{X}_n \right) \left(X_n - \widehat{X}_n \right)^T.\end{aligned}\tag{5.2}$$

Compare the result below with Theorem 3.1 on Lect. 3.

Theorem 5.1.

$$\begin{aligned}\widehat{X}_n &= \widehat{X}_{n,n-1} + \text{Cov}_{n,n-1}(X, Y) \text{Cov}_{n,n-1}^{-1}(Y, Y) \left(Y_n - \widehat{Y}_{n,n-1} \right) \\ P_n &= \text{Cov}_{n,n-1}(X, X) \\ &\quad - \text{Cov}_{n,n-1}(X, Y) \text{Cov}_{n,n-1}^{-1}(Y, Y) \text{Cov}_{n,n-1}^T(X, Y).\end{aligned}\tag{5.3}$$

It the matrix $\text{Cov}_{n,n-1}(Y, Y)$ above is singular, $\text{Cov}_{n,n-1}^{-1}(Y, Y)$ in (5.3) is replaced by the pseudo-inverse matrix $\text{Cov}_{n,n-1}^\oplus(Y, Y)$ in Moore-Penrose sense ¹.

Proof. Introduce a random vector

$$\eta = X_n - \widehat{X}_{n,n-1} + C \left(Y_n - \widehat{Y}_{n,n-1} \right),\tag{5.4}$$

where a deterministic matrix C is chosen such that $\eta \perp \mathcal{M}_n$. Notice that $\mathbf{E}\eta = 0$ an thus η is orthogonal to any constant. Further, for any $j < n$ we have $\mathbf{E}\eta Y_j^T = 0$ since $Y_j \in \mathcal{M}_{n-1}$ while $\eta \perp \mathcal{M}_{n-1}$. So, it remains to choose C such that $\mathbf{E}\eta(Y_n - \widehat{Y}_{n,n-1})^T = 0$.

¹see, Albert, A. (1972): Regression and the Moore-Penrose Pseudoinverse. Academic, New York London

Multiplying the right both sides of (5.4) by $(Y_n - \widehat{Y}_{n,n-1})$ and taking the expectation we find

$$0 = \text{Cov}_{n,n-1}(X, Y) + C \text{Cov}_{n,n-1}(Y, Y). \quad (5.5)$$

Under nonsingular matrix $\text{Cov}_{n,n-1}(Y, Y)$,

$$C = -\text{Cov}_{n,n-1}(X, Y) \text{Cov}_{n,n-1}^{-1}(Y, Y) \quad (5.6)$$

solves (5.5).

Hence, the first statement holds true. To verify the validity of the second one, it suffices to mention that $P_n = \mathbf{E}\eta\eta^T$. Then

$$\begin{aligned} P_n &= \text{Cov}_{n,n-1}(X, X) + C \text{Cov}_{n,n-1}^T(X, Y) + \text{Cov}_{n,n-1}(X, Y) C^T \\ &\quad + C \text{Cov}_{n,n-1}(Y, Y) C^T \end{aligned}$$

and the second statement is provided now by (5.6).

If $\text{Cov}_{n,n-1}(Y, Y)$ is a singular matrix,

$$C = -\text{Cov}_{n,n-1}(X, Y) \text{Cov}_{n,n-1}^\oplus(Y, Y)$$

solves (5.5) as well; the latter is equivalent to an equality

$$\text{Cov}_{n,n-1}(X, Y) = \text{Cov}_{n,n-1}(X, Y) \text{Cov}_{n,n-1}^\oplus(Y, Y) \text{Cov}_{n,n-1}(Y, Y) \quad (5.7)$$

proved in Proposition 5.2 (see Section 5.2). So the first statement is valid. The second statement is proved similarly with a help of formula 1. from the same section. \square

KALMAN FILTER

Let a signal X_n and observation Y_n are vector-valued (of sizes k and ℓ respectively) random sequences defined by the following recurrent equations

$$\begin{aligned} X_n &= a_1(n)X_{n-1} + a_2(n)Y_{n-1} + \varepsilon_n^1 \\ Y_n &= A_1(n)x_{n-1} + A_2(n)Y_{n-1} + \varepsilon_n^2, \end{aligned} \quad (5.8)$$

subject to the initial conditions X_0, Y_0 . In (5.8), the following objects are involved.

$a_1(n), a_2(n)$ and $A_1(n), A_2(n)$ are, dependent on time parameter n , known matrices of sizes $k \times k$, $k \times \ell$, $\ell \times k$, $\ell \times \ell$ respectively.

ε_n^1 and ε_n^2 are vector-valued (of sizes k and ℓ respectively) white noises type random sequences with

$$\begin{aligned} \mathbf{E}\varepsilon_n^1(\varepsilon_m^1)^T &= \begin{cases} B_{11}(n), & n = m, \\ 0, & \text{otherwise} \end{cases} \\ \mathbf{E}\varepsilon_n^2(\varepsilon_m^2)^T &= \begin{cases} B_{22}(n), & n = m, \\ 0, & \text{otherwise} \end{cases} \\ \mathbf{E}\varepsilon_n^1(\varepsilon_m^2)^T &= \begin{cases} B_{12}(n), & n = m, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (5.9)$$

where $B_{11}(n), B_{22}(n), B_{12}(n)$ are n -dependent known matrices of sizes $k \times k, \ell \times \ell, k \times \ell$ respectively.

Random objects (X_0, Y_0) and $(\varepsilon_n^1, \varepsilon_n^2)$ are orthogonal, that is any linear combinations of entries of (X_0, Y_0) and $(\varepsilon_n^1, \varepsilon_n^2)$ have zero correlations; $\mathbf{E}X_0, \mathbf{E}Y_0, \mathbf{Cov}(X_0, X_0), \mathbf{Cov}(Y_0, Y_0), \mathbf{Cov}(X_0, Y_0)$ are known.

Henceforth, for notational simplicity, the n -dependence is omitted, that is we write a_1, a_2, \dots, B_{12} instead of $a_1(n), a_2(n), \dots, B_{12}(n)$. Also, to avoid ‘play’ with pseudoinverse matrices, we assume that matrices

$$\mathbf{Cov}(Y_0, Y_0) \text{ and } B_{22}(n) \text{ are nonsingular.} \quad (5.10)$$

Set \mathcal{M}_n the linear space generated by “ $1, Y_0, Y_1, \dots, Y_n$ ” and

$$\widehat{X}_n = \widehat{\mathbf{E}}(X_n | \mathcal{M}_n), \quad P_n = \mathbf{E}(X_n - \widehat{X}_n)(X_n - \widehat{X}_n)^T.$$

We start with the derivation of the filtering estimate for $n = 0$.

By Theorem 3.1 (Lect. 3), we have

$$\begin{aligned} \widehat{X}_0 &= \mathbf{E}X_0 + \mathbf{Cov}(X_0, Y_0)\mathbf{Cov}^{-1}(Y_0, Y_0)(Y_0 - \mathbf{E}Y_0) \\ P_0 &= \mathbf{Cov}(X_0, X_0) - \mathbf{Cov}(X_0, Y_0)\mathbf{Cov}^{-1}(Y_0, Y_0)\mathbf{Cov}^T(X_0, Y_0). \end{aligned} \quad (5.11)$$

For $n \geq 1$, we apply Theorem 5.1. From (5.1), adapted to (5.8), it follows

$$\begin{aligned} \widehat{X}_{n,n-1} &= a_1\widehat{X}_{n-1} + a_2Y_{n-1} \\ \widehat{Y}_{n,n-1} &= A_1\widehat{X}_{n-1} + A_2Y_{n-1} \\ \mathbf{Cov}_{n,n-1}(X, X) &= a_1P_{n-1}a_1^T + B_{11} \\ \mathbf{Cov}_{n,n-1}(X, Y) &= a_1P_{n-1}A_1^T + B_{12} \\ \mathbf{Cov}_{n,n-1}(Y, Y) &= A_1P_{n-1}A_1^T + B_{22}. \end{aligned} \quad (5.12)$$

Then, by Theorem 5.1 we have

$$\begin{aligned}\widehat{X}_n &= a_1\widehat{X}_{n-1} + a_2Y_{n-1} \\ &\quad + (a_1P_{n-1}A_1^T + B_{12})(A_1P_{n-1}A_1^T + B_{22})^{-1} \\ &\quad \times \left\{ Y_n - (A_1\widehat{X}_{n-1} + A_2Y_{n-1}) \right\}\end{aligned}\quad (5.13)$$

and

$$\begin{aligned}P_n &= a_1P_{n-1}a_1^T + B_{11} \\ &\quad - (a_1P_{n-1}A_1^T + B_{12})(A_1P_{n-1}A_1^T + B_{22})^{-1}(a_1P_{n-1}A_1^T + B_{12})^T.\end{aligned}\quad (5.14)$$

The recurrent equations (5.13), (5.14) supplied by the initial conditions (5.11) constitute the Kalman filter.

5.1.1. **Example.** In many books, a typical model discussed for Kalman's filtering is the following:

$$\begin{aligned}X_n &= aX_{n-1} + \varepsilon_n^1 \\ Y_n &= HX_n + \varepsilon_n^2,\end{aligned}\quad (5.15)$$

where a and H are known matrices and $\varepsilon_n^1, \varepsilon_n^2$ are white noises with $\mathbf{E}\varepsilon_n^1 \equiv 0, \mathbf{E}\varepsilon_n^2 \equiv 0$ and

$$\mathbf{E}\varepsilon_n^1(\varepsilon_n^1)^T \equiv D_{11}, \mathbf{E}\varepsilon_n^2(\varepsilon_n^2)^T \equiv D_{22}, \mathbf{E}\varepsilon_n^1(\varepsilon_n^2)^T \equiv 0;$$

we add only ε_0^2 orthogonal to X_0 . $\mathbf{E}X_0$ and $\text{Cov}(X_0, X_0)$ are assumed to be known. Since $Y_0 = HX_0 + \varepsilon_0^2$. Hence, we have

$$\mathbf{E}Y_0 = H\mathbf{E}X_0,$$

$$\text{Cov}(X_0, Y_0) = H\text{Cov}(X_0, X_0),$$

$$\text{Cov}(Y_0, Y_0) = H\text{Cov}(X_0, X_0)H^T + D_{22}.$$

Further, from (5.15) it follows that

$$Y_n = HaX_{n-1} + \varepsilon_n^1 + \varepsilon_n^2.$$

Now, the model given in (5.15) is completely adapted to (5.8) with $a_1 = a, a_2 = 0, A_1 = Ha, A_2$ and $B_{11} = D_{11}, B_{12} = D_{11}, B_{22} = D_{11} + D_{22}$.

Thus, we have recurrent equations

$$\begin{aligned}\widehat{X}_n &= a\widehat{X}_{n-1} \\ &\quad + (aP_{n-1}(Ha)^T + D_{11})(HaP_{n-1}(Ha)^T + D_{11} + D_{22})^{-1} \\ &\quad \times \left\{ Y_n - Ha\widehat{X}_{n-1} \right\}\end{aligned}$$

and

$$P_n = aP_{n-1}a^T + D_{11} - (aP_{n-1}(Ha)^T + D_{12}) \\ \times (HaP_{n-1}(Ha)^T + D_{11} + D_{22})^{-1}(aP_{n-1}(Ha)^T + D_{12})^T$$

subject to

$$\widehat{X}_0 = \mathbf{E}X_0 + HCov(X_0, X_0) \left\{ HCov(X_0, X_0)H^T + D_{22} \right\}^{-1} (Y_0 - \mathbf{E}Y_0) \\ P_0 = Cov(X_0, X_0) \\ - HCov(X_0, X_0) \left\{ HCov(X_0, X_0)H^T + D_{22} \right\}^{-1} Cov(X_0, X_0)H^T.$$

5.2. Moore-Penrose pseudoinverse matrix. Any $A = A_{m \times n}$ matrix possesses the unique matrix $A^\oplus = A_{n \times m}^\oplus$ defined by two conditions:

1. $AA^\oplus A = A$
2. $A^\oplus = UA^T$ and $A^\oplus = A^T V$ for some matrices U, V .

A^\oplus is named Moore-Penrose pseudoinverse matrix of A and defined as follows: if r ($\leq \min(n, m)$) is the rank of A , then there exist two matrices $B_{n \times r}$ and $C_{r \times m}$ of rank r such that $A = BC$. Then

$$A^\oplus = C^\oplus B^\oplus,$$

where

$$C^\oplus = C^T(CC^T)^{-1} \quad \text{and} \quad B^\oplus = (B^T B)^{-1} B^T.$$

Below we give a list of main properties of A^\oplus .

- (1) $AA^\oplus A = A$, $A^\oplus AA^\oplus = A^\oplus$
- (2) $(A^T)^\oplus = (A^\oplus)^T$
- (3) $(A^\oplus)^\oplus = A$
- (4) $(A^\oplus A)^2 = A^\oplus A$, $(A^\oplus A)^T = (A^\oplus A)$, $(AA^\oplus)^2 = AA^\oplus$, $(AA^\oplus)^T = AA^\oplus$
- (5) $(A^T A)^\oplus = A^\oplus (A^T)^\oplus = A^\oplus (A^\oplus)^T$
- (6) $A^\oplus = (A^T A)^\oplus A^T = A^T (AA^T)^\oplus$
- (7) $A^\oplus AA^T = A^T AA^\oplus = A^T$
- (8) if S is orthogonal matrix ($S^T = S^{-1}$), then $(SAS^T)^\oplus = SA^\oplus S^T$
- (9) if $A_{n \times n}$ is a symmetric nonnegative definite matrix of rank $r < n$ and $A = T^T T$, where $T_{r \times n}$ of rank r , then

$$A^\oplus = T^T (TT^T)^{-2} T.$$

Proposition 5.2. Let x, y be random vectors with correlation matrices $Cov(x, y)$ and $Cov(y, y)$ (singular). Then,

$$Cov(x, y) \left(I - Cov^\oplus(y, y)Cov(y, y) \right) = 0,$$

where I is the unite matrix.

Proof. Without loss of a generality, one can assume that $\mathbf{E}x = 0$ and $\mathbf{E}y = 0$. The use of identity

$$Cov(x, y) \left(I - Cov(y, y)Cov^\oplus(y, y) \right) = \mathbf{E}(x, y^T) \left(I - Cov(y, y)Cov^\oplus(y, y) \right)$$

gives a hint that the desired result holds true provided that

$$\mathbf{E} \left(\left[I - \text{Cov}(y, y) \text{Cov}(y, y)^\oplus \right] y \right) \left(\left[I - \text{Cov}(y, y) \text{Cov}(y, y)^\oplus \right] y \right)^T = 0.$$

Write

$$\begin{aligned} & \mathbf{E} \left(\left[I - \text{Cov}(y, y) \text{Cov}(y, y)^\oplus \right] y \right) \left(\left[I - \text{Cov}(y, y) \text{Cov}(y, y)^\oplus \right] y \right)^T \\ &= \left[I - \text{Cov}(y, y) \text{Cov}(y, y)^\oplus \right] \text{Cov}(y, y) \left(\left[I - \text{Cov}(y, y) \text{Cov}(y, y)^\oplus \right] \right)^T \\ &= \left[\text{Cov}(y, y) - \text{Cov}(y, y) \text{Cov}(y, y)^\oplus \text{Cov}(y, y) \right] \\ & \quad \times \left(\left[I - \text{Cov}(y, y) \text{Cov}^\oplus(y, y) \right] \right)^T = 0, \text{ by property 1.} \end{aligned}$$

□