

Lecture 1. Linear Filtering in Discrete Time 1.1

On a probability space $(\Omega, \mathcal{F}, P)$, consider a pair of square integrable r.v. (X, Y) . We know

$$E[X], E[Y]$$

$$\text{var}(X) = E[(X - E[X])^2], \text{var}(Y) = E[(Y - E[Y])^2]$$

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

We are asked to find constants a'_0 and a'_1 , such that

$$E[(X - a'_0 - a'_1 Y)^2] \leq E[(X - a_0 - a_1 Y)^2] \quad \forall a_0, a_1 \in \mathbb{R}$$

Then, $\hat{X} := a'_0 + a'_1 Y$ is the optimal linear estimate of X given Y .

We can write

$$E[(X - a_0 - a_1 Y)^2] = E[(X - E[X] - a_1(Y - E[Y]) + E[X] - a_1 E[Y] - a_0)^2]$$

$$= E[(X - E[X] - a_1(Y - E[Y]))^2] \quad \text{to}$$

$$+ E[(X - E[X] - a_1(Y - E[Y]))(E[X] - a_1 E[Y] - a_0)]$$

$$+ (E[X] - a_1 E[Y] - a_0)^2$$

$$=: \psi(a_1) + \xi(a_1, a_0)$$

Given a_1 , we can always make $\xi(a_1, a_0) = 0$ by choosing

$$a'_0 = E[X] - a_1 E[Y]$$

On the other hand, we have

$$\psi(a_1) = \text{var}(X) - 2a_1 \text{cov}(X, Y) + a_1^2 \text{var}(Y)$$

and we can minimize ψ over a_1 .

$$0 = \psi'(a_1) = -2 \text{cov}(X, Y) + 2a_1 \text{var}(Y)$$

$$\Rightarrow a'_1 = \frac{\text{cov}(X, Y)}{\text{var}(Y)} \quad \text{if } \text{var}(Y) > 0$$

Then

$$a'_0 = E[X] - \frac{\text{cov}(X, Y)}{\text{var}(Y)} E[Y], \quad a'_1 = \frac{\text{cov}(X, Y)}{\text{var}(Y)}$$

If $\text{var}(Y) = 0$ (or in other words $Y = E[Y]$)

we have that $a'_1 = 0$, $a'_0 = E[X]$

For any $x \in \mathbb{R}$ define

$$x^\oplus = \begin{cases} 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Then, among all linear functionals of $\{1, Y\}$, there is the unique optimal one, given by

$$(1.1) \quad \hat{X} := E[X] + \text{cov}(X, Y) \text{var}(Y)^\oplus (Y - E[Y])$$

with the corresponding minimal M.S.E.

$$E[(X - \hat{X})^2] = \text{var}(X) - \text{cov}^2(X, Y) \text{var}(Y)^\oplus$$

Note that the optimal estimate satisfies the orthogonality property

$$E[(X - \hat{X})1] = 0$$

$$E[(X - \hat{X})Y] = 0,$$

that is, the residual estimation error is

orthogonal to any affine functional of L.3
the observations.

These simple formulas are the basis for
the Kalman-Bucy filter.

1.1. Orthogonal projection and estimation

Let $L^2 = L^2(\Omega, \mathcal{F}, P)$ denote the space of all square integrable
random variables.

L^2 is a Hilbert space with the inner product

$$\langle X, Y \rangle := E[XY] \quad , \quad X, Y \in L^2.$$

and the induced norm $\|X\| := \langle X, X \rangle^{1/2}$.

Theorem 1.1 (Orthogonal projection) Let $\mathcal{L}$ be a closed linear subspace
of L^2 and $X \in L^2$. Then

i) $\exists!$ $\hat{X} \in \mathcal{L}$ such that

$$(1.2) \quad E[(X - \hat{X})^2] = \inf_{Z \in \mathcal{L}} E[(X - Z)^2].$$

ii) $\hat{X} \in \mathcal{L}$ and (1.2) holds $\iff \hat{X} \in \mathcal{L}$ and

$$(1.3) \quad X - \hat{X} \in \mathcal{L}^\perp$$

where $\mathcal{L}^\perp$ is the orthogonal complement of $\mathcal{L}$ in L^2 , i.e.,

$$\mathcal{L}^\perp = \{ Y \in L^2 : E[YZ] = 0, \forall Z \in \mathcal{L} \}$$

The element $\hat{X}$ is called the orthogonal projection
of X onto $\mathcal{L}$.

We will also use the notation $\hat{E}[X|\mathcal{L}]$ for $\hat{X}$.

Proof

1.4

ii) $\Rightarrow$) Suppose that $\hat{x} \in \mathcal{H}$ and satisfies (1.2).

We want to show that $\langle x - \hat{x}, z \rangle = 0, \forall z \in \mathcal{H}$.

Suppose, on the contrary that $\forall \in \mathcal{H}$ is such that

$\langle x - \hat{x}, y \rangle = \delta \neq 0$ and w.l.o.g. $\|y\| = 1$.

Define the vector $\hat{y} := \hat{x} + \delta y$. We then have that

$$\begin{aligned} \|x - y\|^2 &= \|x - \hat{x} - \delta y\|^2 \\ &= \|x - \hat{x}\|^2 - 2\delta \underbrace{\langle x - \hat{x}, y \rangle}_{\delta} + \delta^2 \|y\|^2 \\ &= \|x - \hat{x}\|^2 - \delta^2 < \|x - \hat{x}\|^2 !!! \end{aligned}$$

this proves this implication.

$\Leftarrow$) Suppose that $\hat{x} \in \mathcal{H}$ and $x - \hat{x} \in \mathcal{H}^\perp$

Then, for any $z \in \mathcal{H}$ we have

$$\begin{aligned} \|x - z\|^2 &= \|x - \hat{x} + \hat{x} - z\|^2 \\ &= \|x - \hat{x}\|^2 + \|\hat{x} - z\|^2 \quad \left(\begin{array}{l} \text{because } \hat{x} - z \in \mathcal{H} \\ \text{and } x - \hat{x} \in \mathcal{H}^\perp \end{array} \right) \\ &> \|x - \hat{x}\|^2 \quad \text{for } \hat{x} \neq z \end{aligned}$$

1) Existence of $\hat{x}$

Let $d^2 := \inf_{z \in \mathcal{H}} E[(x - z)^2]$ and let $(z_j)_{j \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $d_j^2 := E[(x - z_j)^2] \rightarrow d^2$.

Note that

$$\|z_i - z_j\|^2 = \|z_i - x\|^2 + \|z_j - x\|^2 - 2 \langle z_i - x, z_j - x \rangle$$

and

$$4 \left\| \frac{z_i + z_j}{2} - x \right\|^2 = \|z_i - x\|^2 + \|z_j - x\|^2 + 2 \langle z_i - x, z_j - x \rangle,$$

which yields

$$\begin{aligned} \|z_i - z_j\|^2 &= 2\|z_i - x\|^2 + 2\|z_j - x\|^2 - 4 \left\| \frac{z_i + z_j}{2} - x \right\|^2 \\ &\leq 2\|z_i - x\|^2 + 2\|z_j - x\|^2 - 4d \xrightarrow{j \rightarrow \infty} 0 \end{aligned}$$

where the inequality holds because $z_i + z_j \in \mathcal{L}$.
 Then $(z_j)_{j \geq 1}$ is a Cauchy sequence in L^2 .

Since the space L^2 is complete, $(z_j)_{j \geq 1}$ converges to a r.v. z_∞ in L^2 . Moreover, since $\mathcal{L}$ is closed, $z_\infty \in \mathcal{L}$.

then

$$\|x - z_\infty\| \leq \|x - z_j\| + \|z_j - z_\infty\| \xrightarrow{j \rightarrow \infty} 0$$

and we have that z_∞ is a version of $\hat{X}$.

Uniqueness of $\hat{X}$

Suppose that $\tilde{X} \in \mathcal{L}$ also satisfies (1.2). Then,

$$\begin{aligned} E[(X - \tilde{X})^2] &= E[(X - \hat{X} + \hat{X} - \tilde{X})^2] \\ &= E[(X - \hat{X})^2] + E[(\hat{X} - \tilde{X})^2] \end{aligned}$$

$$+ 2 E[(X - \hat{X})(\hat{X} - \tilde{X})]$$

0 because $\hat{X} - \tilde{X} \in \mathcal{H}$ and (1.3) holds due to i).

But this implies $E[(\hat{X} - \tilde{X})^2] = 0 \Rightarrow \hat{X} = \tilde{X}$
P-a.s.

Properties of the orthogonal projection

a) $E[\hat{E}[X|\mathcal{H}]] = E[X]$

b) $\hat{E}[X|\mathcal{H}] = X$ if $X \in \mathcal{H}$ and
 $\hat{E}[X|\mathcal{H}] = 0$ if $X \perp \mathcal{H}$.

c) Linearity: For $X_1, X_2 \in L^2$, $c_1, c_2 \in \mathbb{R}$,
 $\hat{E}[c_1 X_1 + c_2 X_2 | \mathcal{H}] = c_1 \hat{E}[X_1 | \mathcal{H}] + c_2 \hat{E}[X_2 | \mathcal{H}]$

d) For two linear subspaces $\mathcal{H}_1 \subseteq \mathcal{H}_2$,
 $\hat{E}[X | \mathcal{H}_2] = \hat{E}[\hat{E}[X | \mathcal{H}_2] | \mathcal{H}_1]$

Proof

a) - c) Follow easily from the definition.

Using Theorem 1.1. ii) we have that d) holds if

$$E[(X - \hat{E}[\hat{E}[X | \mathcal{H}_2] | \mathcal{H}_1])Z] = 0 \quad \forall Z \in \mathcal{H}_1,$$

which holds because we can write

$$E[(X - \hat{E}[\hat{E}[X | \mathcal{H}_2] | \mathcal{H}_1])Z] =$$

$$= E[(X - \hat{E}[X|h_2])z]$$

$$+ E[(\hat{E}[X|h_2] - \hat{E}[\hat{E}[X|h_2]|h_1])z] = \textcircled{1} + \textcircled{2}$$

$$\textcircled{1} = 0 \text{ by (L.3) with } h = h_2 \text{ and } z \in h_2 \subseteq h_2$$

$$\textcircled{2} = 0 \text{ by (L.3) with } X = \hat{E}[X|h_2] \text{ and } h = h_1 \\ \text{and } z \in h_2$$

Remark Theorem 1.1. suggests that the optimal **linear** estimate in the mean square sense of a random variable $X \in L^2$ from the observation of a collection of random variables $Y_j \in L^2$, $j \in J \subseteq \mathbb{Z}_+$ is given by the orthogonal projection of X onto $\mathcal{L}_J^Y := \overline{\text{span}} \{ Y_j, j \in J \}$.

When $|J| < \infty$, the explicit expression for $\hat{E}[X | \mathcal{L}_J^Y]$ is given in the following proposition.

For $|J| = +\infty$, the situation is more involved.

Let A be a positive (semi-)definite matrix, i.e., A is symmetric and $x^T A x \geq 0$.

Example = A covariance matrix $\text{Cov}(Y)$,
 Y being a random vector.

The eigenvalue decomposition for a real symmetric matrix A is given by

$A = U D U^T$, where D is a diagonal matrix whose entries are the eigenvalues of A , and U is an orthogonal matrix ($U U^T = I_d$) whose columns are the eigenvectors of A .

The Moore-Penrose inverse (or pseudoinverse, or generalized inverse) of A is denoted by $A^\oplus$, is given by

$$A^\oplus := U D^\oplus U^T$$

where $D^\oplus$ is the diagonal matrix with entries

$$D^\oplus_{ii} = \begin{cases} 1/D_{ii}, & D_{ii} > 0 \\ 0 & D_{ii} = 0 \end{cases}$$

if $A > 0$, then $A^\oplus = A^{-1}$.

Proposition 1.2 Let $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ be random vectors with square integrable entries. Denote by $\hat{E}[X | \mathcal{L}^Y]$ the orthogonal projection of X onto the linear subspace, spanned by the entries of Y and 1. Then

$$(1.4) \quad \hat{E}[X | \mathcal{L}^Y] = E[X] + \text{cov}(X, Y) \text{cov}(Y)^\oplus (Y - E[Y])$$

and

$$(L.5) \quad E[(X - \hat{E}[X|R^Y])(X - \hat{E}[X|R^Y])^T] = \text{cov}(X) - \text{cov}(X, Y) \text{cov}(Y)^{\oplus} (Y - E[Y]).$$

Proof (Sketch)

Let A be a matrix and a be a vector such that

$$\hat{E}[X|R^Y] = a + AY.$$

Then, applying Theorem 1.1. componentwise

we get

$$E[X - a - AY] = 0 \Rightarrow \boxed{a = E[X] - A E[Y]} \quad (*)$$

and

$$\begin{aligned} 0 &= E[(X - a - AY)(Y - E[Y])^T] \\ &= E[(X - E[X] - A(Y - E[Y]) - a + E[X] - A E[Y])(Y - E[Y])^T] \end{aligned}$$

$$(L.6) = \text{cov}(X, Y) - A \text{cov}(Y)$$

Equation (L.4) follows by checking that if we let $A = \text{cov}(X, Y) \text{cov}(Y)^{\oplus}$, then (L.6) holds (combined with (*))

If $\text{cov}(Y) > 0$, then $\text{cov}(Y)^{\oplus} = \text{cov}(Y)^{-1}$ and

$A = \text{cov}(X, Y) \text{cov}(Y)^{-1}$. That equation (L.6) holds, follows trivially.

If $\text{cov}(Y)$ is only positive semi-definite ($\text{cov}(Y) \geq 0$), the proof is more involved. We know there exist U such that $UU^T = I_d$ and D a diagonal matrix $D \geq 0$ such that $\text{cov}(Y) = U D U^T$ and $\text{cov}(Y)^{\oplus} = U D^{\oplus} U^T$.

Then,

$$\text{cov}(X, Y) - \text{cov}(X, Y) \text{cov}(Y)^{\oplus} \text{cov}(Y)$$

$$= \text{cov}(X, Y) (\text{Id} - \text{cov}(Y)^{\oplus} \text{cov}(Y))$$

1.10

$$= \text{cov}(X, Y) (\mu \mu^T - \mu \underbrace{D^{\oplus} \mu^T}_{\text{Id}} \mu D \mu^T)$$

$$= \text{cov}(X, Y) (\mu \text{Id} \mu^T - \mu D^{\oplus} D \mu)$$

$$(1.7) = \text{cov}(X, Y) \mu (\text{Id} - D^{\oplus} D) \mu^T = \sum_{l: D_{ll} = 0} \text{cov}(X, Y) \mu_l \mu_l^T$$

by the definition of $D^{\oplus}$, where μ_l is the l -th column of μ .

Since μ_l is the eigenvector of $D_{ll} = 0$, we have that

$$\mu_l^T \text{cov}(Y) \mu_l = 0 \Rightarrow \mu_l^T E[(Y - E[Y])(Y - E[Y])^T] \mu_l = 0$$

$$\Rightarrow E[\mu_l^T (Y - E[Y])(Y - E[Y])^T \mu_l] = 0$$

$$\Rightarrow E[|(Y - E[Y])^T \mu_l|_{\ell_m}^2] = 0 \Rightarrow (Y - E[Y])^T \mu_l = 0 \quad \text{p.a.}$$

Hence,

$$\text{cov}(X, Y) \mu_l = E[(X - E[X])(Y - E[Y])^T \mu_l] = 0,$$

and (1.7) (and therefore (1.6)) holds.

Equation (1.5) is verified by substitution of (1.4) and using the properties of the generalised inverse.

2. Recursive orthogonal projection

Consider a pair of processes $(X, Y) = (X_j, Y_j)_{j \in \mathbb{Z}_+}$ with entries in L^2 and let $\mathcal{H}_j^Y = \overline{\text{span}}\{1, Y_0, \dots, Y_j\}$.

Computing $\hat{E}[X_j | \mathcal{H}_j^X]$ by (directly) using the formulae in Proposition 2.1. is inefficient. It would require inverting matrices of sizes growing linearly with j .

The following lemma is key to perform the orthogonal projection algorithm efficiently.

We first introduce some notation.

$$\hat{X}_j := \hat{E}[X_j | \mathcal{L}_j^X], \quad P_j^X := E[(X - \hat{X}_j)(X - \hat{X}_j)^T]$$

$$\hat{X}_{j|j-2} := \hat{E}[X_j | \mathcal{L}_{j-2}^X]$$

$$\hat{Y}_{j|j-2} := \hat{E}[Y_j | \mathcal{L}_{j-2}^Y]$$

$$P_{j|j-2}^X := E[(X_j - \hat{X}_{j|j-2})(X_j - \hat{X}_{j|j-2})^T]$$

$$P_{j|j-2}^Y := E[(Y_j - \hat{Y}_{j|j-2})(Y_j - \hat{Y}_{j|j-2})^T]$$

$$P_{j|j-2}^{XY} := E[(X_j - \hat{X}_{j|j-2})(Y_j - \hat{Y}_{j|j-2})^T]$$

Proposition 1.4 For $j \geq 1$, we have

$$(1.8) \quad \hat{X}_j = \hat{X}_{j|j-2} + P_{j|j-2}^{XY} (P_{j|j-2}^Y)^{\oplus} (Y_j - \hat{Y}_{j|j-2})$$

and

$$(1.9) \quad P_j^X = P_{j|j-2}^X - P_{j|j-2}^{XY} (P_{j|j-2}^Y)^{\oplus} (P_{j|j-2}^{XY})^T$$

Proof (Sketch)

To verify (1.8), we can check that

$$R := X_j - \hat{X}_{j|j-2} + P_{j|j-2}^{XY} (P_{j|j-2}^Y)^{\oplus} (Y_j - \hat{Y}_{j|j-2})$$

is orthogonal to $\mathcal{L}_j^Y$.

Since $x_j - \hat{x}_{j|j-1}$ and $y_j - \hat{y}_{j|j-1}$ are orthogonal to L_{j-1}^Y ,
 then k is orthogonal to L_{j-1}^Y .

Since any $z \in L_j^Y$ can be written as

$$z = \underbrace{a_0 + a_1 y_1 + \dots + a_{j-1} y_{j-1}}_{\substack{\uparrow \\ L_{j-1}^Y}} + a_j y_j$$

We have that $\langle k, z \rangle = a_j \langle k, y_j \rangle$ and
 k is orthogonal to L_j^Y if $k \perp y_j$.

Moreover, a trivial corollary of Theorem 1.1,
 yields that $y_j = \underbrace{\hat{y}_{j|j-1}}_{L_{j-1}^Y} + \underbrace{y_j - \hat{y}_{j|j-1}}_{L_{j-1}^Y \perp}$

And

$$0 = \langle k, y_j \rangle = \underbrace{\langle k, \hat{y}_{j|j-1} \rangle}_0 + \langle k, y_j - \hat{y}_{j|j-1} \rangle$$

That is, $k \perp y_j \Leftrightarrow k \perp y_j - \hat{y}_{j|j-1}$.

Then,

$$\begin{aligned} E[k(y_j - \hat{y}_{j|j-1})] &= P_{j|j-1}^{xy} - P_{j|j-1}^{xy} (P_{j|j-1}^y)^{\oplus} P_{j|j-1}^y \\ &= P_{j|j-1}^{xx} (I_d - (P_{j|j-1}^y)^{\oplus} P_{j|j-1}^y) \stackrel{(*)}{=} 0 \end{aligned}$$

Where (*) is verified as in (1.7).

Equation (1.9) is obtained similarly to (1.5).

3. the Kalman - Bucy filter in discrete time

Consider a pair of processes $(X, Y) = (X_j, Y_j)_{j \geq 0}$ generated by the linear recursive equations

$$(2.10) \quad X_j = a_0(j) + a_1(j) X_{j-1} + a_2(j) Y_{j-1} + b_1(j) \varepsilon_j + b_2(j) \zeta_j$$

$$(2.11) \quad Y_j = A_0(j) + A_1(j) X_{j-1} + A_2(j) Y_{j-1} + B_1(j) \varepsilon_j + B_2(j) \zeta_j$$

for $j \geq 1$, where

- $X_j \in \mathbb{R}^m$ and $Y_j \in \mathbb{R}^n$.
- $\varepsilon = (\varepsilon_j)_{j \geq 1}$ and $\zeta = (\zeta_j)_{j \geq 1}$ are orthogonal (discrete time) white noises with values in $\mathbb{R}^l$ and $\mathbb{R}^k$, i.e.,

$$E[\varepsilon_j] = 0 \quad E[\varepsilon_j \varepsilon_j^T] = \begin{cases} I_l & i=j \\ 0 & i \neq j \end{cases} \in \mathbb{R}^{l \times l}$$

$$E[\zeta_j] = 0 \quad E[\zeta_j \zeta_j^T] = \begin{cases} I_k & i=j \\ 0 & i \neq j \end{cases} \in \mathbb{R}^{k \times k}$$

and

$$E[\varepsilon_j \zeta_j^T] = 0 \quad \forall j \geq 0.$$

- The coefficients $a_0(j), a_1(j), \dots$ are **deterministic (known)** sequences of matrices of appropriate dimension. From now on, for brevity, we will omit the time dependency.
- The initial conditions X_0 and Y_0 are random with known means and covariances and independent of ε and ζ .

Denote the optimal linear estimate of X_j , given (L.14)

$$h_j^y = \overline{\text{span}} \{1, Y_1, \dots, Y_j\}, \text{ by}$$

$$\hat{X}_j = \hat{E}[X_j | h_j^y]$$

and the corresponding error covariance matrix by

$$P_j = E[(X_j - \hat{X}_j)(X_j - \hat{X}_j)^T]$$

Theorem 2.5 The estimate $\hat{X}_j$ and the error covariance P_j satisfy the equations

$$(L.12) \quad \hat{X}_j = a_0 + a_1 \hat{X}_{j-1} + a_2 Y_{j-1} + (a_1 P_{j-1} A_1^T + b_0 B)(A_1 P_{j-1} A_1^T + B_0 B)^{\oplus} (Y_j - A_0 - A_1 \hat{X}_{j-1} - A_2 Y_{j-1})$$

$$(L.13) \quad P_j = a_1 P_{j-1} A_1^T + b_0 b_0^T - (a_1 P_{j-1} A_1^T + b_0 B)(A_1 P_{j-1} A_1^T + B_0 B)^{\oplus} (a_1 P_{j-1} A_1^T + b_0 B)^T,$$

where

$$b_0 b_0^T = b_1 b_1^T + b_2 b_2^T, \quad b_0 B = b_1 B_1^T + b_2 B_2^T, \quad B_0 B = B_1 B_1^T + B_2 B_2^T$$

and

$$\hat{X}_0 = E[X_0] + \text{cov}(X_0, Y_0) \text{cov}(Y_0)^{\oplus} (Y_0 - E[Y_0])$$

$$P_0 = \text{cov}(X_0) - \text{cov}(X_0, Y_0) \text{cov}(Y_0)^{\oplus} \text{cov}(X_0, Y_0)^T$$

Proof (Very sketchy idea)

Apply the formulae of Proposition 2.4 and the properties of orthogonal projection. For example,

$$\hat{X}_{j|j-1} = \hat{E}[a_0 + a_1 X_{j-1} + a_2 Y_{j-1} + b_1 \varepsilon_j + b_2 \eta_j | h_{j-1}^y]$$

$$\stackrel{(\oplus)}{=} a_0 + a_1 \hat{E}[X_{j-1} | h_{j-1}^y] + a_2 Y_{j-1} = a_0 + a_1 \hat{X}_{j-1} + a_2 Y_{j-1}$$

where the equality (*) holds since ε_j and η_j are L-11 orthogonal to $\mathcal{L}_{j-2}^\perp$.

And proceed similarly with the rest of the families.

Example 1.6 Consider an autoregressive scalar signal generated by

$$X_j = a X_{j-1} + \varepsilon_j, \quad X_0 = 0,$$

where a is a constant and ε is a white noise sequence.

Suppose the observation process is given by

$$Y_j = X_{j-1} + \eta_j$$

where η is another white noise sequence, orthogonal to ε .

Applying the equations from Theorem 1.5 one gets

$$\hat{X}_j = a \hat{X}_{j-1} + \frac{a P_{j-1}}{P_{j-1} + 1} (Y_j - \hat{X}_{j-1}), \quad \hat{X}_0 = 0$$

where

$$P_j = a^2 P_{j-1} + \frac{a^2 P_{j-1}}{P_{j-1} + 1}, \quad P_0 = 0.$$

3.1. Properties of the Kalman-Bay filter

[L.] The equation for P_j is called difference (discrete time) Riccati equation.

It does not depend on the observations and so can be solved off-line (before the filter is applied to the data).

Even if all the coefficients of the system (L.10) and (L.11) are constant (do not depend on j), the

optimal filter has in general time varying coefficients, (because it depends on P_j). 2.16

② $\exists!$ and strict positiveness of the limit $P_\infty := \lim_{j \rightarrow \infty} P_j$ is a difficult question in general.

If the limit exists and is unique, one can use the stationary version of the filter. That means, to replace P_{j-1} with P_∞ in computing $\hat{X}_j$. In this case, the error matrix of this "suboptimal" filter converges to P_∞ as well.

③ The propagation of $\hat{X}_j$ and P_j is sometimes regarded in two stages:

Prediction

$$\hat{X}_{j|j-1} = a_0 + a_1 \hat{X}_{j-1} + a_2 Y_{j-1}$$

$$\hat{Y}_{j|j-1} = A_0 + A_1 \hat{X}_{j-1} + A_2 Y_{j-1}$$

Update

$$\hat{X}_j = \hat{X}_{j|j-1} + K_j (Y_j - \hat{Y}_{j|j-1})$$

where

$$K_j := (a_1 P_{j-1} A_1^T + b \circ B) (A_1 P_{j-1} A_1^T + B \circ B)^{-1}$$

i) called the Kalman gain matrix.

Similar interpolation is possible for P_j .

④ the sequence $\bar{\varepsilon}_j := Y_j - A_0 - A_1 \hat{X}_{j-1} - A_2 Y_{j-1}$

turns out to be an orthogonal sequence and is called the "innovations": It is the residual "information" borne by Y_j after its prediction on the basis past information is subtracted