6. KALMAN FILTER

Consider the discrete-time system

$$X(k+1) = A(k)X(k) + B(k)U(k) + \tilde{F}(k)N_k'$$

$$X(k) \in \mathbb{R}^n, \quad U(k) \in \mathbb{R}^p$$

$$Y(k+1) = C(k)X(k) + D(k)U(k) + \tilde{G}(k)N_k''$$

$$Y(k) \in \mathbb{R}^m, \qquad (S)$$

where $\{N_k'\}$, $\{N_k''\}$ are sequences of random vectors, $U(k)$ are the deterministic inputs.

Our problem is to estimate $X(j)$ from the observations $\{Y(0), Y(1), \dots, Y(k)\}$. If $j > k$ we are predicting, if $j = k$ we are filtering and if $j < k$ we are interpolating.

REMARK. We may assume that $E\{N_k'\} = 0$ and $E\{N_k''\} = 0$.

If not simply redefine

$\overline{N}_k' \triangleq N_k' - E\{N_k'\}$, $\overline{B}(k) \triangleq [B(k) \vdots F(k) \vdots 0]$

$\overline{N}_k'' \triangleq N_k'' - E\{N_k''\}$, $\overline{D}(k) \triangleq [D(k) \vdots 0 \vdots \tilde{G}(k)]$

$$\overline{U}(k) \triangleq \begin{pmatrix} U(k) \\ E\{N_k'\} \\ E\{N_k''\} \end{pmatrix}$$

and rewrite (S) as

$$\begin{cases} X(k+1) = A(k)X(k) + \overline{B}(k)\overline{U}(k) + \tilde{F}(k)\overline{N}_k' \\ Y(k) = C(k)X(k) + \overline{D}(k)\overline{U}(k) + \tilde{G}(k)\overline{N}_k'' \end{cases}$$

where $\overline{N}_k'$, $\overline{N}_k''$ have zero mean.

First of all, to solve our problem, we will split the state $X(k)$ into a deterministic and stochastic component

$$X(k) = X_d(k) + X_s(k)$$

$X_d(k)$: it depends only from $\{U(k)\}$

$X_s(k)$: it depends only from $\{N_k'\}$

$X_s(k) \triangleq X(k) - E\{X(k)\}$

$X_d(k) \triangleq E\{X(k)\}$

Notice that $E\{X_s(k)\} = 0$.
Moreover, the evolution of $X_d(k)$ is given by

$$\boxed{X_d(k+1)} = E\{X(k+1)\} = E\{A(k)X(k) + B(k)U(k) + \tilde{F}(k)N_k'\} = \boxed{A(k)X_d(k) + B(k)U(k)}$$

$$\boxed{X_d(0) = E\{X(0)\}}$$

and

$$\boxed{X_s(k+1)} \doteq X(k+1) - X_d(k+1)$$
$$= \boxed{A(k)X_s(k) + \tilde{F}(k)N_k'}$$

$$\boxed{X_s(0) = X(0) - E\{X(0)\}}$$

Similarly, $Y(k) \doteq Y_d(k) + Y_s(k)$
where

$$\boxed{Y_d(k) = C(k)X_d(k) + D(k)U(k)}$$
$$\boxed{Y_s(k) = C(k)X_s(k) + \tilde{G}(k)N_k''}$$

Notice $E\{Y_s(k)\} = 0$

REMARK We also can assume that the covariances of N'_k, N''_k are the identity matrices. If not:

$$E\{N'_k N'^T_k\} \triangleq \Psi'_k > 0, \quad E\{N''_k N''^T_k\} \triangleq \Psi''_k > 0$$

Consider the new sequences

$$\bar{N}'_k \triangleq (\Psi'_k)^{-\frac{1}{2}} N'_k, \quad \bar{N}''_k \triangleq (\Psi''_k)^{-\frac{1}{2}} N''_k$$

with

$$\bar{F}(k) \triangleq \tilde{F}(k)(\Psi'_k)^{\frac{1}{2}}, \quad \bar{G}(k) \triangleq \tilde{G}(k)(\Psi''_k)^{\frac{1}{2}}$$

It follows

$$E\{\bar{N}'_k \bar{N}'^T_k\} = (\Psi'_k)^{-\frac{1}{2}} E\{N'_k N'^T_k\} (\Psi'_k)^{-\frac{1}{2}} = I$$

$$E\{\bar{N}''_k \bar{N}''^T_k\} = (\Psi''_k)^{-\frac{1}{2}} E\{N''_k N''^T_k\} (\Psi''_k)^{-\frac{1}{2}} = I$$

The contribution of $\bar{N}'_k, \bar{N}''_k$ on the measurements and states is the same as N'_k, N''_k:

$$\bar{F}(k)\bar{N}'_k = \tilde{F}(k)N'_k, \quad \bar{G}(k)\bar{N}''_k = \tilde{G}(k)N''_k$$

Clearly, to estimate $X(k)$ it is sufficient to estimate $X_s(k)$ and then summing $X_d(t)$ to it.

To simplify further the representation of the stochastic system

$$\begin{cases} X_s(k+1) = A(k)X_s(k) + \tilde{F}(k)N_k' \\ X_s(0) = X(0) - E\{X(0)\} \\ Y_s(k) = CX_s(k) + \tilde{G}(k)N_k'' \end{cases} \quad (SS)$$

we will define

$$N_k = \begin{pmatrix} N_k' \\ N_k'' \end{pmatrix} \qquad \begin{array}{l} F(k) = (\tilde{F}(k) \mid 0) \\ G(k) = (0 \mid \tilde{G}(k)) \end{array}$$

so that (SS) becomes

$$\begin{cases} X_s(k+1) = A(k)X_s(k) + F(k)N_k \\ X_s(0) = X(0) - E\{X(0)\} \\ Y_s(k) = C(k)X_s(k) + G(k)N_k \end{cases} \quad (SSF)$$

We will assume that $\{N_k\}$ is white and gaussian with

$$\begin{cases} E\{N_k N_j^T\} = \delta_{kj} I \\ E\{N_k\} = 0 \end{cases} \qquad \forall k, j$$

(also $F(k)N_k$ and $G(j)N_j$, $\forall j,k$, are uncorrelated)

<u>REMARK</u>. It is also possible to assume that the covariance of $G(k)N_k$, i.e. $G(k)G^T(k)$, is nonsingular. ◀

Let $(\Omega, \mathcal{F}, P)$ be the probability space for the problem we are considering, $\mathcal{F}_k^{Y_S}$ be the σ-algebra generated by the aggregate of observations up to time k:

$$Y_{S,k} \triangleq \begin{pmatrix} Y_S(0) \\ \vdots \\ Y_S(k) \end{pmatrix}$$

$$\mathcal{F}_k^{Y_S} \triangleq \sigma(Y_{S,k}) \subset \mathcal{F}$$

Moreover, clearly

$$\mathcal{F}_k^{Y_S} \subseteq \mathcal{F}_{k+1}^{Y_S}$$

i.e. $\{\mathcal{F}_k^{Y_S}\}$ is a <u>filtration</u>. Our estimation problem of $X_S(k)$ relies on a suitable manipulation of $Y_{S,k}$

$$\hat{X}_S(k) = f(Y_{S,k})$$

and if we choose $f(\cdot)$, for example, in order to minimize

$$J(f(Y_{S,k})) = E\{\|f(Y_{S,k}) - X_S(k)\|^2\}$$

(error variance) then, as we know,

$$\boxed{\hat{X}_S(k) = E\{X_S(k) \mid \mathcal{Y}_k^{Y_S}\}}$$

This estimate, as we know, also minimises the error covariance

$$E\{(X_S(k) - \hat{X}_S(k))(X_S(k) - \hat{X}_S(k))^T\}$$

REMARK. It is also possible, using $\hat{X}(k)$, to give an optimal estimate of $s(k) \triangleq C(k) X_S(k)$, the "useful" part of $Y_S(k)$:

$$\hat{s}(k) = E\{s(k) \mid \mathcal{Y}_k^{Y_S}\} = C(k) \hat{X}_S(k)$$

The solution of our problem relies in the computation of $E\{X_S(k) \mid \mathcal{Y}_k^{Y_S}\}$ for which we need $p_{X_S(k) \mid Y_{S,k}}(x_k, y_k)$, the conditional density of $X_S(k)$ given

$Y_{S,k}$. This task is very complex from a computational points of view since $Y_{S,k}$ is a vector with large dimension as k becomes large. Recursive solutions of this problem are much more efficient. This kind of implementation is possible when $\{N_k\}$ is gaussian and white and it is what we call KALMAN FILTER.

Before going to the technical details of this implementation we sum up the main assumptions on the system (CSSF)

1) $\{N_k\}$ is gaussian, with $E\{N_k\}=0$ and $E\{N_k N_j^T\}=\delta_{kj} I$, $k \neq j$.
2) $X_S(0)$ is gaussian and independent from $\{N_k\}$ (and, thus, uncorrelated).

We notice at once that it is possible to obtain a <u>recursive</u> expression for the covariance $\Psi_{X_S(k)}$ of $X_S(k)$ from the initial data $\Psi_{X_S(0)}$:

$$\Psi_{X_S(k+1)} = E\{X_S(k+1)X_S^T(k+1)\}$$

$$= E\{(A(k)X_S(k)+F(k)N_k)\cdot (A(k)X_S(k)+F(k)N_k)^T\}$$

$$= A(k)E\{X_S(k)X_S^T(k)\}A^T(k)$$

$$+ A(k)E\{X_S(k)N_k^T\}F^T(k)$$

$$+ F(k)E\{N_k X_S^T(k)\}A^T(k)$$

$$+ F(k)E\{N_k N_k^T\}F^T(k)$$

but since $E\{X_S(k)N_k^T\} = 0 \Rightarrow$

$$\boxed{\Psi_{X_S(k+1)} = A(k)\Psi_{X_S(k)}A^T(k) + F(k)F^T(k)}$$

This recursive equation will be helpful in the sequel.

6.1 DEFINITION OF INNOVATION SEQUENCES

An important preliminary passage is to redefine some sequences of random vectors which will be used to implement the optimal recursive filter.

DEFINITION (STATE INNOVATIONS)

We define the SEQUENCE of STATE INNOVATIONS:

$$\boxed{\begin{aligned} \nu_S(0) &= \hat{X}_S(0) = 0 \\ \nu_S(k) &= \hat{X}_S(k) - E\{\hat{X}_S(k) \mid \mathcal{Y}_{k-1}^{Y_S}\} \end{aligned}}$$

DEFINITION (OUTPUT INNOVATIONS)

$$\boxed{\begin{aligned} \mu_S(0) &= \hat{Y}_S(0) = 0 \\ \mu_S(k) &= \hat{Y}_S(k) - E\{\hat{Y}_S(k) \mid \mathcal{Y}_{k-1}^{Y_S}\} \end{aligned}}$$

(recall that by definition:

$\hat{X}_S(k) = E\{X_S(k) \mid \mathcal{Y}_k^{Y_S}\}$, $\hat{Y}_S(k) = E\{Y_S(k) \mid \mathcal{Y}_k^{Y_S}\}$)

Notice that

$$E\{\hat{X}_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\} = E\{E\{X_s(k) \mid \mathcal{Y}_k^{Y_s}\} \mid \mathcal{Y}_{k-1}^{Y_s}\}$$

$$= E\{X_s(k) \mid \mathcal{F}_{k-1}^{Y}\} \triangleq \hat{X}_s(k|k-1)$$

which is the optimal estimate of $X_s(k)$ given the observations $Y_s(k), \dots, Y_s(k-1)$ (<u>one-step prediction</u>).

Therefore, $\nu_s(k)$ is a $\mathcal{F}_k^{Y_s}$- measurable random vector and represents the "innovative" contribution of $Y_s(k)$ in the estimation of $X_s(k)$. Similar remark can be made on $\mu_s(k)$.

Also note that, since $Y_s(k)$ is $\mathcal{Y}_k^{Y_s}$- measurable

$$\hat{Y}_s(k) = Y_s(k)$$

In conclusion, we have the following definitions:

$$\boxed{\begin{aligned} \nu_s(k) &\triangleq \hat{X}_s(k) - E\{X_s(k) \mid \mathcal{F}_{k-1}^{Y_s}\} \\ \mu_s(k) &\triangleq Y_s(k) - E\{Y_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\} \end{aligned}}$$

An important property of the state innovations sequences is the following.

FACT #1.

$$E\{\nu_s(k)\}=0$$

$$E\{\nu_s(k)\nu_s^T(j)\}=0,\ k\neq j$$

Proof. We have

$$E\{\nu_s(k)\}=E\{\overbrace{E\{X_s(k)|\mathcal{Y}_k^{Y_s}\}}^{\hat{X}_s(k)}\}$$

$$-E\{E\{X_s(k)|\mathcal{Y}_{k-1}^{Y}\}\}$$

$$=E\{X_s(k)\}-E\{X_s(k)\}=0.$$

Moreover,

$$E\{\nu_s(k)\nu_s^T(j)\}=$$

$$E\{[\hat{X}_s(k)-E\{X_s(k)|\mathcal{Y}_{k-1}^{Y_s}\}][\hat{X}_s(j)-E\{X_s(j)|\mathcal{Y}_{j-1}^{Y_s}\}]^T\}$$

Notice that if $k\neq j$, we have either $k>j$ or $j>k$. We prove only the case $\boxed{k<j}$.

It follows $\mathcal{Y}_k^{Y_s}\subseteq\mathcal{Y}_j^{Y_s}$ and for $k<j$:

$$E\{\nu_s(k)\nu_s^T(j)\} = E\{E\{\nu_s(k)\nu_s^T(j)\mid \mathcal{F}_k^{Y_s}\}\}.$$

Moreover, $\nu_s(k)$ is $\mathcal{F}_k^{Y_s}$-measurable since $\hat{X}_s(k)$ is $\mathcal{F}_k^{Y_s}$-measurable and $E\{X_s(k)\mid \mathcal{F}_{k-1}^{Y_s}\}$ is $\mathcal{F}_{k-1}^{Y_s}(\subseteq \mathcal{F}_k^{Y_s})$-measurable. It follows

$$E\{\nu_s(k)\nu_s^T(j)\} = E\{E\{\nu_s(k)\nu_s^T(j)\mid \mathcal{F}_k^{Y_s}\}\}$$
$$= E\{\nu_s(k)\,E\{\nu_s^T(j)\mid \mathcal{F}_k^{Y_s}\}\}.$$

But, since $\mathcal{F}_k^{Y_s} \subseteq \mathcal{F}_{j-1}^{s}$:

$$E\{\nu_s(j)\mid \mathcal{F}_k^{Y_s}\} =$$
$$= E\{\hat{X}_s(j) - E\{X_s(j)\mid \mathcal{F}_{j-1}^{Y_s}\}\mid \mathcal{F}_k^{Y_s}\}$$
$$= E\{E\{X_s(j)\mid \mathcal{F}_j^{Y_s}\} - E\{X_s(j)\mid \mathcal{F}_{j-1}^{Y_s}\}\mid \mathcal{F}_k^{Y_s}\}$$
$$= E\{X_s(j)\mid \mathcal{F}_k^{Y_s}\} - E\{X_s(j)\mid \mathcal{F}_k^{Y_s}\} = 0$$

Therefore, for $k<j$

$$E\{\nu_s(k)\nu_s^T(j)\} = 0$$

FACT. #2.

$$E\{\mu_s(k)\} = 0$$

$$E\{\mu_s(k)\mu_s^T(j)\}, \quad \forall k \neq j.$$

FACT. #3.

$$E\{\nu_s(k)\mu_s^T(j)\} = 0 \quad \forall k \neq j.$$

Proof. Notice that $\nu_s(k)$ is $\mathcal{Y}_k^{Y_s}$-measurable and $\mu_s(j)$ is $\mathcal{Y}_j^{Y_s}$-measurable. For $k > j$,

$$E\{\nu_s(k)\mu_s^T(j)\} = E\{E\{\nu_s(k)\mu_s^T(j) \mid \mathcal{Y}_j^{Y_s}\}\}$$

$$= E\{E\{\nu_s(k) \mid \mathcal{Y}_j^{Y_s}\}\mu_s^T(j)\}$$

But

$$E\{\nu_s(k) \mid \mathcal{Y}_j^{Y}\} = E\{\hat{X}_s(k) - E\{X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\} \mid \mathcal{Y}_j^{Y_s}\}$$

$$= E\{E\{X_s(k) \mid \mathcal{Y}_k^{Y_s}\} - E\{X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\} \mid \mathcal{Y}_j^{Y_s}\}$$

and since $\mathcal{Y}_j^{Y_s} \subseteq \mathcal{Y}_{k-1}^{Y_s} \subseteq \mathcal{Y}_k^{Y_s}$:

$$= E\{X_s(k) \mid \mathcal{F}_j^{Y_s}\} - E\{X_s(k) \mid \mathcal{F}_j^{Y_s}\} = 0$$

which implies for $k > j$:

$$E\{\nu_s(k)\,\mu_s^T(j)\} = 0$$

FACT. #4.

$$\nu_s(k) = \hat{X}_s(k) - A(k-1)\hat{X}_s(k-1)$$

Proof.

$$\nu_s(k) = \hat{X}_s(k) - E\{X_s(k) \mid \mathcal{F}_{k-1}^{Y_s}\}$$

$$= \hat{X}_s(k) - E\{A(k-1)X_s(k-1) + F(k-1)N_{k-1} \mid \mathcal{F}_{k-1}^{Y_s}\}$$

$$= \hat{X}_s(k) - A(k-1)\hat{X}_s(k-1)$$
$$\quad - E\{F(k-1)N_{k-1} \mid \mathcal{F}_{k-1}^{Y_s}\}$$

But

$$Y_s(k-1) = C(k-1)X_s(k-1) + G(k-1)N_{k-1}$$

$$= C(k-1)A(k-2)X_s(k-2) +$$
$$+ C(k-1)F(k-2)N_{k-2} + G(k-1)N_{k-1}$$

Moreover, $F(k-1)N_{k-1}$, $F(j-2)N_{j-2}$ and $G(j-1)N_{j-1}$ are all independent for $j \leq k$. Therefore,

$$E\{F(k-1)N_{k-1} \mid \mathcal{F}^{Y_s}_{k-1}\} =$$
$$E\{F(k-1)N_{k-1}\} = 0$$

which implies our thesis

REMARK. From fact #4 and the definition of $\nu_s(k)$ it follows that

$$\boxed{\hat{X}_s(k|k-1) = A(k-1)\hat{X}_s(k-1)},$$

the optimal one-step prediction:
Compare this estimate with the state

$$X_s(k) = A(k-1)X_s(k-1) + F(k-1)N_k$$

Fact #5.

$$\mu_s(k) = Y_s(k) - C(k)A(k-1)\hat{X}_s(k-1)$$

Proof. $\mu_s(k) = Y_s(k) - E\{Y_s(k) \mid \mathcal{F}^{Y_s}_{k-1}\}$

$$= Y_s(k) - E\{C(k)X_s(k) + G(k)N_k \mid \mathcal{F}^{Y_s}_{k-1}\}$$
$$= Y_s(k) - C(k)\hat{X}_s(k|k-1) - E\{G(k)N_k \mid \mathcal{F}^{Y_s}_{k-1}\}$$
$$= Y_s(k) - C(k)A(k-1)\hat{X}_s(k-1)$$

<u>REMARK</u>. From fact #5 and the definition of $\mu_s(k)$:

$$E\{Y_s(k) \mid \mathcal{F}_{k-1}^{Y_s}\} = C(k)A(k-1)\hat{X}_s(k-1),$$

the one-step optimal prediction of the output $Y_s(k)$. Compare with

$$Y_s(k) = C(k)A(k-1)X_s(k-1) + C(k)F(k-1)N_{k-1} + G(k)N_k$$

Fact #6. The innovation sequences $\{\nu_s(k)\}$, $\{\mu_s(k)\}$ are gaussian random vectors.

First of all, notice that $\forall\, k \geq 1$:

$$X_s(k) = \Phi(k,0)X_s(0) + \sum_{j=0}^{k-1} \Phi(k,j+1)F(j)N_j$$

where

$$\Phi(k,k) = I$$

$$\Phi(k,j) = A(k-1)\cdots A(j), \quad k>j.$$

Therefore, $X_s(k)$ is a linear combination of gaussian vectors with zero mean.

It follows that $X_s(k)$ is gaussian with zero mean. Also

$$Y_s(j) = C(j)X_s(j) + G(j)N_j$$
$$j=0,\dots,k,$$

so that $Y_s(j)$ is gaussian with zero mean.

But also $\hat{X}_s(k)$ and $\hat{X}_s(k|k-1)$ are gaussian. Indeed, since

$$\hat{X}_s(k) = E\{X_s(k) \mid \mathcal{Y}_k^{Y_s}\}$$

is an affine function of the (gaussian) observations $Y_s(k)$, it is gaussian.

For the same reasons,

$$\hat{X}_s(k|k-1) = E\{X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\}$$

is gaussian. And also, $E\{Y_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\}$ is gaussian. It follows, by the definition of $\{\nu_s(k)\}$ and $\{\mu_s(k)\}$, that they are gaussian random vectors. ◀

Fact # 7 $\mathcal{F}_k^{\mu_s} = \mathcal{F}_k^{Y_s}$

where

$$\mathcal{F}_k^{\mu_s} = \sigma\{\mu_s(j) : 0 \le j \le k\}$$

Since

$$\mu_0(0) = Y_s(0)$$

$$\mu_s(k) = Y_s(k) - E\{Y_s(k) \mid \mathcal{F}_{k-1}^{Y_s}\}$$

it follows that $\mu_s(j)$, $0 \le j \le k$, is $\mathcal{F}_k^{Y_s}$ - measurable, since $\mathcal{F}_j^{Y_s} \subseteq \mathcal{F}_k^{Y_s}$.
But $\mathcal{F}_k^{\mu_s}$ is the smallest σ-algebra for which $\mu_s(j)$, $j = 0, \dots, k$, are $\mathcal{F}_k^{\mu_s}$ - measurable. It follows

$$\mathcal{F}_k^{\mu_s} \subseteq \mathcal{F}_k^{Y_s}$$

We will prove now

$$\mathcal{F}_k^{Y_s} \subseteq \mathcal{F}_k^{\mu_s}$$

so that equality follows.

But $Y_s(j)$ is a measurable function of $\mu_s(i)$ for $i \leq j$. Indeed,

$$Y_s(0) = \mu_s(0)$$

$$\Rightarrow \mathcal{F}_0^{Y_s} = \mathcal{F}_0^{\mu_s}.$$

~~Also~~

$$Y_s(1) = \mu_s(1) + \overbrace{E\{Y_s(1) \mid \mathcal{F}_0^{Y_s}\}}^{\text{meas. function of } Y_s(0) = Y_s(0)}$$

$\Rightarrow Y_s(1)$ is a measurable function of $\mu_s(1), \mu_s(0)$.

Also,

$$Y_s(2) = \mu_s(2) + \overbrace{E\{Y_s(2) \mid \mathcal{F}_1^{Y_s}\}}^{\text{meas. function of } Y_s(0),\, Y_s(1)}$$

$\Rightarrow Y_s(2)$ is a measurable function of $\mu_s(0), \mu_s(1), \mu_s(2)$

It follows by induction that $Y_s(k)$ is a measurable function of $\mu_s(0), \ldots, \mu_s(k) \Rightarrow$

$$\mathcal{F}_k^{Y_s} \subseteq \mathcal{F}_k^{\mu_s}$$

6.1. KALMAN FILTER EQUATIONS

Let us consider the relation we proved before:

$$\nu_s(k) = \hat{X}_s(k) - A(k-1)\hat{X}_s(k-1)$$

$\boxed{\begin{array}{l}\underline{\text{Fact \# 8}}.\\ \hat{\nu}_s(k) \triangleq E\{\nu_s(k) \mid \mathcal{Y}_k^{Y_s}\} = \nu_s(k)\end{array}}$

Indeed,

$$\hat{\nu}_s(k) = E\{\nu_s(k) \mid \mathcal{Y}_k^{Y_s}\} =$$

$$= E\{\hat{X}_s(k) - E\{X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\} \mid \mathcal{Y}_k^{Y_s}\}$$

$$= E\{E\{X_s(k) \mid \mathcal{Y}_k^{Y_s}\} - E\{X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\} \mid \mathcal{Y}_k^{Y_s}\}$$

$$= \hat{X}_s(k) - E\{X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\} = \nu_s(k) \quad \blacktriangleleft$$

From fact # 7

$$\boxed{\nu_s(k) = E\{\nu_s(k) \mid \mathcal{Y}_k^{\mu_s}\}}$$

Since

$$\begin{bmatrix} \nu_s(k) \\ \mu_s(0) \\ \vdots \\ \mu_s(k) \end{bmatrix}$$

is gaussian with zero mean, it follows that

$$E\{\mu_s(k) \mid \mathcal{Y}_k^{\nu_s}\}$$

(the optimal estimate of $\mu_s(k)$ given $\mathcal{Y}_k^{\mu_s}$) is a linear function of

$$\begin{bmatrix} \mu_s(0) \\ \vdots \\ \mu_s(k) \end{bmatrix} :$$

$$E\{\nu_s(k) \mid \mathcal{Y}_k^{\nu_s}\} = \Pi(k) \begin{bmatrix} \mu_s(0) \\ \vdots \\ \mu_s(k) \end{bmatrix}$$

for some $\Pi(k)$. Let us give $\Pi(k)$ the following structure

$$\Pi(k) = \left[\Pi_0(k) \,\vdots\, \cdots\cdots \,\vdots\, \Pi_k(k) \right]$$

We want to show that

$$\boxed{\Pi_j(k) = 0 \quad \forall\, j < k.}$$

Indeed,

$$\hat{\nu}_s(k) = E\{\nu_s(k) \mid \mathcal{F}_k^{\mu_s}\} = \sum_{i=0}^{k} \Pi_i(k)\mu_s(i)$$

Post-multiplying by $\mu_s^T(j)$, $j<k$, and applying expectation:

$$E\{\hat{\nu}_s(k)\mu_s^T(j)\} = \sum_{i=0}^{k} \Pi_i(k) E\{\mu_s(i)\mu_s^T(j)\}$$

But we know that

$$\left.\begin{aligned} &E\{\nu_s(k)\mu_s^T(j)\} = 0 \quad k \neq j, \\ &E\{\mu_s(i)\mu_s(j)\} = 0 \quad i \neq j \end{aligned}\right\}$$

$\Rightarrow$

$$0 = E\{\nu_s(k)\mu_s^T(j)\} = \Pi_j(k)\underbrace{E\{\mu_s(j)\mu_s^T(j)\}}_{\left(\substack{>0 \text{ since} \\ \nu_s(j) \text{ is} \\ \text{gaussian}}\right)}$$

$\Rightarrow \Pi_j(k) = 0 \quad \forall\, j<k$ ◀

We conclude that

$$\boxed{\underline{\text{Fact \# 9}}.\quad \nu_s(k) = \underbrace{\Pi_k(k)}_{\text{gain matrix}}\mu_s(k)}$$

Therefore,

$$\hat{X}_s(k) - A(k-1)\hat{X}_s(k-1)$$

$$= \nu_s(k) = \Pi_k(k)\mu_s(k)$$

$$= \Pi_k(k)\left[Y_s(k) - C(k)A(k-1)\hat{X}_s(k-1)\right]$$

$$\Rightarrow$$

$$\boxed{\begin{aligned}\hat{X}_s(k) &= \overbrace{A(k-1)\hat{X}_s(k-1)}^{\hat{X}_s(k|k-1)} \\ &\quad + \underbrace{\Pi_k(k)}_{\text{gain matrix}}\Big(Y_s(k) - C(k)\underbrace{A(k-1)\hat{X}_s(k-1)}_{\hat{X}_s(k|k-1)}\Big)\end{aligned}} \qquad (*)$$

and also

$$\hat{X}_s(k) = \big(I - \Pi_k(k)C(k)\big)A(k)\hat{X}_s(k-1) + \Pi_k(k)Y_s(k) \qquad (**)$$

6.2. COMPUTATION OF $\Pi_k(t)$

We start from the equation

$$\nu_s(k) = \Pi_k(t)\,\mu_s(t)$$

Multiply on the right both members by $\mu_s^T(t)$ e apply expectations

$$\underbrace{E\{\nu_s(k)\mu_s^T(k)\}}_{\Psi_{\nu_s(k),\mu_s(k)}} = \Pi_k(t)\,\underbrace{E\{\mu_s(k)\mu_s^T(k)\}}_{\Psi_{\mu_s(t)}}$$

$$\Rightarrow \quad \boxed{\Pi_k(t) = \Psi_{\nu_s(k),\mu_s(k)}\,\Psi^{-1}_{\mu_s(t)}}$$

In order to refine this formula, we compute $E\{e_s(t)e_s^T(k)\}$, where $e_s(t) \triangleq X_s(t) - \hat{X}_s(t)$ (the estimation error).

6.2.1. Calculus of $E\{e_s(k)e_s^T(k)\}$

Since

$$E\{\hat{X}_s(k)\} = E\{E\{X_s(k) \mid \mathcal{Y}_k^{Y_s}\}\}$$

$$= E\{X_s(k)\}$$

then $E\{e_s(k)\} = 0$ and

$P(k) \triangleq E\{e_s(k)e_s^T(k)\}$ is the estimation error covariance.

Using (**) (pg. 180)

$$e_s(k) = X_s(k) - [I - \Pi_k(k)C(k)]A(k-1)\hat{X}_s(k-1)$$
$$- \Pi_k(k)Y_s(k)$$

and since from system equations

$$\begin{cases} X_s(k) = A(k-1)X_s(k-1) + F(k-1)N_{k-1} \\ Y_s(k) = C(k)A(k-1)X_s(k-1) + C(k)F(k-1)N_{k-1} \\ \quad + G(k)N_k \end{cases}$$

we obtain

$$e_s(k) = \overbrace{[I - \Pi_k(k)C(k)]A(k-1)e_s(k-1)}^{(I)}$$

$$+ \underbrace{[I - \Pi_k(k)C(k)]F(k-1)N_{k-1}}_{(II)}$$

$$- \underbrace{\Pi_k(k)G(k)N_k}_{(III)}$$

The three summands (I), (II), (III) are pairwise independent:

- (II), (III) are independent since N_{k-1} and N_k are independent $\Rightarrow E\{(II)(III)^T\} = E\{(II)\}E\{(III)^T\} = 0$ (notice $E\{(II)\}=0$ and $E\{(III)\}=0$).
- (I), (III) are independent since $e_s(k-1)$ does not depend on N_k (causality of the model) but only on $N_0, \dots, N_{k-1}, X_s(0)$ $\Rightarrow E\{(I)(III)^T\} = E\{(I)\}E\{(III)^T\} = 0$ (notice $E\{(I)\}=0$, $E\{(III)\}=0$).

- (I), (II) are independent since $e_s(k-1)$ does not depend on $F(k-1)H_{k-1}$:

indeed, $X_s(k-1)$ depend from $F(k-2)H_{k-2}, \ldots, F(0)H_0, X_s(0)$; $\hat{X}_s(k-1)$ depends from $G(k-1)N_{k-1}, \ldots, G(0)N_0$; $F(k-2)N_{k-2}, \ldots, F(0)N_0$ and $X_s(0)$. All the above variable are independent with $F(k-1)N_{k-1} \Rightarrow e_s(k-1) = X_s(k-1) - \hat{X}_s(k-1)$ is independent with $F(k-1)N_{k-1}$.

It follows that

$$\boxed{P(k)} = E\{e_s(k)e_s^T(k)\} =$$

$$= E\{(I)(I)^T\} + E\{(II)(II)^T\}$$

$$+ E\{(III)(III)^T\} =$$

(recall that $E\{N_k N_k^T\} = I$ and $E\{e_s(k-1)e_s^T(k-1)\} = P(k-1)$)

$$= [(I - \Pi_k(k) C(k)] A(k-1) P(k-1) \cdot$$

$$\cdot A^T(k-1) [I - \Pi_k(k) C(k)]^T$$

$$+ [I - \Pi_k(k) C(k)] F(k-1) F^T(k-1) \cdot$$

$$\cdot [I - \Pi_k(k) C(k)]^T$$

$$+ \Pi_k(k) G(k) G^T(k) \Pi_k^T(k)$$

$$\boxed{\begin{aligned} = & [I - \Pi_k(k) C(k)] \Lambda_p(k) [I - \Pi_k(k) C(k)]^T \\ & + \Pi_k(k) G(k) G^T(k) \Pi_k^T(k) \end{aligned}}$$

where

$$\boxed{\begin{aligned} \Lambda_p(k) \triangleq & A(k-1) P(k-1) A^T(k-1) \\ & + F(k-1) F^T(k-1) \end{aligned}}$$

REMARK. The matrix $\Lambda_p(k)$ is the error prediction covariance at time k:

$$\boxed{\Lambda_p(k) = E\{ e_{s,p}(k) e_{s,p}(k) \}}$$

where $\boxed{e_{s,p}(k) \triangleq X_s(k) - E\{ X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s} \}}$

is the prediction error.

Indeed,

$$E\{e_{s,p}(k)\} = E\{X_s(k)\}$$

$$- E\{E\{X_s(k) \mid \mathcal{Y}^{Y_s}_{k-1}\}\}$$

$$= E\{X_s(k)\} - E\{X_s(k)\} = 0$$

Also (since $E\{F(k-1)N_{k-1} \mid \mathcal{Y}^{Y_s}_{k-1}\} = 0$):

$$e_{s,p}(k) = X_s(k) - A(k-1)\hat{X}_s(k-1)$$

$$= \underbrace{A(k-1)e_{s,p}(k)}_{(I)} + \underbrace{F(k-1)N_{k-1}}_{(II)}$$

Since (I) and (II) are independent:

$$\Psi_{e_{s,p}(k)} \triangleq E\{e_{s,p}(k)e_{s,p}(k)\} = \Lambda_p(k)$$

Next, using the formulas for $P(k)$ we compute $E\{\nu_s(k)\mu_s^T(k)\}$ and $E\{\mu_s(k)\mu_s^T(k)\}$.

6.2.2. Calculus of $E\{\mu_s(k)\mu_s^T(k)\}$ and $E\{\nu_s(k)\mu_s^T(k)\}$

Let's begin from

$$\mu_s(k) = \underbrace{C(k)X_s(k) + G(k)N_k}_{Y_s(k)} \underbrace{- C(k)A(k-1)\hat{X}_s(k-1)}_{E\{Y_s(k)\mid \mathcal{F}_{k-1}^{Y_s}\}} =$$

(using the system equations)

$$= C(k)A(k-1)e_s(k-1) + C(k)F(k-1)N_{k-1} + G(k)N_k$$

The above three summands are independent (by similar reasons as before)

$\Rightarrow$

$$\boxed{E\{\mu_s(k)\mu_s^T(k)\}} =$$

$$C(k)A(k-1)P(k-1)A^T(k-1)C^T(k)$$

$$+ C(k)F(k-1)F^T(k-1)C^T(k)$$

$$+ G(k)G^T(k) =$$

$$= \boxed{C(k)\Lambda_p(k)C^T(k) + G(k)G^T(k)}$$

On the other hand, using the system equations

$$\nu_s(k) = \hat{X}_s(k) - E\{X_s(k) \mid \mathcal{Y}_{k-1}^{Y_s}\}$$

$$= \hat{X}_s(k) - X_s(k) + X_s(k)$$

$$- A(k-1)\hat{X}_s(k-1) =$$

$$= -e_s(k) + A(k-1)e_s(k-1) + F(k-1)N_{k-1}$$

(notice that $E\{F(k-1)N_{k-1} \mid \mathcal{Y}_{k-1}^{Y_s}\} = 0$)

$$\Rightarrow E\{\nu_s(k)\mu_s^T(k)\} = \overbrace{-E\{e_s(k)\mu_s^T(k)\}}^{(I)}$$

$$\overbrace{+ E\{A(k-1)e_s(k-1)\mu_s^T(k)\}}^{II}$$

$$\underbrace{+ E\{F(k-1)N_{k-1}\mu_s^T(k)\}}_{(III)}$$

But $(I) = 0$, i.e.

$$\boxed{E\{e_s(k)\mu_s^T(k)\} = 0}$$. Indeed, since $\mu_s(k)$ is $\mathcal{F}_k^{\mu_s} (\equiv \mathcal{F}_k^{Y_s})$ - measurable, we have

$$E\{e_s(k)\mu_s^T(k)\} = E\{E\{e_s(k)\mu_s^T(k) | \mathcal{F}_k^{Y_s}\}\}$$

$$= E\{E\{e_s(k) | \mathcal{F}_k^{Y_s}\} \mu_s^T(k)\}$$

Moreover,

$$E\{e_s(k) | \mathcal{F}_k^{Y_s}\} = E\{X_s(k) | \mathcal{F}_k^{Y_s}\}$$

$$- E\{\hat{X}_s(k) | \mathcal{F}_k^{Y_s}\} = \hat{X}_s(k) - \hat{X}_s(k) = 0.$$

Moreover,

$$(II) = A(k-1) E\{e_s(k-1)e_s^T(k-1)\} A^T(k-1) C^T(k)$$

$$= A(k-1) P(k-1) A^T(k-1)$$

$$(III) = F(k-1) E\{N_{k-1} N_{k-1}^T\} F^T(k-1) C^T(k)$$

$$= F(k-1) F^T(k-1) C^T(k)$$

It follows

$$\boxed{E\{\nu_s(k)\mu_s^T(k)\}} = A(k-1)P(k-1)A^T(k-1)C^T(k)$$

$$+ F(k-1)F^T(k-1)C^T(k)$$

$$= \boxed{\Lambda_p(k)C^T(k)}.$$

Making the proper substitutions in the formula of $\Pi_k(k)$ (pg. 182):

$$\boxed{\Pi_k(k) = \Lambda_p(k)C^T(k) \cdot \Big(\underbrace{C(k)\Lambda_p(k)C^T(k) + G(k)G^T(k)}_{= E\{\mu_s(k)\mu_s^T(k)\}}\Big)^{-1}}$$

Remark. This requires $E\{\mu_s(k)\mu_s^T(k)\}$ to be invertible. In general, however it is always true that (pg. 182)

$$\Lambda_p(k)C^T(k) = \Pi_k(k)\big(C(k)\Lambda_p(k)C^T(k) + G(k)G^T(k)\big) \qquad (*)$$

If $\Lambda_p(k)$ is positive definite then $E\{\mu_s(k)\mu_s^T(k)\}$ is nonsingular if and only if the rows of $[C(k)\ G(k)]$ are independent. It is possible to give conditions on the system to guarantee that $\Lambda_p(k)$ is positive definite (see later discussions).

Before giving the complete form of the Kalman filter, we want to further develop the expression of $P(k)$ on pg. 186-187:

$$\boxed{P(k)} = (I - \Pi_k(k)C(k)]\Lambda_p(k)[I - \Pi_k(k)C(k)]^T$$
$$+ \Pi_k(k)G(k)G^T(k)\Pi_k^T(k)$$
$$= (I - \Pi_k(k)C(k))\Lambda_p(k) +$$
$$- \left([I - \Pi_k(k)C(k)]\Lambda_p(k)C^T(k) - \Pi_k(k)G(k)G^T(k)\right)\Pi_k^T(k)$$

$$= \boxed{[I - \Pi_k(k)C(k)]\Lambda_p(k)}$$

since by (*) on pg. 191:

$$[I-\Pi_k(k)C(k)]\Lambda_p(k)C^T(k) - \Pi_k(k)G(k)G^T(k) = 0$$

6.3 KALMAN FILTER EQUATIONS

Notice first that for the filter equation (*) on pg. 180 we need the initialization value at $k=0$:

$$\hat{X}_s(0|-1) = E\{X_s(0)|\mathcal{F}_{-1}^{Y_s}\} = E\{X_s(0)|\mathcal{F}_m\} =$$

(since $\mathcal{F}_0^{Y_s} = \mathcal{F}_m$: no observations at time 0)

$$= E\{X_s(0)\} = 0$$

Moreover, for the error covariance $P(\cdot)$ equation (pg. 186) we need the initialization value

$$\Lambda_P(0) = E\{e_{S,P}(0)\, e_{S,P}^T(0)\}$$

$$= E\{(X_S(0) - X_S(0|-1)) \cdot (X_S(0) - X_S(0|-1))^T\}$$

$$= E\{X_S(0) X_S^T(0)\} = \Psi_{X_S(0)} .$$

$\boxed{\text{KALMAN FILTER}}$

$\boxed{\text{Step I.}}$ Initialization:

$$\hat{X}_S(0|-1) = E\{X_S(0)\} = 0$$

$$\Lambda_P(0) = \Psi_{X_S(0)}$$

$$k = 0$$

$\boxed{\text{Step II}}$. error prediction covariance:

$$\Lambda_P(k) = A(k-1) P(k-1) A^T(k-1) + F(k-1) F^T(k-1)$$

$\boxed{\text{Step III}}$. gain matrix :

$$\Pi_k(k) = \Lambda_p(k) C^T(k) \cdot$$
$$\cdot \left(\underbrace{C(k)\Lambda_p(k)C^T(k) + G(k)G^T(k)}_{q \times q \text{ matrix}} \right)^{-1}$$

$\boxed{\text{Step IV}}$. error covariance :

$$P(k) = \left[I - \Pi_k(k) C(k) \right] \Lambda_p(k)$$

$\boxed{\text{Step V}}$. optimal prediction :

$$\hat{X}_s(k|k-1) = A(k-1)\hat{X}_s(k)$$

$\boxed{\text{Step VI}}$. optimal estimate :

$$\hat{X}_s(k) = \hat{X}_s(k|k-1)$$
$$+ \Pi_k(k)\left(Y_s(k) - C(k)\hat{X}_s(k|k-1)\right)$$

$\boxed{\text{Step VII}}$. $k \to k+1$. Go to step II.

REMARKS

#1. The Kalman filter gives at each time k the optimal estimate of $x_s(k)$ (and its optimal prediction) together with the error covariance.

#2. The gain matrix varies with time, even if $A(k)$, $B(k)$, $F(k)$, $G(k)$ are constant.

#3. The error covariance $P(k)$ can be evaluated <u>off-line</u>, while $\hat{x}_s(k)$ can be evaluated only <u>on-line</u> (when data become available)

#4. The initialization step

$$\begin{cases} \hat{x}_s(0) = 0 \\ \Lambda_+(0) = \Psi_{x_s(0)} \end{cases}$$

is crucial for optimality, i.e. optimality must be guaranteed at $k=0$.

We will see later that this exact initialization can be relaxed. By doing this, we lose optimality. Under certain hypothesis, the resulting kalman filter is ASYMPTOTICALLY optimal.

- if $G(k)=0 \Rightarrow$

$$\Pi_k(k) = \Lambda_p(k)C^T(k)\left(C(k)\Lambda_p(k)C^T(k)\right)^{-1}$$

In this case

$$C(k)\left[I-\Pi_k(k)C(k)\right]=0, \quad C(k)\Pi_k(k)=I,$$

and the filter equation is

$$\hat{X}_s(k) = \left[I-\Pi_k(k)C(k)\right]A(k-1)\hat{X}_s(k-1) + \Pi_k(k)Y_s(k)$$

with

$$C(k)\hat{X}_s(k) = Y_s(k).$$

The matrix $I-\Pi_k(k)C(k)$ is the linear operator which projects onto $\ker C(k)$ with the scalar product $\langle\cdot,\cdot\rangle_{\Lambda_p^{-1}(k)}$.

6.4. KALMAN PREDICTOR

It is possible to rewrite the Kalman algorithm in such a way to compute at each step the prediction. Since

$$\hat{X}_s(k+1|k) = A(k)\hat{X}_s(k)$$

then

$$\hat{X}_s(k) = \hat{X}_s(k|k-1) + \Pi_k(k)\left(Y_s(k) - C(k)\hat{X}_s(k|k-1)\right)$$

and it follows

$$\boxed{\begin{aligned}\hat{X}_s(k+1|k) &= A(k)\hat{X}_s(k|k-1) \\ &+ \underbrace{A(k)\Pi_k(k)}_{\text{predictor gain } \triangleq \overline{\Pi}_k(k)}\left(Y_s(k) - C(k)\hat{X}_s(k|k-1)\right)\end{aligned}}$$

Moreover, from $P(k)$ on pg. 192-193,

$$P(k) = (I - \Pi_k(k)C(k))\Lambda_p(k)$$

$$\Lambda_p(k+1) = A(k)P(k)A^T(k) + F(k)F^T(k)$$

It follows

$$\Lambda_p(k+1) = A(k)\left(I - \Pi_k(k)C(k)\right) \cdot$$
$$\cdot \Lambda_p(k)A^T(k) + F(k)F^T(k)$$

and finally

$$\boxed{\begin{aligned}\Lambda_p(k+1) &= A(k)\Lambda_p(k)A^T(k) \\ &\quad - \overline{\Pi}_k(k)C(k)\Lambda_p(k)A^T(k) + F(k)F^T(k) \\ \text{with} & \\ \overline{\Pi}_k(k) &= A(k)\Lambda_p(k)C^T(k)\left(C(k)\Lambda_p(k)C^T(k) \right. \\ &\qquad \left. + G(k)G^T(k)\right)^{-1}\end{aligned}}$$

KALMAN PREDICTOR

(step I) INITIALIZATION :

$$\hat{X}_S(0|-1) = E\{X_S(0)\} = 0$$

$$\Lambda_p(0) = \Psi_{X_S(0)}$$

$$k = 0$$

(step II) PREDICTOR GAIN :

$$\overline{\Pi}_k(k) = A(k)\Lambda_p(k)C^T(k) \cdot \left(C(k)\Lambda_p(k)C^T(k) + G(k)G^T(k)\right)^{-1}$$

(step III) OPTIMAL PREDICTION :

$$\hat{X}_S(k+1|k) = A(k)\hat{X}_S(k|k-1) + \overline{\Pi}_k(k) \cdot \left(Y_S(k) - C(k)\hat{X}_S(k|k-1)\right)$$

(step IV) COVARIANCE :

$$\Lambda_p(k+1) = A(k)\Lambda_p(k)A^T(k) - \overline{\Pi}_k(k)C(k)\Lambda_p(k)A^T(k) + F(k)F^T(k)$$

- $k \longrightarrow k+1$

goto II

6.5. FILTRO DI KALMAN CON INGRESSO DETERMINISTICO

We want to write the equations of the kalman filter for the optimal estimate of

$$X(k) = \underbrace{X_s(k) + X_d(k)}$$

where

$$\begin{cases} X_d(k+1) = A(k)X_d(k) + B(k)U(k) \\ Y_d(k) = C(k)X_d(k) + D(k)U(k) \\ X_d(0) = E\{X(0)\} \\ X_s(k+1) = A(k)X_s(k) + F(k)N_k \\ Y_s(k) = C(k)X_s(k) + G(k)N_k \left(= Y(k) - Y_d(k)\right) \\ X_s(0) = X(0) - E\{X(0)\}, \quad E\{X_s(0)X_s^T(0)\} = \Psi_{X(0)} \end{cases}$$

The Kalman filter for $\hat{X}_s(k)$ is:

$$\begin{cases} \hat{X}_s(k+1) = A(k)\hat{X}_s(k) + \Pi_{k+1}(k+1) \cdot \\ \quad \cdot \left(Y_s(k+1) - C(k+1)A(k)\hat{X}_s(k) \right) \\ \hat{X}_s(0) = 0 \end{cases}$$

We obtain with $\hat{X}(k) = \hat{X}_s(k) + X_d(k)$
(and since $E\{X_d(k) + X_s(k) \mid \mathcal{F}_k^Y\} = X_d(k) + \hat{X}_s(k)$)

$$\boxed{\hat{X}(k+1)} = A(k)\left(X_d(k) + \hat{X}_s(k)\right) + B(k)U(k)$$
$$+ \Pi_{k+1}(k+1)\Big(Y(k+1) - D(k+1)U(k+1)$$
$$- C(k+1)\left(A(k)X_d(k) + B(k)U(k)\right)$$
$$- C(k+1)A(k)\hat{X}_s(k) \Big)$$

$$\boxed{\begin{aligned} &= A(k)\hat{X}(k) + B(k)U(k) \\ &+ \Pi_{k+1}(k+1)\Big(Y(k+1) - D(k+1)U(k+1) \\ &\quad - C(k+1)\left(A(k)\hat{X}(k) + B(k)U(k)\right)\Big) \end{aligned}}$$

$$\boxed{\hat{X}(0) = E\{X(0)\}}$$

$\Rightarrow$

$$\hat{X}(k+1) = \hat{X}(k+1|k) + \Pi_{k+1}(k+1) \cdot$$

$$\cdot \left(Y(k+1) - \hat{Y}(k+1|k) \right)$$

where

$$\hat{X}(k+1|k) \triangleq E\{X(k+1) | \mathcal{F}_k^Y\}$$

$$\hat{Y}(k+1|k) \triangleq E\{Y(k+1) | \mathcal{F}_k^Y\}.$$

6.6. STEADY STATE KALMAN FILTER

The Kalman Filter is optimal if $\hat{X}_s(0)$ and $P(0)$ are selected exactly equal to 0 and, respectively, $\Psi_{X_s(0)} = \Psi_{X(0)}$. Under this regard, the Kalman filter is not "robust" with respect to initialization errors. We want to study under which conditions the covariance sequence $\{P(k)\}$ has a steady state value (insensitive with respect to $P(0)$). Moreover, it is also important to study conditions under which the filter is asymptotically stable because in this case its evolution has a steady state behaviour which insensitive with respect to $\hat{X}_s(0)$.

For the analysis of the steady state properties of the filter and covariance, we will restrict to the class of system

$$\begin{cases} X_s(k+1) = AX_s(k) + FN_k & (S) \\ Y_s(k) = CX_s(k) + GN_k \\ E\{X_s(0)\} = 0, \quad E\{X_s(0)X_s^T(0)\} = \Psi_{X(0)} \\ FG^T = 0 \end{cases}$$

The Kalman filter is:

(alternative version: notice that
$P(k)C^T = \underbrace{(I - \Pi_k(k)C)\Lambda_p(k)}C^T = \Pi_k(k)GG^T$)

(I) $\hat{X}_s(0) = 0$,

$P(0) = \Psi_{x(0)}$,

$k = 0$.

(alternative initialization:
$\hat{X}_s(0|-1) = 0$
$\Lambda_p(0) = \Psi_{x(0)}$
$k = -1$)

(II)
$$\begin{cases} \Lambda_p(k+1) = A P(K) A^T + FF^T \\ \\ P(k+1) = \left(I + \Lambda_p(k+1) C^T \underbrace{(GG^T)^{-1}} C\right)^{-1} \Lambda_p(k+1) \end{cases}$$

(here GG^T is assumed nonsingular)

(III) GAIN:

$$\Pi_{k+1}(k+1) = P(k+1)C^T(GG^T)^{-1}$$

(IV) Optimal estimate:

$$\hat{X}_s(k+1) = A(k)\hat{X}_s(k) + \Pi_{k+1}(k+1) \cdot$$
$$\cdot \left(Y_s(k+1) - CA\hat{X}_s(k)\right)$$

(V) $k \to k+1$

goto (II)

6.6.1 STEADY STATE COVARIANCE

The steady state values P_∞ and $\Lambda_{p,\infty}$ of $P(t)$ and, respectively, $\Lambda_p(t)$ must satisfy the following equations:

$$\begin{cases} P_\infty = \left(I + \Lambda_{p,\infty} C^T (GG^T)^{-1} C\right)^{-1} \Lambda_{p,\infty} \\ \Lambda_{p,\infty} = A P_\infty A^T + FF^T \end{cases}$$

By replacing $\Lambda_{p,\infty}$ with its expression given by the second equation in the first equation we get

$$P_\infty = \left[I + (AP_\infty A^T + FF^T) C^T (GG^T)^{-1} C\right]^{-1} \cdot$$
$$\cdot (AP_\infty A^T + FF^T)$$

which can be rewritten as

$$\boxed{0 = P_\infty + (AP_\infty A^T + FF^T)(C^T (GG^T)^{-1} C P_\infty - I)}$$

(RICCATI EQUATION (I))

Remark. By using the algorithm on page 194-195 we have the steady state equations:

$$
\begin{cases}
\Lambda_{p,\infty} = A P_\infty A^T + F F^T \\
P_\infty = (I - K_\infty C)\Lambda_{p,\infty} \\
K_\infty = \Lambda_{p,\infty} C^T (C \Lambda_{p,\infty} C^T + G G^T)^{-1}
\end{cases}
$$

which together give

$$
\boxed{
\begin{aligned}
0 = {} & P_\infty - A P_\infty A^T - F F^T \\
& + (A P_\infty A^T + F F^T) C^T \cdot \\
& \cdot \left(C (A P_\infty A^T + F F^T) C^T + G G^T \right)^{-1} C \cdot \\
& \cdot (A P_\infty A^T + F F^T)
\end{aligned}
}
$$

(RICCATI EQUATION (II))

This equation has a symmetry which is not apparent from the Riccati equation (I). ◀

For simplifying our analysis, we will assume $GG^T = I$.

This is always possible in (S) (page 204) if G has full row rank: simply re-define the output vector $Y_s(k)$ as

$$\bar{Y}_s(k) = \bar{C} X_s(k) + \bar{G} N_k$$

where $\bar{Y}_s(k) = (GG^T)^{-\frac{1}{2}} Y_s(k)$

and $\bar{C} = (GG^T)^{-\frac{1}{2}} C$

$$\bar{G} = (GG^T)^{-\frac{1}{2}} G$$

The Riccati equations (II) on page 205 become

$$\begin{cases} \Lambda_p(k+1) = A P(k) A^T + FF^T \\ P(k+1) = (I + \Lambda_p(k) C^T C)^{-1} \Lambda_p(k+1) \end{cases} \quad \text{(RE)}$$

We will study the (steady-state) solution of these equations.

6.6.2. INVERTIBILITY OF $I + \Lambda_p(k+1)C^TC$

In this section we will study conditions to guarantee that the matrix $I + \Lambda_p(k+1)C^TC$ be nonsingular, which is necessary to implement (RE) on page 208.

Notice that, since $P(k) \geq 0 \ \forall k \geq 0$,

$$\Lambda_p(k+1) \doteq AP(k)A^T + FF^T \geq 0 \qquad \forall k \geq 0$$

and then

$$I + \underbrace{\Lambda_p(k+1)}_{\geq 0} \underbrace{C^TC}_{\geq 0} > 0, \quad \forall k \geq -1$$

$$\Rightarrow \boxed{I + \Lambda_p(k+1)C^TC \text{ nonsingular } \forall k \geq -1.}$$

6.6.3. POSITIVE DEFINITENESS OF P(k)

The matrices $P(k)$ are, as covariances, always positive semidefinite and symmetric. We want to study conditions for which $P(k)>0$ $\forall\, k\geq 0$.

FACT #1. If (A,F) is a reachable pair (i.e. $\mathrm{rank}\,[F \,\vdots\, AF \,\vdots\, \dots \,\vdots\, A^{n-1}F]=n$), $\boxed{P(k)>0 \Rightarrow P(k+1)>0}$, $\forall\, k\geq 0$.

Proof. $P(k+1)$ is nonsingular if and only if $\Lambda_p(k+1)$ is nonsingular, since $I+\Lambda_p(k+1)C^TC>0$ $\forall\, k\geq -1$ (section 6.6.2). The nonsingularity of

$$\Lambda_p(k)=AP(k)A^T+FF^T$$

follows from $P(k) > 0$

and $\operatorname{rank}\left[F \,\vdots\, AF \,\vdots\, \cdots \,\vdots\, A^{n-1}F\right] = n$

Remark. From the proof above it follows that to guarantee $P(k) > 0$ $\Rightarrow P(k+1) > 0$ it is sufficient that

$$\operatorname{rank}\left[F \,\vdots\, A\right] = n$$

which is weaker than

$$\operatorname{rank}\left[F \,\vdots\, AF \,\vdots\, \cdots \,\vdots\, A^{n-1}F\right] = n$$

or equivalently (Hautus test)

$$\operatorname{rank}\left[\lambda I - A \,\vdots\, F\right] = n \quad \forall \lambda \in \sigma(A)$$

By iterated application of fact #1 :

Fact #2. If (A, F) is a reachable pair (or, simpler, if $\operatorname{rank}[F \vdots A] = n$) then :

$$\boxed{P(k) > 0 \Rightarrow P(k+h) > 0 \quad \forall h > 0 \quad \forall k \geq 0.}$$

This fact implies for $k = 0 \Rightarrow$

$$\boxed{P(0) > 0 \Rightarrow P(k) > 0 \quad \forall k > 0}$$

6.6.4. MONOTONICITY OF $\{P(k)\}$

In this section we prove a monotone property of $\{P(k)\}$.

Define

$$\Phi(P) \triangleq (I + \Psi(P) C^T C)^{-1} \Psi(P)$$

$$\Psi(P) \triangleq A P A^T + F F^T$$

Notice that Ψ transforms positive semidefinite matrices into positive semidefinite matrices.

The equations (RE) on page 208 are written as

$$P(k+1) = \Phi(P(k))$$

with steady-state equation

$$P_\infty = \Phi(P_\infty) .$$

FACT #3.

P, Q symmetric and positive semidefinite $\Rightarrow$

$$\Phi(P+Q) \geq \Phi(P)$$

Proof. (Sketch)

- $$\Phi(P+Q) - \Phi(P) = \int_0^1 \frac{d}{d\lambda} \Phi(P+\lambda Q)\, d\lambda$$

- if $\quad S(\lambda) = I + \Psi(P+\lambda Q) C^T C$:

$$\begin{cases} \dfrac{dS}{d\lambda} = AQA^T C^T C \\[2ex] \dfrac{dS^{-1}}{d\lambda} = -S^{-1}(\lambda)\left(\dfrac{dS}{d\lambda}(\lambda)\right) S^{-1}(\lambda) \end{cases}$$

$$= -S^{-1}(\lambda) AQA^T C^T C S^{-1}(\lambda)$$

- $$\frac{d}{d\lambda}\Phi(P+\lambda Q) = \frac{d}{d\lambda}\left(S^{-1}(\lambda)\Psi(P+\lambda Q)\right)$$

$$= S^{-1}(\lambda) AQA^T S^{-T}(\lambda) \geq 0$$

◀

A direct consequence of fact #3 is that:

$$\boxed{\text{if there exists } k_0 : \quad P(k_0+1) \geq P(k_0)}$$

$$\Rightarrow \underbrace{\Phi(P(k_0+1))}_{=P(k_0+2)} \geq \underbrace{\Phi(P(k_0))}_{=P(k_0+1)}$$

$$\Rightarrow \underbrace{\Phi(P(k_0+2))}_{=P(k_0+3)} \geq \underbrace{\Phi(P(k_0+1))}_{=P(k_0+2)}$$

$\Rightarrow$ etc. etc.

$$\Rightarrow \boxed{P(k+1) \geq P(k) \,, \quad \forall k \geq k_0}$$

Similarly,

$$\boxed{\text{if there exists } k_0 : \quad P(k_0+1) \leq P(k_0)}$$

$$\Rightarrow \boxed{P(k+1) \leq P(k) \,, \quad \forall k \geq k_0}$$

It follows that to know if $\{P(k)\}$ is monotone nondecreasing / nonincreasing it is sufficient to see if

$$P(0) \leq P(1) \; / \; P(0) \geq P(1) .$$

In particular, $P(0) = 0$ sets off a nondecreasing sequence $\{P(k)\}$, since

$$P(0) = 0 \Rightarrow \Lambda_p(0) = FF^T \geq 0 \Rightarrow$$
$$P(1) = (I + FF^T C^T C)^{-1} FF^T \geq 0 \Rightarrow$$
$$P(1) \geq P(0) = 0$$

6.6.5 BOUNDEDNESS OF $\{P(k)\}$

(I) RECALLS. Equivalent conditions:

i) A has all its eigenvalues in $\mathbb{S}^1$ (unit circle): $\sigma(A) \subset \mathring{\mathbb{S}}^1$;

ii) $\lim_{k \to +\infty} A^k = 0$;

iii) $\exists\ \mu > 0,\ \lambda \in (0,1)$:
$$\|A^k\| \leq \mu \lambda^k, \quad k \geq 0;$$

iv) $\forall\ Q$ symmetric and positive definite, $\exists!\ P$ symmetric and positive definite:
$$\boxed{P = APA^T + Q} \quad \text{(Lyapunov equation)}.$$

II) RECALLS. A state x is

$$\boxed{\text{A-STABLE if } \lim_{k \to +\infty} A^k x = 0}$$

Equivalent conditions for $\begin{cases} x(k+1) = Ax(k) \\ y(k) = Cx(k) \end{cases}$:

i) (C, A) is detectable;

ii) all the unobservable states are A-stable, i.e.

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} x = 0 \Rightarrow x \text{ is A-stable};$$

iii) $\operatorname{rank} \begin{bmatrix} A - \lambda I \\ C \end{bmatrix} = n$

$\forall \lambda \in \mathbb{C} : |\lambda| \geq 1$

iv) $\exists\ k : \sigma(A - kC) \subset S^1$

<u>FACT #4</u>. If (C, A) is detectable, for each $P(0) \geq 0$ there exists a sequence $\{S(k)\}$ of positive semidefinite matrices such that $P(k) \leq S(k) \ \forall\ k \geq 0$ and

$$\lim_{k \to \infty} S(k) = S_\infty$$

with S_∞ depending on $P(0)$

<u>Proof</u>. Construct a Luenberger observer with error covariance $\{S(k)\}$ such that $\lim_{k \to +\infty} S(k) = S_\infty$

It follows that, if $\{P(k)\}$ is monotone nondecreasing, since by fact #4 $P(k) \leq S(k) < +\infty \ \forall\ k \geq 0$, then $\lim_{k \to +\infty} P(k) = P_\infty \geq 0$. On the other hand, if $\{P(k)\}$ is monotone nonincreasing, since $\{P(k)\}$ is lower bounded by 0, then $\lim_{k \to +\infty} P(k) = P_\infty \geq 0$.

However, the steady state P_∞ may depend on $P(0)$.

Therefore, it is of interest to understand under which conditions P_∞ is unique.

<u>FACT #5</u>. Let $P_\infty \geq 0$ a symmetric solution of the Riccati equation (I) on page 206. If (F^T, A^T) is detectable, P_∞ is the unique solution of (I) and

$$\sigma\left(\left(I - P_\infty C^T C\right)A\right) \subset S^1$$

6.6.6. STEADY STATE KALMAN FILTER EQUATIONS (SSKF)

Collecting FACTS # 4 & 5 and monotonicity of $\{P(k)\}$, we can state and prove the following conclusive result.

(SSKF). If (C, A) and (F^T, A^T) are detectable:

1) UNIQUENESS OF STEADY STATE ERROR COVARIANCE:
there exists a unique symmetric solution $P_\infty \geq 0$ of the Riccati equation (I) and $\lim_{k \to +\infty} P(k) = P_\infty$;

2) STABILITY OF SSKF:

$$\sigma((I - P_\infty C^T C)A) \subset S^1 \quad ;$$

3) ASYMPTOTIC OPTIMALITY of SSKF:
the SSKF filter

$$\hat{Z}_s(k+1) = A\hat{Z}_s(k) + P_\infty C^T(Y_s(k+1) - CA\hat{Z}_s(k))$$

is asymptotically optimal, in the sense that if

$$P_E^{\hat{Z}_s}(k) \triangleq E\{(\hat{Z}_s(k) - X_s(k))(\hat{Z}_s(k) - X_s(k))^T\}$$

then
$$\lim_{k \to +\infty} P^{\hat{Z}_s}(k) = P_\infty .$$

<u>Proof.</u> 1) and 2) follow from facts #4 & #5 and monotonicity of $\{P(k)\}$.
We have only to prove
$$\lim_{k \to +\infty} P^{\hat{Z}_s}(k) = P_\infty .$$
Notice that, if $K_\infty \triangleq P_\infty C^T$,
$$X_s(k+1) - \hat{Z}_s(k+1) = (I - K_\infty C)(A(X_s(k) - \hat{Z}_s(k)) + FN_k) - K_\infty G N_{k+1}$$
$$= (I) + (II) + (III)$$
With standard arguments, it is possible to see that (I), (II) and (III) are pairwise independent. It follows, with similar calculations used to obtain the recursive equation for $P(k)$, that

$$P^{\hat{z}_s}(k+1) = (I - K_\infty C)(A P^{\hat{z}_s}(k) A^T + FF^T) \cdot$$
$$\cdot (I - K_\infty C)^T + K_\infty G G^T K_\infty^T$$

But $P_\infty \geq 0$ is solution of :

$$0 = P_\infty - (I - K_\infty C)(A P_\infty A^T + FF^T)(I - K_\infty C)^T$$
$$+ K_\infty G G^T K_\infty^T \qquad (\text{by } (1))$$

Subtracting the two above equations

$$P^{\hat{z}_s}(k+1) - P_\infty = [(I - K_\infty C)A] \cdot (P^{\hat{z}_s}(k) - P_\infty) \cdot$$
$$\cdot [(I - K_\infty C)A]^T$$

which, solved backwards, give

$$P^{\hat{z}_s}(k) - P_\infty = [(I - K_\infty C)A]^k \cdot (P^{\hat{z}_s}(0) - P_\infty) \cdot$$
$$\cdot ([(I - K_\infty C)A]^k)^T$$

But by 2) : $\sigma((I - K_\infty C)A) \subset S^1$

so that

there exist $\mu > 0, \lambda \in (0,1)$ (see iii) page 217)

$$\| P^{\hat{z}_s} - P_\infty \|$$

$$\leq \| [(I - K_\infty C)A]^k \| \cdot \| P^{\hat{z}_s}(0) - P_\infty \| \cdot$$

$$\cdot \| ([(I - K_\infty C)A]^k)^T \|$$

$$\leq \mu^2 \lambda^{2k} \| P^{\hat{z}_s}(0) - P_\infty \|$$

which implies $\lim\limits_{k \to \infty} P^{\hat{z}_s} = P_\infty$ ◀

7. EXTENDED KALMAN FILTER

In this section we consider nonlinear systems

$$(S)\quad X(k+1) = f(X(k), U(k), k) + F(k)N_k$$
$$(M)\quad Y(k) = h(X(k), U(k), k) + G(k)N_k$$
$$F(j)^T G(k) = 0, \forall j, k.$$

For computing a state estimate (not optimal) an extended version of the Kalman filter is used (EXTENDED KF). The filter is applied to the linearization of the system around the current values of the state estimate.

We denote by $\hat{X}(k)$ the state estimate using the measurements up to time k; $\hat{Y}(k)$ the output estimate (which coincides with $Y(k)$); $\hat{X}(k+1|k)$ and $\hat{Y}(k+1|k)$ the respective predictions:

$$(P)\begin{cases} \hat{X}(k+1|k) = f(\hat{X}(k), U(k), k) \\ \hat{Y}(k+1|k) = h(\hat{X}(k+1|k), U(k+1), k+1) \end{cases}$$

The state and output innovations are defined as

$$\nu_s(k+1) = \hat{X}(k+1) - \hat{X}(k+1|k)$$

$$\mu_s(k+1) = \hat{Y}(k+1) - \hat{Y}(k+1|k)$$

It is reasonable to look for a filter for which

$$\nu_s(k+1) = \Pi_{k+1}(k+1)\,\mu_s(k+1)$$

and

$$\hat{X}_s(k+1) = \hat{X}_s(k+1|k) + \Pi_{k+1}(k+1)\big(Y(k+1) - \hat{Y}(k+1|k)\big)$$

To obtain the gain $\Pi_{k+1}(k+1)$ we will linearize the state equation (S) around $\hat{X}(k)$ and the output equation (M) around $\hat{X}(k|k-1)$:

$$
\begin{aligned}
X(k+1) \approx\; & f(\hat{X}(k), U(k), k) \\
& + A(k)(X(k) - \hat{X}(k)) \\
& + F(k) N_k
\end{aligned}
$$

$$
\begin{aligned}
Y(k) \approx\; & h(\hat{X}(k|k-1), U(k), k) \\
& + C(k)(X(k) - \hat{X}(k|k-1)) \\
& + G(k) N_k
\end{aligned}
$$

where

$$A(k) \triangleq \left. \frac{\partial f}{\partial x} \right|_{\hat{X}(k)}$$

$$C(k) \triangleq \left. \frac{\partial h}{\partial x} \right|_{\hat{X}(k|k-1)}$$

and using the predictions (P)

$$X(k+1) \approx A(k)X(k) + \left[\hat{X}(k+1|k) - A(k)\hat{X}(k)\right] + F(k)N_k$$

$$Y(k) \approx C(k)X(k) + \left[\hat{Y}(k|k-1) - C(k)\hat{X}(k|k-1)\right] + G(k)N_k$$

Define the following deterministic inputs:

$$U_s(k) \triangleq \hat{X}(k+1|k) - A(k)\hat{X}(k|k)$$

$$U_o(k) \triangleq \hat{Y}(k|k-1) - C(k)\hat{X}(k|k-1)$$

we obtain

$$X(k+1) \approx A(k)X(k) + U_s(k) + F(k)N_k$$

$$Y(k) \approx C(k)X(k) + U_o(k) + G(k)N_k$$

The EKF (extended kalman filter) will be given by (see the KF with deterministic inputs)

$$\hat{X}(k+1) = A\hat{X}(k) + U_s(k) + \Pi_{k+1}(k+1) \cdot \left[Y(k+1) - U_0(k+1) - C(k+1) \cdot \left(A(k)\hat{X}(k) + U_s(k) \right) \right]$$

and $\Pi_{k+1}(k+1)$ is the same as the kalman gain in KF.

But

$$A(k)\hat{X}(k) + U_s(k) = \hat{X}(k+1|k)$$

and

$$U_0(k+1) + C(k+1)\left(A(k)\hat{X}(k) + U_s(k)\right) = \hat{Y}(k+1|k)$$

$$\Rightarrow \boxed{\hat{X}(k+1) = \hat{X}(k+1|k) + \Pi_{k+1}(k+1) \cdot \left(Y(k+1) - Y(k+1|k) \right)}$$

EXTENDED KF

(I) $\hat{X}(0|-1) = E\{X(0)\}$,

$\Lambda_p(0) = \Psi_{X(0)}$,

$k = -1$

(II) OUTPUT PREDICTION:

$$\hat{Y}(k+1|k) = h(\hat{X}(k+1|k), U(k+1), k+1)$$

(III) CALCUS OF $C(k+1)$:

$$C(k+1) = \frac{\partial h}{\partial x}(x, U(k+1), k+1)\Big|_{x=\hat{X}(k+1|k)}$$

(IV) KALMAN GAIN:

$$\Pi_{k+1}(k+1) = \Lambda_p(k+1)C^T(k+1) \cdot$$

$$\cdot \left(C(k+1)\Lambda_p(k+1)C^T(k+1) + G(k+1)G^T(k+1)\right)^{-1}$$

(V) ESTIMATION ERROR COVARIANCE:

$$P(k+1) = (I - \Pi_{k+1}(k+1)C(k+1))\Lambda_p(k+1)$$

(VI) STATE ESTIMATE:

$$\hat{X}(k+1) = \hat{X}(k+1|k) + \Pi_{k+1}(k+1) \cdot$$

$$\cdot [Y(k+1) - Y(k+1|k)]$$

(VII) $k \to k+1$

(VIII) STATE PREDICTION:

$$\hat{x}(k+1|k) = f(\hat{x}(k), U(k), k)$$

(IX) CALCULUS OF $A(k)$:

$$A(k) = \frac{\partial f}{\partial x}(x, U(k), k)\Big|_{x=\hat{x}(k)}$$

(X) PREDICTION ERROR COVARIANCE:

$$\Lambda_p(k+1) = A(k)P(k)A^T(k) + F(k)F^T(k)$$

(XI) goto (II).