The Kalman Filter

Chapter 4

Professor Michel Gevers

CESAME-UCL

Louvain-la-Neuve

Belgium

Outline

1. The State Estimation Problem
2. Innovations
3. The Kalman Predictor
4. The Kalman Filter
5. The Steady-State Kalman Predictor
6. Optimal Predictor in Input-Output Form

State Estimation: The Gauss-Markov Model

Linear state-space model

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + Gw_k \\ y_k = Cx_k + v_k \end{cases}$$

where

- $x_k \in \mathbb{R}^n, \quad u_k \in \mathbb{R}^m, \quad w_k \in \mathbb{R}^q, \quad y_k \in \mathbb{R}^p, \quad v_k \in \mathbb{R}^p$
- u_k deterministic, x_0 Gaussian
- $\{w_k\}$ and $\{v_k\}$ are white Gaussian noises, mutually independent, and independent of x_0
- $x_0 \sim N(\bar{x}_0, P_0), \quad w_k \sim N(0, Q), \quad v_k \sim N(0, R)$

Comments about Model Assumptions

- u_k is the "control input". In many applications it is zero.
- w_k is often called process noise; v_k is measurement noise.
- For simplicity of notation : A, B, C, G time-invariant. Linear time-varying model causes no problem.
- Correlation between w_k and v_k is easy to handle.
- Gaussian assumption : not required. It will make the Kalman estimator an optimal (conditional mean) estimator.

Formulation of the State Estimation Problem

Consider that at time k the following information is available:

- the model, i.e. $A, B, C, G, \bar{x}_0, P_0, Q, R$
- the information vector (= the data)

$$Z^k = (y_k^T, u_k^T, y_{k-1}^T, u_{k-1}^T, \ldots, y_0^T, u_0^T)^T$$

Construct an optimal estimator $\hat{x}_{j|k}$ for x_j given Z^k. We have the following 3 situations:

1) if $j = k$: optimal filtering
2) if $j > k$: optimal prediction
3) if $j < k$: optimal smoothing

Preliminary Lemma

Let X and Y be jointly Gaussian random vectors, with $\boldsymbol{E}\{X\} = m_X, \boldsymbol{E}\{Y\} = m_Y$, and

$$Cov \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix}$$

Then the Linear Min. Variance Unbiased estimator is (see Chapter 3):

$$\hat{X} = m_X + C_{XY} C_{YY}^{-1} (y - m_Y) = \boldsymbol{E}\{X|y\} \tag{1}$$

In addition the error $\tilde{X} \triangleq X - \hat{X}$ is independent of Y and hence of $\hat{X}$:

$$\boldsymbol{E}\{\tilde{X} Y^T\} = 0, \quad \boldsymbol{E}\{\tilde{X} \hat{X}^T\} = 0$$

The covariance of the error is

$$Cov(\tilde{X}) = C_{XX} - C_{XY} C_{YY}^{-1} C_{YX} \tag{2}$$

Proof

1. Observe first that $\boldsymbol{E}\{\hat{X}\} = m_X \quad \Rightarrow \quad \boldsymbol{E}\{\tilde{X}\} = 0$.

2. $\boldsymbol{E}\{\tilde{X}(Y - m_Y)^T\} = \boldsymbol{E}\{(X - m_X)(Y - m_Y)^T\} - C_{XY} = 0$

3. Thus $\tilde{X}$ and Y are independent, and hence $\boldsymbol{E}\{\tilde{X}Y^T\} = 0$
$\Longrightarrow \boldsymbol{E}\{X|Y\} = \boldsymbol{E}\{\hat{X} + \tilde{X}|Y\} = \boldsymbol{E}\{\hat{X}|Y\} = \hat{X}$

4.
$$\begin{aligned} Cov(X) &= \boldsymbol{E}\{(X - m_X)(X - m_X)^T\} \\ &= \boldsymbol{E}\{(X - \hat{X} + \hat{X} - m_X)(X - \hat{X} + \hat{X} - m_X)^T\} \\ &= \boldsymbol{E}\{\tilde{X}\tilde{X}^T\} + \boldsymbol{E}\{(X - \hat{X})(\hat{X} - m_X)^T\} \\ &\quad +\boldsymbol{E}\{(\hat{X} - m_X)(X - \hat{X})^T\} + \boldsymbol{E}\{(\hat{X} - m_X)(\hat{X} - m_X)^T\} \\ &= Cov(\tilde{X}) + Cov(\hat{X}) \end{aligned}$$

5. $\Longrightarrow Cov(\tilde{X}) = Cov(X) - Cov(\hat{X}) = C_{XX} - C_{XY}C_{YY}^{-1}C_{YX}$

Innovations

Denote: $\hat{x}_{j|k} \triangleq \boldsymbol{E}\{x_j | Z^k\}$ and $\hat{y}_{j|k} \triangleq \boldsymbol{E}\{y_j | Z^k\}$

We focus on **one-step-ahead predictors** first.

By noise assumptions:

$$\hat{y}_{k|k-1} = C\hat{x}_{k|k-1}$$

since v_k is white noise. Denote: $\tilde{y}_{k|k-1} \triangleq y_k - \hat{y}_{k|k-1}$

By the preliminary lemma:

$$\boldsymbol{E}\{\tilde{y}_{k|k-1} | Z^{k-1}\} = 0$$

Therefore $\{\tilde{y}_{k|k-1}\}$ is a sequence of independent white Gaussian random variables that contains the same information as the sequence $\{y_k\}$. The sequence $\{\tilde{y}_{k|k-1}\}$ is called the **innovations sequence** of $\{y_k\}$.

Computation of $\boldsymbol{E}\{x_{k+1}|Z^k\}$

Denote: $$Z^k = \begin{pmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_0 \end{pmatrix} \quad \text{and} \quad \tilde{Z}^k = \begin{pmatrix} \tilde{y}_{k|k-1} \\ \tilde{y}_{k-1|k-2} \\ \vdots \\ \tilde{y}_{0|-1} \end{pmatrix}$$

By the Lemma, one can show that:

$$\begin{aligned} \hat{x}_{k+1|k} &\triangleq \boldsymbol{E}\{x_{k+1}|Z^k\} = \boldsymbol{E}\{x_{k+1}|\tilde{Z}^k\} \\ &= \boldsymbol{E}\{x_{k+1}\} + C_{x_{k+1}\,\tilde{Z}^k} C^{-1}_{\tilde{Z}^k\,\tilde{Z}^k} \tilde{Z}^k \\ &= \boldsymbol{E}\{x_{k+1}\} + \sum_{j=0}^{k} \boldsymbol{E}\{x_{k+1}\,\tilde{y}^T_{j|j-1}\}\, C_j^{-1}\, \tilde{y}_{j|j-1} \end{aligned}$$

Denote: $K_k \triangleq \boldsymbol{E}\{x_{k+1}\,\tilde{y}^T_{k|k-1}\}\, C_k^{-1}$ with $C_k \triangleq \boldsymbol{E}\{\tilde{y}_{k|k-1}\tilde{y}^T_{k|k-1}\}$

Observe that:

1) $\boldsymbol{E}\{x_{k+1}\} = A E\{x_k\} + B u_k$

2) $\boldsymbol{E}\{\tilde{y}_{j|j-1} w_k^T\} = 0$ and $\boldsymbol{E}\{\tilde{y}_{j|j-1} u_k^T\} = 0$ for $j \le k$

Then:

$$
\begin{aligned}
\hat{x}_{k+1|k} &= A\boldsymbol{E}\{x_k\} + Bu_k + K_k \tilde{y}_{k|k-1} \\
&+ \sum_{j=0}^{k-1} \boldsymbol{E}\{(Ax_k + Bu_k + Gw_k)\tilde{y}_{j|j-1}^T\} C_j^{-1} \tilde{y}_{j|j-1} \\
&= A[\boldsymbol{E}\{x_k\} + \sum_{j=0}^{k-1} \boldsymbol{E}\{x_k \tilde{y}_{j|j-1}^T\} C_j^{-1} \tilde{y}_{j|j-1}] + Bu_k + K_k \tilde{y}_{k|k-1} \\
&= A\hat{x}_{k|k-1} + Bu_k + K_k \tilde{y}_{k|k-1}
\end{aligned}
$$

Computation of $\boldsymbol{E}\{x_{k+1}\tilde{y}_{k|k-1}^T\}$

Remember: $K_k \triangleq \boldsymbol{E}\{x_{k+1}\tilde{y}_{k|k-1}^T\}C_k^{-1}$. First compute $\boldsymbol{E}\{x_{k+1}\tilde{y}_{k|k-1}^T\}$.

$$\begin{aligned}
\boldsymbol{E}\{x_{k+1}\tilde{y}_{k|k-1}^T\} &= \boldsymbol{E}\{(Ax_k + Bu_k + Gw_k)\tilde{y}_{k|k-1}^T\} \\
&= AE\{x_k\tilde{y}_{k|k-1}^T\} \\
&= AE\{(x_k - \hat{x}_{k|k-1})\tilde{y}_{k|k-1}^T\} \\
&= AE\{\tilde{x}_{k|k-1}(C\tilde{x}_{k|k-1} + v_k)^T\} \\
&\triangleq AP_{k|k-1}C^T
\end{aligned}$$

$$\left.\begin{aligned}
y_k &= Cx_k + v_k \\
\hat{y}_{k|k-1} &= C\hat{x}_{k|k-1}
\end{aligned}\right\} \Rightarrow \tilde{y}_{k|k-1} = C\tilde{x}_{k|k-1} + v_k$$

$$P_{k|k-1} \triangleq \boldsymbol{E}\{\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^T\}$$

Computation of C_k and K_k

$$
\begin{aligned}
C_k &\triangleq \boldsymbol{E}\{\tilde{y}_{k|k-1}\tilde{y}_{k|k-1}^T\} \\
&= \boldsymbol{E}\{(C\tilde{x}_{k|k-1}+v_k)(C\tilde{x}_{k|k-1}+v_k)^T\} \\
&= CP_{k|k-1}C^T + R \\
K_k &\triangleq \boldsymbol{E}\{x_{k+1}\tilde{y}_{k|k-1}^T\}C_k^{-1} \\
&= AP_{k|k-1}C^T(CP_{k|k-1}C^T+R)^{-1}
\end{aligned}
$$

Remaining problem: $P_{k|k-1}$?

Recursive equation for $P_{k|k-1}$

- $x_{k+1} = Ax_k + Bu_k + Gw_k$
- $\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + K_k\tilde{y}_{k|k-1}$

Therefore:

$\tilde{x}_{k+1|k} = A\tilde{x}_{k|k-1} + Gw_k - K_k\tilde{y}_{k|k-1}$

Also:

$\tilde{y}_{k|k-1} = C\tilde{x}_{k|k-1} + v_k$

and w_k, v_k uncorrelated with $\tilde{x}_{k|k-1}$ and $\tilde{y}_{k|k-1}$

Therefore:

$$
\begin{aligned}
P_{k+1|k} &\triangleq \boldsymbol{E}\{\tilde{x}_{k+1|k}\tilde{x}_{k+1|k}^T\} \\
&= AP_{k|k-1}A^T + GQG^T + K_kC_kK_k^T \\
&\quad -A\boldsymbol{E}\{\tilde{x}_{k|k-1}\tilde{y}_{k|k-1}^T\}K_k^T - K_k\boldsymbol{E}\{\tilde{y}_{k|k-1}\tilde{x}_{k|k-1}^T\}A^T \\
&= AP_{k|k-1}A^T + GQG^T + K_kC_kK_k^T \\
&\quad -AP_{k|k-1}C^TK_k^T - K_kCP_{k|k-1}A^T \\
&= AP_{k|k-1}A^T + GQG^T \\
&\quad -AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T
\end{aligned}
$$

The Kalman Predictor Equations

Collecting all equations:

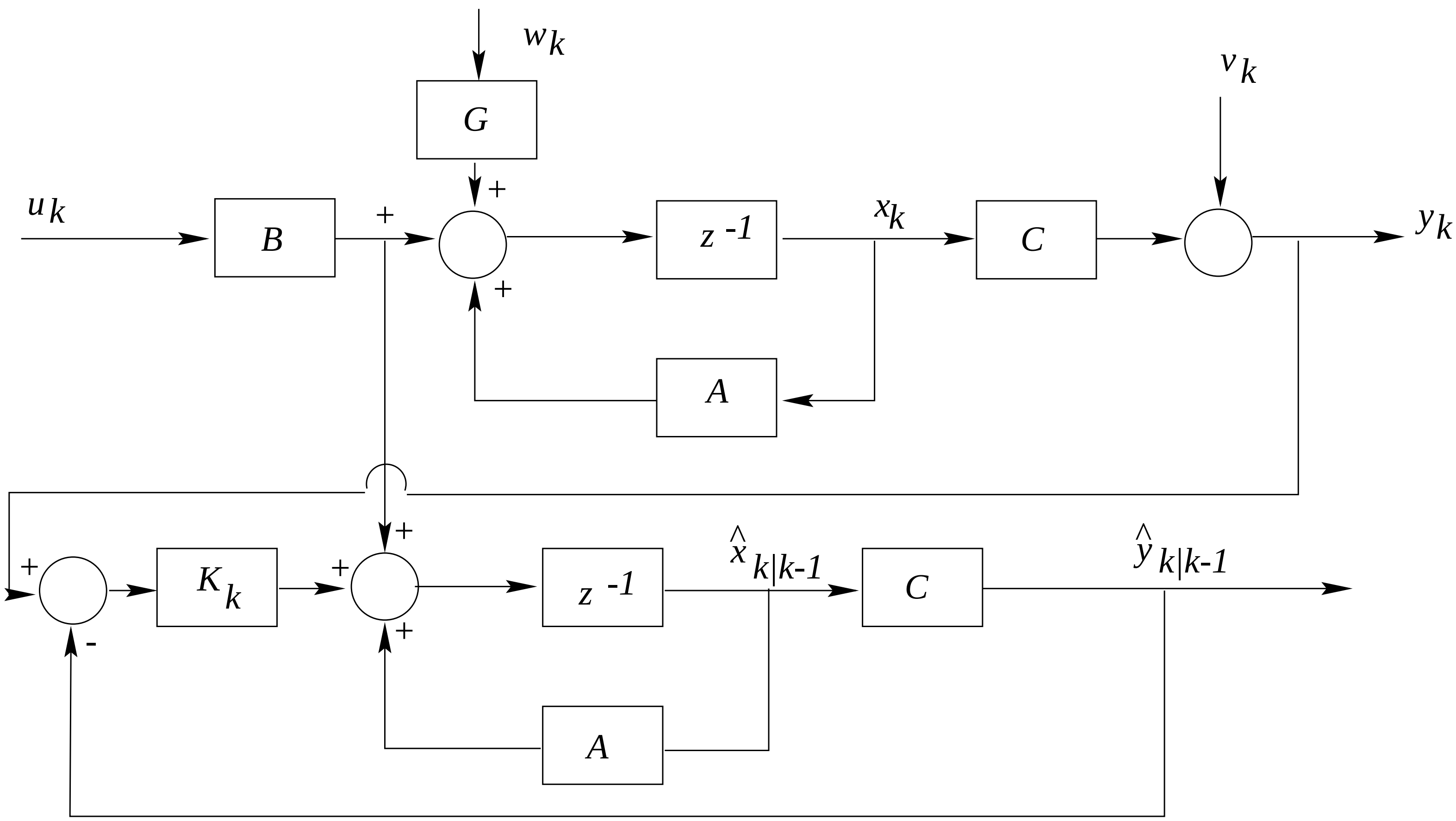
$$\begin{aligned}
\hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + Bu_k + K_k(y_k - C\hat{x}_{k|k-1}) \\
&= (A - K_kC)\hat{x}_{k|k-1} + Bu_k + K_ky_k \\
\hat{x}_{0|-1} &= \bar{x}_0
\end{aligned}$$

Kalman gain: $K_k = AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}$

Riccati equation:

$$\begin{aligned}
P_{k+1|k} &= AP_{k|k-1}A^T + GQG^T \\
&\quad -AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T \\
P_{0|-1} &= P_0
\end{aligned}$$

Block diagram representation of Kalman predictor



Some Properties

- The Kalman predictor is a 2-input 1-output filter
- $\hat{x}_{k+1|k} = \boldsymbol{E}\{x_{k+1}|Z^k\}$ = conditional mean estimate
- $P_{k+1|k} = \boldsymbol{E}\{(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^T|Z^k\}$
  $=$ conditional covariance of $\hat{x}_{k+1|k}$
- If no Gaussian assumption, then $\hat{x}_{k+1|k}$ is the Linear Minimum Variance Unbiased estimate.
- The Riccati equation is independent of the measured data. Depends on A, B, C, G, Q, R, P_0. Can be computed a priori.

Application[a]

Consider noisy measurements of an exponentially decaying signal:

$$\begin{cases} x_{k+1} = a x_k & |a| < 1 \\ y_k = x_k + v_k & \boldsymbol{E}\{v_k\} = 0, \quad \boldsymbol{E}\{v_k^2\} = R \end{cases}$$

Kalman predictor for $\hat{x}_{k+1|k}$:

$$\begin{aligned} \hat{x}_{k+1|k} &= a\hat{x}_{k|k-1} + K_k(y_k - \hat{x}_{k|k-1}), \hat{x}_{0|-1} = 0 \\ K_k &= \frac{aP_{k|k-1}}{P_{k|k-1} + R} \\ P_{k+1|k} &= a^2 P_{k|k-1} - \frac{a^2 P_{k|k-1}^2}{P_{k|k-1} + R}, \quad P_{0|-1} = P_0 \end{aligned}$$

[a] From T. Söderström: "Discrete-time stochastic systems: estimation and control", Prentice Hall International, 1994

Comparison with Constant Gain Filter

An alternative is to use a constant gain filter:

$$\hat{x}_{k+1|k} = a\hat{x}_{k|k-1} + K[y_k - \hat{x}_{k|k-1}]$$

Estimation error $\tilde{x}_k \triangleq x_k - \hat{x}_{k|k-1}$ for this filter:

$$\tilde{x}_{k+1} = (a-K)\tilde{x}_k - Kv_k, \;\; \tilde{x}_0 = x_0$$

Asymptotic variance:

$$\lim_{k\to\infty} \boldsymbol{E}\{\tilde{x}_k^2\} = \frac{K^2R}{1-(a-K)^2}$$

For rapid convergence, we want $K \simeq a$; for small steady state error we want $K \simeq 0$. With Kalman filter: $\lim_{k\to\infty} P_{k|k-1} = 0$.

Illustration with $a = 0.9,\quad P_0 = 1,\quad R = 0.04,\quad x(0) = 1,\quad \hat{x}_{0|-1} = 0$

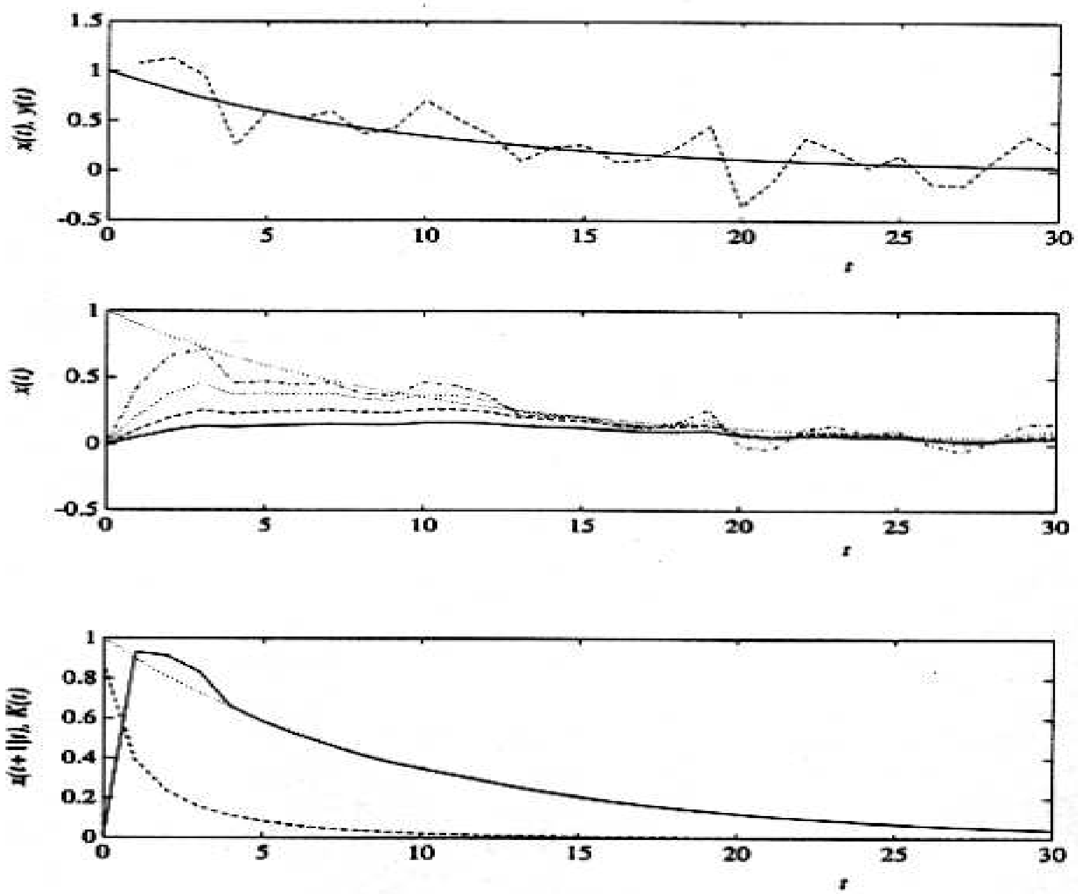


Figure 6.1 Illustration of optimal and time-invariant estimators, Example 6.2.2. (a) exponential decay $x(t)$ — solid line, measurements $y(t)$ — dotted line; (b) estimates based on time-invariant gains $K = 0.05$ (solid), 0.1 (dashed), 0.2 (dotted), 0.4 (dash–dotted); (c) optimal estimate $\hat{x}(t+1|t)$ (solid line) and optimal time-varying gain $K(t)$ (dashed line).

Computation of Kalman Filter

Applying the Lemma again, we have:

$$
\begin{aligned}
\hat{x}_{k|k} &= \boldsymbol{E}\{x_k\} + C_{x_k \tilde{Z}_k} \cdot C^{-1}_{\tilde{Z}_K \tilde{Z}_k} \cdot \tilde{Z}^k \\
&= \boldsymbol{E}\{x_k\} + \sum_{j=0}^{k} \boldsymbol{E}\{x_k \tilde{y}^T_{j|j-1}\} \cdot C_j^{-1} \tilde{y}_{j|j-1} \\
&= \boldsymbol{E}\{x_k\} + \sum_{j=0}^{k-1} \boldsymbol{E}\{x_k \tilde{y}^T_{j|j-1}\} . C_j^{-1} \tilde{y}_{j|j-1} + \underbrace{\boldsymbol{E}\{x_k \tilde{y}^T_{k|k-1}\} C_k^{-1}}_{K_k^f} \tilde{y}_{k|k-1} \\
&= \hat{x}_{k|k-1} + K_k^f \tilde{y}_{k|k-1}
\end{aligned}
$$

Computation of K_k^f

$$\begin{aligned} K_k^f &\triangleq \boldsymbol{E}\{x_k \tilde{y}_{k|k-1}^T\} C_k^{-1} \\ &= \boldsymbol{E}\{(x_k - \hat{x}_{k|k-1}) \tilde{y}_{k|k-1}^T\} C_k^{-1} \\ &= P_{k|k-1} C^T C_k^{-1} \\ \Longrightarrow K_k^f &= P_{k|k-1} C^T (C P_{k|k-1} C^T + R)^{-1} \end{aligned}$$

Observe:

- $K_k = A K_k^f$
- $$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k} + Bu_k \\ &= A\hat{x}_{k|k-1} + Bu_k + K_k \tilde{y}_{k|k-1} \end{aligned}$$

Computation of the Filter Error Covariance

Filtering error:

$$x_k - \hat{x}_{k|k} = x_k - \hat{x}_{k|k-1} - K_k^f \tilde{y}_{k|k-1}$$

Therefore:

$$\begin{aligned}
P_{k|k} &\triangleq \boldsymbol{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k}))^T\} \\
&= P_{k|k-1} - \boldsymbol{E}\{\tilde{x}_{k|k-1}\tilde{y}_{k|k-1}^T\}(K_k^f)^T \\
&\quad -K_k^f \boldsymbol{E}\{\tilde{y}_{k|k-1}(x_k - \hat{x}_{k|k-1})^T\} + K_k^f C_k (K_k^f)^T \\
&= P_{k|k-1} - K_f^k C_k (K_k^f)^T \\
&= P_{k|k-1} - P_{k|k-1} C^T (C P_{k|k-1} C^T + R)^{-1} C P_{k|k-1}
\end{aligned}$$

Observe that $P_{k|k} \leq P_{k|k-1}$. The new information improves the quality of the estimate.

The Kalman Filter Equations

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k^f \tilde{y}_{k|k-1}$$

where:

- $\hat{x}_{k|k-1}$ is computed from Kalman predictor
- $K_k^f = P_{k|k-1} C^T (C P_{k|k-1} C^T + R)^{-1}$
- $P_{k|k} \triangleq \mathrm{cov}(\hat{x}_{k|k})$
  $= P_{k|k-1} - P_{k|k-1} C^T (C P_{k|k-1} C^T + R)^{-1} C P_{k|k-1}$
- $P_{k|k-1}$ is computed from the Riccati equation

Optimal j-step ahead predictor

Objective: construct the optimal estimator $\boldsymbol{E}\{x_{k+j}|Z^k\}$ for $j > 0$.

From:

$$x_{k+j} = Ax_{k+j-1} + Bu_{k+j-1} + Gw_{k+j-1}$$

and the noise assumption on $\{w_k\}$, we get:

$$\begin{aligned} \hat{x}_{k+j|k} &\triangleq \boldsymbol{E}\{x_{k+j}|Z^k\} \\ &= A\boldsymbol{E}\{x_{k+j-1}|Z^k\} + Bu_{k+j-1} \\ &= A\hat{x}_{k+j-1|k} + Bu_{k+j-1} \end{aligned}$$

Iterating this step j times, we get:

$$\hat{x}_{k+j|k} = A^j \hat{x}_{k|k} + \sum_{l=0}^{j-1} A^{j-1-l} Bu_{k+l}$$

Covariance of j-step ahead predictor error

$$
\begin{aligned}
P_{k+j|k} &\triangleq \boldsymbol{E}\{(x_{k+j} - \hat{x}_{k+j|k})(x_{k+j} - \hat{x}_{k+j|k})^T\} \\
&= \boldsymbol{E}\left\{[A(x_{k+j-1} - \hat{x}_{k+j-1|k}) + Gw_{k+j-1}]\right. \\
&\qquad \left.[A(x_{k+j-1} - \hat{x}_{k+j-1|k}) + Gw_{k+j-1}]^T\right\} \\
&= AP_{k+j-1|k}A^T + GQG^T \\
&= A^j P_{k|k}(A^j)^T + \sum_{l=0}^{j-1} A^l GQG^T (A^l)^T
\end{aligned}
$$

Optimal Smoothing

Smoothing means the computation of $\boldsymbol{E}\{x_k|Z^j\}$ for $j > k$. There are different types of smoothers.

- **Fixed-point smoothing**

  Estimate $\boldsymbol{E}\{x_k|Z^j\}$ for some fixed k and for increasing j. Example: estimation of initial condition ($k = 0$).

- **Fixed-lag smoothing**

  Estimate $\boldsymbol{E}\{x_k|Z^{k+N}\}$ for all k and some fixed N. Example: telecommunications.

- **Fixed-interval smoothing**

  Estimate $\boldsymbol{E}\{x_k|Z^N\}$ for all k and for some fixed N. Example: estimate the signal sequence x_k measured with noise, given the whole data set.

Comments on Smoothers

- Often using the whole data set, or at least some window of future data, leads to significant improvement of the quality of the estimates.
- Improvement depends on correlation of signal and on signal/noise ratio.
- All optimal smoothing formulas can be written as a correction term added to the optimal filter formula.
- Here we present only the optimal fixed-point smoother.

Optimal Fixed-Point Smoothing

Consider the model:

$$\begin{cases} x_{k+1} = Ax_k + Gw_k & \text{(i.e. } u_k = 0) \\ y_k = Cx_k + v_k & \end{cases}$$

We construct $\hat{x}_{k|j} \triangleq \boldsymbol{E}\{x_k | Z^j\}$ for some fixed but arbitrary k. Define:

$$\xi_j \triangleq \begin{pmatrix} x_j \\ x_j^a \end{pmatrix} \quad \text{where } x_j^a = x_k^a = x_k \;\; \forall j \geq k$$

Optimal Fixed-Point Smoothing (cont'd)

The model for this augmented state is:

$$\begin{cases} \xi_{j+1} &= \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \xi_j + \begin{pmatrix} G \\ 0 \end{pmatrix} w_j, \quad \xi_k = \begin{pmatrix} x_k \\ x_k \end{pmatrix} \\ y_j &= (C \;\; 0) \;\; \xi_j + v_j \end{cases}$$

$$\Longrightarrow \hat{\xi}_{j+1|j} = \begin{pmatrix} \hat{x}_{j+1|j} \\ \hat{x}_{k|j} \end{pmatrix} \qquad \forall j \geq k$$

Computation of $\hat{x}_{k|j}$ for all $j \geq k$

Compute the one-step ahead Kalman predictor for the augmented state ξ_j . The lower half of $\hat{\xi}_{j+1|j}$ is $\hat{x}_{k|j}$. This yields:

- $\hat{x}_{k|j} = \hat{x}_{k|k} + \sum_{l=k+1}^{j} K_l^{(2)}(y_l - C\hat{x}_{l|l-1})$
- $K_l^{(2)} = P_{l|l-1}^{(21)} C^T (C P_{l|l-1}^{(11)} C^T + R)^{-1}$
- $P_{l|l-1}^{(21)}$ and $P_{l|l-1}^{(11)}$ are the submatrices of the solution of the Riccati equation for the augmented system.

Observation : $\hat{x}_{k|j}$ is the filtered estimate $\hat{x}_{k|k}$ plus a weighted sum of the future innovations.

Augmented Riccati Equation

$$P_{l|l-1} = \begin{pmatrix} P^{(11)}_{l|l-1} & P^{(12)}_{l|l-1} \\ P^{(21)}_{l|l-1} & P^{(22)}_{l|l-1} \end{pmatrix} \text{ with}$$

$$\begin{aligned} P_{l+1|l} &= \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} P_{l|l-1} \begin{pmatrix} A^T & 0 \\ 0 & I \end{pmatrix} + \begin{pmatrix} GQG^T & 0 \\ 0 & 0 \end{pmatrix} \\ & - \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} P_{l|l-1} \begin{pmatrix} C^T \\ 0 \end{pmatrix} \left[(C \;\; 0) P_{l|l-1} \begin{pmatrix} C^T \\ 0 \end{pmatrix} + R \right]^{-1} \\ & \times (C \;\; 0) P_{l|l-1} \begin{pmatrix} A^T & 0 \\ 0 & I \end{pmatrix} \end{aligned}$$

Steady-State Kalman Predictor

The Kalman predictor is a 2-input 1-output filter. It can be written as:

$$\hat{x}_{k+1|k} = (A - K_k C)\hat{x}_{k|k-1} + Bu_k + K_k y_k$$

Question: When does this filter converge to a time-invariant stable filter?

We need:

- $\lim_{k \to \infty} K_k = K_\infty$
- $|\lambda_i(A - K_\infty C)| < 1 \quad \forall\, i$

where $K_k = AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}$

Convergence of Riccati Solution

Consider the Riccati Difference Equation (RDE):

$$\begin{aligned} P_{k+1|k} &= AP_{k|k-1}A^T + GQG^T \\ &\quad -AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T \\ P_{0|-1} &= P_0 \geq 0 \end{aligned}$$

Without loss of generality we can assume $Q > 0$.

Theorem: If $R > 0$, (A, G) is stabilizable and (A, C) is detectable, then $\lim_{k\to\infty} P_{k|k-1} = P_\infty$, where P_∞ is the unique nonnegative definite solution of the Algebraic Riccati Equation (ARE):

$$P = APA^T + GQG^T - APC^T(CPC^T + R)^{-1}CPA^T$$

$$\text{In addition:} \qquad |\lambda_i(A - K_\infty C)| < 1 \quad \forall\, i$$

$$\text{with} \qquad K_\infty = AP_\infty C^T(CP_\infty C^T + R)^{-1}$$

Comments

1. (A, G) stabilizable $\iff \exists\, K$ such that:

$$|\lambda_i(A - GK)| < 1 \ \ \forall\, i$$

   (A, C) detectable $\iff \exists\, K$ such that:

$$|\lambda_i(A - KC)| < 1 \ \ \forall\, i$$

2. The conditions $R > 0$, (A, G) stabilizable, (A, C) detectable can be relaxed. Necessary and sufficient conditions are given in R.R. Bitmead and M. Gevers, "Riccati Difference and Differential Equations : Convergence, monotonicity and stability", in 'The Riccati equation', S. Bittanti, A.J. Laub and J.C. Willems (Eds), Springer Verlag, 1991, pp. 263-291.

Kalman Predictor = Innovations model

Assume that the Kalman gain converges to K_∞ (not essential). We can then rewrite the Kalman predictor as a time-invariant state space model for the process y_k.

Steady state Kalman predictor equations:

$$\begin{aligned}
\hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + Bu_k + K_\infty(y_k - \hat{y}_{k|k-1}) \\
K_\infty &= AP_\infty C^T(CP_\infty C^T + R)^{-1} \\
P_\infty &= AP_\infty A^T + GQG^T - AP_\infty C^T(CP_\infty C^T + R)^{-1}CP_\infty A^T
\end{aligned}$$

Denote $\epsilon_k \triangleq y_k - \hat{y}_{k|k-1}$. Then the Kalman predictor can be rewritten as:

$$\begin{cases} \hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + K_\infty \epsilon_k \\ y_k = C\hat{x}_{k|k-1} + \epsilon_k \end{cases}$$

with $\boldsymbol{E}\{\epsilon_k \epsilon_k^T\} = CP_\infty C^T + R$. This is a new model for the process y_k, driven by the process ϵ_k. It is called the **innovations model**.

More on the Innovations Model

Suppose we start off from a stable *"physical model"* for the process y_k, driven by two independent noise sources, the process noise w_k and the measurement noise v_k:

$$\begin{cases} x_{k+1} & = & Ax_k & + & Bu_k & + & Gw_k \\ y_k & = & Cx_k & + & v_k & & \end{cases}$$

Via the Kalman predictor we have obtained a new *innovations model* for y_k driven by a unique noise source ϵ_k:

$$\begin{cases} \hat{x}_{k+1|k} & = & A\hat{x}_{k|k-1} + Bu_k + K_\infty \epsilon_k \\ y_k & = & C\hat{x}_{k|k-1} + \epsilon_k \end{cases}$$

These are two equivalent models in the sense that the outputs of these two models have identical mean and covariance functions. (It's a useful exercise to verify this !)

Innovations Model in Input-Output Form

The input-output equation of the state space innovations model is:

$$\begin{aligned} y_k &= C(zI-A)^{-1}Bu_k + [C(zI-A)^{-1}K_\infty + I]\epsilon_k \\ &= G(z)u_k + H(z)\epsilon_k \end{aligned}$$

where:

$$G(z) = C(zI-A)^{-1}B \quad \text{and} \quad H(z) = I + C(zI-A)^{-1}K_\infty.$$

The two transfer functions $G(z)$ and $H(z)$ have the same poles, and $H(z)$ is monic. This model is equivalent to an **ARMAX model:**

$$A(z^{-1})y_k = B(z^{-1})u_k + C(z^{-1})\epsilon_k$$

Thus we have justified the generality of ARMAX models for the representation of stationary linear Markov processes, whether these processes have more than one noise source or not.

Properties of the Innovations Model in ARMAX Form

$$\frac{C(z^{-1})}{A(z^{-1})} = H(z) = I + C(zI - A)^{-1}K_\infty$$

$$\frac{B(z^{-1})}{A(z^{-1})} = G(z) = C(zI - A)^{-1}B$$

- Therefore the roots of $C(z^{-1})$ are stable, because they are the poles of
  $$[I + C(zI - A)^{-1}K_\infty]^{-1} = I - C(zI - A + K_\infty C)^{-1}K$$
  and by the Theorem $(A - K_\infty C)$ is a stable matrix.
- Thus the innovations model is stably invertible, i.e. the transfer function $H(z)^{-1} = I - C(zI - A + K_\infty C)^{-1}K$ from the observations sequence y_k to the innovations sequence ϵ_k is stable.
- The sequences $\{y_j, u_j, j \leq k\}$ and $\{\epsilon_j, u_j, j \leq k\}$ contain the same information.

Prediction with ARMAX Models

Consider the ARMAX model:

$$A(z^{-1})y_k = B(z^{-1})u_k + C(z^{-1})e_k$$

where:

$$\begin{aligned} A(z^{-1}) &\triangleq 1 + a_1 z^{-1} + \ldots + a_n z^{-n} \\ B(z^{-1}) &\triangleq b_1 z^{-1} + \ldots + b_m z^{-m} \\ C(z^{-1}) &\triangleq 1 + c_1 z^{-1} + \ldots + c_p z^{-p} \end{aligned}$$

and e_k is white noise with variance σ^2. For simplicity we shall assume that the processes y_k and u_k are scalar. We now construct the best one-step ahead predictor of y_k given the data vector

$$Z^{k-1} \triangleq \{y_j, u_j, j \leq k-1\}$$

The ARMAX model can be re-written as:

$$y_k = [1 - A(z^{-1})]y_k + B(z^{-1})u_k + [C(z^{-1}) - 1]e_k + e_k$$

Since $C(z)$ is monic and since $\{e_j, u_j, j \leq k-1\}$ contains the same information as $\{y_j, u_j, j \leq k-1\}$, the term $[C(z^{-1}) - 1]e_k$ is entirely known at time $k-1$. Therefore:

$$\begin{aligned} \hat{y}_{k|k-1} &= \boldsymbol{E}\{y_k | Z^{k-1}\} \\ &= \boldsymbol{E}\{[1 - A(z^{-1})]y_k + B(z^{-1})u_k + [C(z^{-1}) - 1]e_k + e(k) | Z^{k-1}\} \\ &= [1 - A(z^{-1})]y_k + B(z^{-1})u_k + [C(z^{-1}) - 1]e_k \end{aligned}$$

By replacing e_k above by: $e_k = \frac{A(z^{-1})}{C(z^{-1})}y_k - \frac{B(z^{-1})}{C(z^{-1})}u_k$ we get, after some calculations:

$$C(z^{-1})\hat{y}_{k|k-1} = [C(z^{-1}) - A(z^{-1})]y_k + B(z^{-1})u_k$$

Properties of the One-Step-Ahead Predictor

- The one step ahead predictor

$$C(z^{-1})\hat{y}_{k|k-1} = [C(z^{-1}) - A(z^{-1})]y_k + B(z^{-1})u_k$$

must be interpreted as (assuming $p = n$ for simplicity):

$$\begin{aligned} \hat{y}_{k|k-1} + c_1\hat{y}_{k-1|k-2} + \ldots + c_n\hat{y}_{k-n|k-n-1} &= \\ (c_1 - a_1)y_{k-1} + \ldots + (c_n - a_n)y_{k-n} &+ \\ b_1 u_{k-1} + \ldots + b_m u_{k-m} & \end{aligned}$$

- It can also be rewritten as:

$$\hat{y}_{k|k-1} = [1 - \frac{A(z^{-1})}{C(z^{-1})}]y_k + \frac{B(z^{-1})}{C(z^{-1})}u_k$$

- **It is the optimal steady state Kalman filter for the ARMAX model.**

Properties of the One-Step-Ahead Predictor (cont'd)

- Recall that the roots of $C(z^{-1})$, which are the poles of the Kalman filter, are given by

$$\text{roots of } \; C(z^{-1}) = \; \text{eigenvalues of } \; (zI - A + K_\infty C)$$

- For general input-output models of the form

$$y_k = G(z)u_k + H(z)e_k \quad \text{with} \quad e_k \quad \text{white noise,}$$

the optimal one-step-ahead Kalman predictor is:

$$\hat{y}_{k|k-1} = H(z)^{-1}G(z)u_k + [1 - H(z)^{-1}]y_k.$$

Observe that this is consistent with the special case of an ARMAX model.

Stochastic Convergence Notions

1. Convergence in probability

Definition:
Let $X_k \in R^n, k \in Z_+$ be a random sequence. X_k is said to converge in probability to X if

$$\forall \, \epsilon > 0 : \lim_{k \to \infty} p[||X_k - X|| > \epsilon] = 0$$

One often writes $p \lim_{k \to \infty} X_k = X$.
Convergence in probability is the weakest convergence concept.

2. Mean-Square Convergence

Definition:
Let $X_k \in R^n, k \in Z_+$ be a random sequence. X_k is said to converge in mean square to X if

$$\lim_{k \to \infty} \boldsymbol{E}\{||X_k - X||^2\} = 0$$

One often writes $l.i.m.X_k = X$ where $l.i.m.$ means "limit in the mean". Note that if X_k is an estimator of Θ and if $X = \Theta$, then mean square convergence of $\hat{\Theta}_k$ to Θ implies that both the bias and the variance tend to zero asymptotically.

3. Almost Sure Convergence

Every realization of a random sequence is produced by an event ω in the space of events Ω. Therefore, for a given $\omega, x_1(\omega), x_2(\omega), \ldots, x_k(\omega)$ is a sequence of numbers that may (or may not) converge to a number $x(\omega)$. If for all ω, the realization $x_k(\omega)$ converges to some $x(\omega)$, possibly different for each ω, then the random sequence X_k is said to converge X_k everywhere. This definition, called **convergence everywhere** is too strong and is replaced by **convergence almost everywhere**, or "almost sure convergence" or "convergence with probability one".

3. Almost Sure Convergence (cont'd)

Definition:

Let $X_k \in R^n, k \in Z_+$ be a random sequence. X_k converges to X almost everywhere if $p[X_k(\omega) \to_{k\to\infty} X(\omega)] = 1$. More precisely:

$$x_k(\omega) \to_{k\to\infty} x(\omega) \;\; \forall\, \omega \in A \subset \Omega, \quad with\, P(A) = 1$$

.

One often writes: $X_k \to X \;\; a.s., \; a.e. \; ou \; w.p.1$

$(a.s. =$ almost surely$, a.e. =$ almost everywhere$, w.p.1 =$
with probability one.