Chapter 4

Kalman Filter

The problem of stochastic filtering is to estimate the state of a random sequence X_t at a given time t basing on related data Y_t obtained by noisy observations until time t. $P(X_t|Y_0, \dots, Y_t)$.

For a *linear state and observation model*, this problem have been solved by R. Kalman in discrete time case (1960) and by Kalman and Bucy (1961) in continuous time case.

The solution is called the Kalman filter. It has many applications in engineering and economics.

The main advantages of Kalman filter:

1. Recursive form of the filtering equations.
2. Easy extension to vector and non stationary cases.

4.1 Linear State and Observation Model

4.1.1 Scalar Case

First we consider a scalar process $\{X_t\}$, which is described by linear difference equation,

$$X_{t+1} = aX_t + V_{t+1}, \quad E(X_0) = m_0, \quad Var(X_0) = \gamma_0, \tag{4.1}$$

process $\{X_t\}$ is said to be a *state process*. Another process $\{Y_t\}$ referred to as an *observation process* is also described by linear difference equation

$$Y_{t+1} = bX_t + W_{t+1}, \quad Y_0 = 0. \tag{4.2}$$

In (4.1), (4.2) the initial condition X_0 and noises $W_t \sim WN(0, \sigma_W^2)$ and $V_t \sim WN(0, \sigma_V^2)$ are uncorrelated, and a, b are known constants.

The pair of equations (4.1), (4.2) constitutes the *linear state and observation model*.

Filtering problem

We assume that process $\{X_t\}$ is unobservable and we have to find the best linear predictor of X_t given $Y_1, \ldots, Y_t = \mathbf{Y}_t$, $P(X_t|\mathbf{Y}_t)$.

If the joint distribution of noises and initial condition X_0 is Gaussian, then $P(X_t|\mathbf{Y}_t) = E(X_t|\mathbf{Y}_t)$.

Recursive Filtering Equations

Denote the set of observations up to time t as

$$\mathbf{Y}_t = (Y_1, \ldots, Y_t),$$

the best predictor

$$\hat{X}_t = E(X_t|\mathbf{Y}_t),$$

and the mean squared estimation error

$$v_t^2 = E((X_t - \hat{X}_t)^2|\mathbf{Y}_t).$$

Using the linear model, and the theorem on the orthogonal projection for the best predictor (see Theorem 1.7), we get a recursion for $(\hat{X}_t, v_t^2)$.

From linear model we have

$$E(X_{t+1}|\mathbf{Y}_t) = a\hat{X}_t,\ E(Y_{t+1}|\mathbf{Y}_t) = b\hat{X}_t,$$

and

$$\begin{aligned} X_{t+1} - E(X_{t+1}|\mathbf{Y}_t) &= a(X_t - \hat{X}_t) + V_{t+1}, \\ Y_{t+1} - E(Y_{t+1}|\mathbf{Y}_t) &= b(X_t - \hat{X}_t) + W_{t+1}. \end{aligned}$$

Then, taking into account that $X_t, \hat{X}_t, \mathbf{Y}_t$ and V_{t+1}, W_{t+1} are uncorrelated (independent in Gaussian case) we can calculate the conditional covariance matrix:

$$\begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

where

$$d_{11} = cov(X_{t+1}, X_{t+1}|\mathbf{Y}_t) = E\{(X_{t+1} - E(X_{t+1}|\mathbf{Y}_t))^2|\mathbf{Y}_t)\}$$

$$E\{(a(X_t - \hat{X}_t) + V_{t+1})^2|\mathbf{Y}_t\} = a^2 v_t^2 + \sigma_V^2.$$

and in the same way

$$d_{12} = cov(X_{t+1}, Y_{t+1}|\mathbf{Y}_t) = abv_t^2,$$
$$d_{22} = cov(Y_{t+1}, Y_{t+1}|\mathbf{Y}_t) = b^2 v_t^2 + \sigma_W^2.$$

Using the best linear prediction theorem (Theorem 1.7 of normal correlation in Gaussian case) we get the following recurrent equations

$$\hat{X}_{t+1} = E(X_{t+1}|\mathbf{Y}_t, Y_{t+1}) = E(X_{t+1}|\mathbf{Y}_t) + \frac{d_{12}}{d_{22}}(Y_{t+1} - E(Y_{t+1}|\mathbf{Y}_t))$$

$$= a\hat{X}_t + \frac{abv_t^2}{b^2 v_t^2 + \sigma_W^2}(Y_{t+1} - b\hat{X}_t), \tag{4.3}$$

$$v_{t+1}^2 = cov(X_{t+1}, X_{t+1}|\mathbf{Y}_t, Y_{t+1}) = d_{11} - \frac{d_{12}^2}{d_{22}} = a^2 v_t^2 + \sigma_v^2 - \frac{(abv_t^2)^2}{b^2 v_t^2 + \sigma_W^2}$$

known as *discrete Kalman filter*.

Equations (4.3) must be solved for $t = 1, 2, \ldots$ with initial conditions

$$\hat{X}_0 = E(X_0) = m_0, \quad v_0^2 = Var(X_0) = \gamma_0.$$

Remarkable Features of Kalman Filter

1. Recursive form of equations permits to realize them without huge amount of memory. In order to calculate the next estimate $(\hat{X}_{t+1}, v_{t+1}^2)$ one have to know only the previous estimate $(\hat{X}_t, v_t^2)$ and update them when the new observation Y_{t+1} comes.

2. Equation for v_t^2 does not depend on observations, so the second equation in Kalman filter (called the *discrete Riccati equation*) can be solved in advance. Solution of this equation permits to estimate the possible mean squared error of filtering and therefore to design the experiment.

3. In the Gaussian case equations (4.3) give the best estimate, however, the same equations give the best linear estimate in non-gaussian case. It means that Kalman filter is robust with respect to the noise distribution.

4. The same type of filtering equations can be obtained even in non-stationary case for vector processes.

Other form of Kalman filter

Notice that first equation (4.3) it is a weighted average of the previous estimate and current observation:

$$\hat{X}_{t+1} = \frac{\sigma_W^2}{b^2 v_t^2 + \sigma_W^2} \hat{X}_t + \frac{abv_t^2}{b^2 v_t^2 + \sigma_W^2} Y_{t+1}.$$

Steady-state Solution

It can be shown that v_t^2 converges to $v^2 < \infty$ (steady-state) as $t \to \infty$.

Denote $v^2 = \Gamma$ then Γ satisfies the equation

$$\Gamma = a^2\Gamma + \sigma_V^2 - \frac{(ab)^2\Gamma^2}{b^2\Gamma + \sigma_W^2},$$

which can be reduces to quadratic equation

$$\Gamma^2 + \left[\frac{\sigma_W^2(1-a^2)}{b^2} - \sigma_V^2\right]\Gamma - \frac{\sigma_V^2\sigma_W^2}{b^2} = 0.$$

If $\sigma_V^2 > 0, \sigma_W^2 > 0, b^2 > 0$ (otherwise the filtering problem is singular), this equation always has two real solutions $\Gamma_1 > 0$ and $\Gamma_2 < 0$, and the solution $\Gamma_1 = \lim_{t\to\infty} v_t^2$.

Then one can replace the time-varying coefficient in the filter equation by its steady-state value, which gives

$$\frac{abv_t^2}{b^2 v_t^2 + \sigma_W^2} \to \frac{ab\Gamma_1}{b^2\Gamma_1 + \sigma_W^2} := \delta$$

and the filter itself can be reduced to its steady-state form

$$\hat{X}_{t+1} = a\hat{X}_t + \delta(Y_{t+1} - b\hat{X}_t).$$

Example 4.1 Random Walk with Noise.

This is a case $a = 1$ in the state equation

$$X_{t+1} = X_t + V_{t+1},$$

and $b = 1$ in the observation equation

$$Y_{t+1} = X_t + W_{t+1}.$$

Then the equation for steady-state solution has a form

$$\Gamma^2 - \sigma_V^2\Gamma - \sigma_V^2\sigma_W^2 = 0.$$

Solving this equation we obtain the positive solution

$$\Gamma = \frac{1}{2}\left(\sigma_V^2 + \sqrt{\sigma_V^2 + 4\sigma_V^2\sigma_W^2}\right),$$

and

$$\delta = \frac{v^2}{v^2 + \sigma_W^2} = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_W^2}.$$

The best predictor at time t of X_t based on steady-state filter is

$$\hat{X}_t = \hat{X}_{t-1} + \delta(Y_t - \hat{X}_{t-1})$$

or

$$\hat{X}_t = (1-\delta)\hat{X}_{t-1} + \delta Y_t.$$

But this is the equation for exponential smoothing filter (see [4] p. 27).

4.2 Vector Case

4.2.1 Linear State-space Model

In practice we deal with systems that described by vector parameters, and such systems are just discussed below.

We use the notation $\{\mathbf{W}_t\} \sim WN(\mathbf{0}, \{R_t\})$ for multidimensional white noise with mean $\mathbf{0}$ and covariance matrix

$$\mathrm{E}(\mathbf{W}_s\mathbf{W}_t^T) = \begin{cases} R_t, & \text{if } s = t, \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

In the case of two dimensions for white noises we have the following. Let

$$\mathbf{W}_t = \begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix}$$

where $W_{1,t}$ and $W_{2,t}$ are two white noises, both with mean 0 and with covariance matrix defined as follows.

$$\mathrm{E}\mathbf{W}_t\mathbf{W}_t^T = \begin{pmatrix} \mathrm{E}W_{1,t}^2 & \mathrm{E}W_{1,t}W_{2,t} \\ \mathrm{E}W_{1,t}W_{2,t} & \mathrm{E}W_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2(t) & \sigma_{1,2}(t) \\ \sigma_{1,2}(t) & \sigma_2^2(t) \end{pmatrix} = R_t, \tag{4.5}$$

and $\mathrm{E}\mathbf{W}_s\mathbf{W}_t^T = 0$ (null matrix) as $s \neq t$.

Notice, that unlike the case studied before the variances $\mathrm{E}W_{1,t}^2$ and $\mathrm{E}W_{2,t}^2$ as well as the covariance $\mathrm{E}W_{1,t}W_{2,t}$ all are assumed to be dependent of t.

However in the most examples below the white noises are considered to be time independent nevertheless.

State-Space Representations

We consider to vector processes $\mathbf{X}_t$ and $\mathbf{Y}_t$. For the vector process $\mathbf{Y}_t$ we consider two equations defining *time-space model*. The first equation is called the **observation equation**:

$$\mathbf{Y}_{t+1} = G_t\mathbf{X}_t + \mathbf{W}_{t+1}, \quad t = 0, 1, 2, \dots, \qquad \mathbf{Y}_0 = 0, \tag{4.6}$$

where $\{W_t\} \sim WN(\mathbf{0}, \{R_t\})$, and G_t is a sequence of matrices. The dimension of G_t is $w \times v$ (w is the dimension of white noise, v is the number of observations).

The second equation is

$$\mathbf{X}_{t+1} = F_t\mathbf{X}_t + \mathbf{V}_{t+1}, \quad t = 0, 1, 2, \dots, \qquad \mathbf{X}_0 \sim N(m_0, \gamma_0), \tag{4.7}$$

where $\{F_t\}$ is a sequence of $v \times v$ matrices, and $\{\mathbf{V}_t\} \sim WN(\mathbf{0}, \{Q_t\})$, and $\{\mathbf{V}_t\}$ is assumed to be uncorrelated with $\{\mathbf{W}_t\}$. That is $\mathrm{E}W_tV_s' = 0$ (null-matrix) for all s and t. It is assumed additionally that $\mathbf{X}_0$ is uncorrelated with all of the noise terms $\{\mathbf{V}_t\}$ and $\{\mathbf{W}_t\}$. The second equation (4.7) is called the **state equation**.

Statement of the Optimal Filtering Problem

The problem of the best in mean square prediction of $\mathbf{X}_t$ based on previous observations $\mathbb{Y}_t = (\mathbf{Y}_1, ..., \mathbf{Y}_t)$, is said to be the *optimal filtering problem.* Solution of this problem is given by

$$\hat{\mathbf{X}}_t = E(\mathbf{X}_t|\mathbb{Y}_t).$$

In Gaussian case this estimate is a linear function of $(\mathbf{Y}_1, ..., \mathbf{Y}_t)$. In the non-gaussian cases the same formula gives the best *linear* mean square prediction.

The formulas for the estimator have a recurrent form and are represented by Kalman filter.

Kalman Filter (Vector Case)

Assume that $\hat{\mathbf{X}}_t = E(\mathbf{X}_t|\mathbb{Y}_t)$ and

$$\Gamma_t = cov[(\mathbf{X}_t - \hat{\mathbf{X}}_t)(\mathbf{X}_t - \hat{\mathbf{X}}_t)^T|\mathbb{Y}_t]$$

are calculated and are based on observations up to time t.

Then the estimate of $\hat{\mathbf{X}}_{t+1} = E(\mathbf{X}_{t+1}|\mathbb{Y}_{t+1})$ and its conditional covariance Γ_{t+1} that are based on $\mathbb{Y}_{t+1}$ are the solutions of the following recurrent equations

$$\begin{aligned}
\hat{\mathbf{X}}_{t+1} &= F_t\hat{\mathbf{X}}_t + F_t\Gamma_t G_t^T[R_{t+1} + G_t\Gamma_t G_t^T]^{-1}(Y_{t+1} - G_t\hat{\mathbf{X}}_t), \\
\Gamma_{t+1} &= F_t\Gamma_t F_t^T + Q_{t+1} - F_t\Gamma_t G_t^T[R_{t+1} + G_t\Gamma_t G_t^T]^{-1}G_t\Gamma_t F_t^T.
\end{aligned} \tag{4.8}$$

These equations have to be solved for $t = 0, 1, ...$ with initial conditions

$$\hat{\mathbf{X}}_0 = m_0, \quad \Gamma_0 = \gamma_0.$$

In the Gaussian case these equations give the best mean square estimation of $\mathbf{X}_t$ based on $\mathbb{Y}_t$. In non-gaussian case the same equations give the best mean square linear prediction based on $\mathbb{Y}_t$.

4.2.2 h-step Prediction

The problem of the best in mean square prediction of $\mathbf{X}_{t+h}$ based on previous observations $\mathbb{Y}_t = (\mathbf{Y}_1, ..., \mathbf{Y}_t)$, is said to be the *optimal h-step prediction problem.* Solution of this problem is given by

$$\hat{\mathbf{X}}_t = E(\mathbf{X}_{t+h}|\mathbb{Y}_t).$$

In Gaussian case this estimate is a linear function of $(\mathbf{Y}_1, ..., \mathbf{Y}_t)$. In the non-gaussian cases the same formula gives the best *linear* mean square prediction.

The formulas for the estimator have the same recurrent form as Kalman filter.

Formulas for h-step Prediction

The linear state-observation model gives the simple set of equations, having the same linear form. Assume that $\hat{\mathbf{X}}_t = E(\mathbf{X}_t|\mathbb{Y}_t)$ and

$$\Gamma_t = cov[(\mathbf{X}_t - \hat{\mathbf{X}}_t)(\mathbf{X}_t - \hat{\mathbf{X}}_t)^T|\mathbb{Y}_t]$$

are calculated and are based on observations up to time t. Then for $k = 1, ..., h$ the estimate $\hat{\mathbf{X}}_{t+k} = E(\mathbf{X}_{t+h}|\mathbb{Y}_t)$ and

$$\Gamma_{t+k} = cov[(\mathbf{X}_{t+k} - \hat{\mathbf{X}}_{t+k})(\mathbf{X}_{t+k} - \hat{\mathbf{X}}_{t+k})^T|\mathbb{Y}_t]$$

can be calculated with the aid of following *deterministic* systems of recurrent equations

$$\begin{aligned}
\hat{\mathbf{X}}_{t+k} &= F_{t-1+k}\hat{\mathbf{X}}_{t-1+k} \\
\Gamma_{t+k} &= F_{t-1+k}\Gamma_{t-1+k}F_{t-1+k}^T + Q_{t+1},
\end{aligned} \tag{4.9}$$

with *random* initial conditions

$$\hat{\mathbf{X}}_t \quad \text{and} \quad \Gamma_t.$$

4.2.3 Application of Kalman Filter to the Parameter Estimation

Vector parameter estimation problem arises in image restoration, physical measurements, identification etc.

Problem Statement and Relation with Kalman Filtering

Assume that unknown (unobservable) vector θ is a random vector in R^n with the parameters

$$E(\theta) = m_\theta, \quad cov(\theta, \theta) = \Gamma_\theta.$$

Observations are the sequence of values

$$Y_{t+1} = a_t^T \theta + W_{t+1},$$

where a_t is a sequence of known row-vectors in R^n and $W_t \sim WN(0, \sigma^t)$. The problem is to find best mean square linear estimation of θ basing on observations available up to time t.

To reduce this problem to a general case take the state equation as

$$\theta_{t+1} = \theta_t,$$

then this model has a standard linear state-space representation with

$$F_t \equiv I \quad \text{identity matrix}, \quad Q_t \equiv 0 \quad \text{null matrix}, \quad G_t = a_t^T.$$

The best mean square estimation

$$\hat{\theta}_t = E(\theta|\mathbb{Y}_t)$$

and its conditional covariance

$$\Gamma_t = cov[(\theta, \theta)|\mathbb{Y}_t] = E[(\theta - \hat{\theta}_t)(\theta - \hat{\theta}_t)^T|\mathbb{Y}_t]$$

can be calculated from the following system of recurrent equations as a particular case of general equations (4.8)

$$\begin{aligned} \hat{\theta}_{t+1} &= \hat{\theta}_t + \Gamma_t a_t[\sigma_t^2 + a_t^T \Gamma_t a_t]^{-1}(Y_{t+1} - a_t^T \hat{\theta}_t), \\ \Gamma_{t+1} &= \Gamma_t - \Gamma_t a_t[\sigma_t^2 + a_t^T \Gamma_t a_t]^{-1} a_t^T \Gamma_t. \end{aligned} \tag{4.10}$$

These equations have to be solved for $t = 0, 1, ...$ with initial conditions

$$\hat{\theta}_0 = m_0, \quad \Gamma_0 = \gamma_0.$$

Explicit Solution for Parameter Estimation

Equations (4.10) admit the explicit solution which helps to investigate either the parameter estimation procedure is consistent or not. Solution to the equations (4.10) has a form

$$\hat{\theta}_{t+1} = \left[I + \gamma_0 \sum_{k=0}^{t} a_k a_k^T \sigma^{-2} \right]^{-1} \left[m_0 + \gamma_0 \sum_{k=0}^{t} a_k \sigma^{-2} Y_{k+1} \right],$$

$$\Gamma_{t+1} = \left[I + \gamma_0 \sum_{k=0}^{t} a_k a_k^T \sigma^{-2} \right]^{-1} \gamma_0. \tag{4.11}$$

Definition 4.1 The estimation procedure is *consistent* if

$$cov(\theta|\mathbb{Y}_t) \to 0$$

as $t \to \infty$.

Proposition 4.1 If all diagonal elements of matrix $\sum_{k=0}^{t} a_k a_k^T$ tend to infinity as $t \to \infty$ then the estimation procedure is consistent.

PROOF: In this case the all diagonal elements of Γ_t tend to zero as $t \to \infty$ and it implies that $\Gamma_t \to 0$.

□

Example 4.2 Parameter Estimation (Scalar Parameter)
In this case

$$\theta \in R^1, \quad a_t \equiv a \in R^1 \quad \text{and} \quad \sigma_t^2 = sigma^2.$$

The estimation procedure for the parameter θ gives

$$\hat{\theta}_{t+1} = \hat{\theta}_{t+1} \left[1 - \frac{a^2 \Gamma_t}{\sigma^2 + a^2 \Gamma_t} \right] + \frac{a \Gamma_t}{\sigma^2 + a^2 \Gamma_t} Y_{t+1},$$

$$\Gamma_{t+1} = \frac{\sigma^2 \Gamma_t}{\sigma^2 + a^2 \Gamma_t}. \tag{4.12}$$

From equations (4.11) we obtain the explicit solution

$$\hat{\theta}_{t+1} = \frac{m_0 \sigma^2 + \gamma_0 a \sum_{k=1}^{t+1} Y_k}{\sigma^2 + \gamma_0 a^2 (t+1)}, \tag{4.13}$$

$$\Gamma_{t+1} = \frac{\gamma_0 \sigma^2}{\sigma^2 + \gamma_0 a^2 (t+1)}.$$

Therefore, $\Gamma_t \to 0$ as $t \to \infty$.

Example 4.3 In Gaussian case how many observation we need to achieve the given level of accuracy ε with confident level 95%?

SOLUTION: The conditional distribution of $\theta - \hat{\theta}_t \sim N(0, \Gamma_t)$. If we need

$$P\{|\theta - \hat{\theta}_t| \leq \varepsilon\} \geq 0.95,$$

then it means that

$$1.96(\Gamma_t)^{1/2} \leq \varepsilon,$$

and therefore we have to choose t, from condition

$$\frac{\gamma_0 \sigma^2}{\sigma^2 + \gamma_0 a^2 t} \leq \frac{\varepsilon^2}{(1.96)^2} \approx \frac{\varepsilon^2}{4}.$$

□