Discrete- and Continuous-Time Kalman Filters *

(COMS 4770/5770 Notes)

Yan-Bin Jia

May 18, 2026

1 Discrete-Time System

A *discrete-time system* is a process that transforms input discrete-time signals into output discrete-time signals. Simply stated, it takes an input sequence and produces an output sequence. Such a system usually takes the form of

$$\boldsymbol{x}_i = F_{i-1}\boldsymbol{x}_{i-1} + G_{i-1}\boldsymbol{u}_{i-1} + \boldsymbol{w}_{i-1}, \tag{1}$$

where the n-vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_{i-1}$ are the states at the current and previous time steps, the l-vector $\boldsymbol{u}_i$ is a known input, and the n-vector $\boldsymbol{w}_i$ is white noise with zero mean and covariance Q_i. The matrices F_{i-1} and G_{i-1} have dimensions $n \times n$ and $n \times l$, respectively. We are interested in how the mean of the state $\boldsymbol{x}_i$ evolves with time.

From now on, we denote the mean of a random variable with an overbar '$\bar{}$'. Take the expected value of both sides of equation (1):

$$\bar{\boldsymbol{x}}_i = E(\boldsymbol{x}_i) = F_{i-1}\bar{\boldsymbol{x}}_{i-1} + G_{i-1}\boldsymbol{u}_{i-1}. \tag{2}$$

How does the covariance of $\boldsymbol{x}_i$ evolve with time? Subtraction of (2) from (1) yields

$$\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i = F_{i-1}(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1}) + \boldsymbol{w}_{i-1}.$$

The covariance P_i as an $n \times n$ matrix is then computed as follows:

$$\begin{aligned}
P_i &= E\left((\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i)^\top\right) \\
&= E\left(\Big(F_{i-1}(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1}) + \boldsymbol{w}_{i-1}\Big)\Big(F_{i-1}(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1}) + \boldsymbol{w}_{i-1}\Big)^\top\right) \\
&= E\Big(F_{i-1}(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1})(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1})^\top F_{i-1}^\top + \boldsymbol{w}_{i-1}\boldsymbol{w}_{i-1}^\top + \\
&\qquad F_{i-1}(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1})\boldsymbol{w}_{i-1}^\top + \boldsymbol{w}_{i-1}(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1})^\top F_{i-1}^\top\Big).
\end{aligned}$$

*The material is adapted from Chapters 4.1, 5.1–5.3, and 8.1–8.2 in Dan Simon's book *Optimal State Estimation* [1].

That $\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1}$ and $\boldsymbol{w}_{i-1}$ are uncorrelated implies

$$\begin{aligned} E\left((\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1})\boldsymbol{w}_{i-1}^\top\right) &= E(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1})E(\boldsymbol{w}_{i-1}^\top) \\ &= 0, \qquad\qquad (\text{since } E(\boldsymbol{w}_{i-1}^\top) = 0) \end{aligned}$$

and thus

$$E\left(\boldsymbol{w}_{i-1}(\boldsymbol{x}_{i-1} - \bar{\boldsymbol{x}}_{i-1})^\top\right) = 0.$$

Two of the summands in the equation for P_i vanish. The recursive covariance equation reduces to

$$P_i = F_{i-1}P_{i-1}F_{i-1}^\top + Q_{i-1}, \tag{3}$$

where $Q_{i-1} = E(\boldsymbol{w}_{i-1}\boldsymbol{w}_{i-1}^\top)$ is the covariance of $\boldsymbol{w}_{i-1}$. The above is called a *discrete-time Lyapunov equation* or a *Stein equation*.

EXAMPLE 1. The following is a linear system that describes the population of a predator x_1 and that of its prey x_2, where the second subscript of each item denotes the time step:

$$\begin{aligned} x_{1,i+1} &= x_{1,i} - 0.8x_{1,i} + 0.4x_{2,i} + w_{1,i}, \\ x_{2,i+1} &= x_{2,i} - 0.4x_{1,i} + u_i + w_{2,i}. \end{aligned}$$

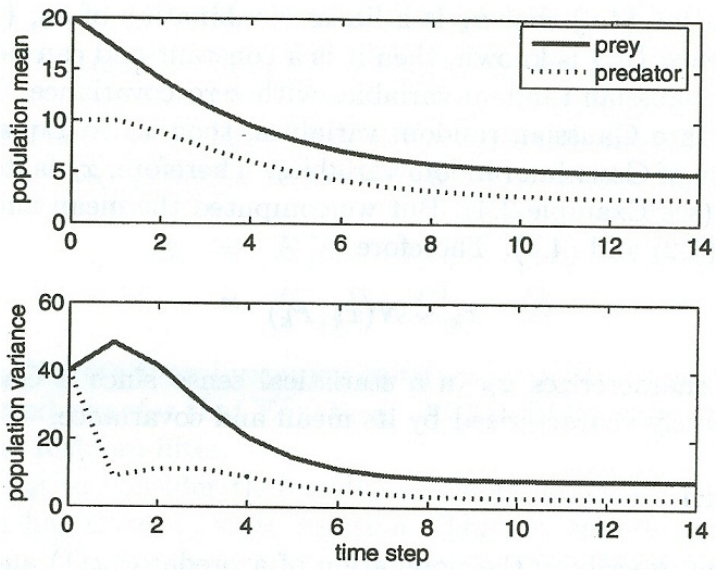
The first equation says that the predator population decreases due to overcrowding as well as the reduction in the prey population. The second equation says that the prey population decreases due to the predator population and increases on external food supply u_i. Both populations are subject to random disturbances (with variances 1 and 2, respectively) due to environmental factors. We rewrite the system in the state space form as

$$\begin{aligned} \boldsymbol{x}_{i+1} &= \begin{pmatrix} 0.2 & 0.4 \\ -0.4 & 1 \end{pmatrix}\boldsymbol{x}_i + \begin{pmatrix} 0 \\ 1 \end{pmatrix}u_i + \boldsymbol{w}_i, \\ \boldsymbol{w}_i &\sim (0, Q), \end{aligned}$$

where $Q = \mathrm{diag}(1, 2)$.

The mean and covariance of the populations evolve with time according to equations (2) and (3). Set $u_i = 1$ and the initial conditions as $\bar{x}_0 = (10, 20)^\top$ and $P_0 = \mathrm{diag}(40, 40)$. The figure[1] below depicts the two means and the two diagonal elements of the covariance matrix for the first few steps.

[1]Figure 4.1 on p. 110 in [1].



From the figure, we see that the mean and covariance eventually reach the steady-state values. The steady state value of the mean is the solution of the linear system after setting $\boldsymbol{w}_i = 0$:

$$\bar{\boldsymbol{x}} = \begin{pmatrix} 2.5 \\ 5 \end{pmatrix}.$$

The steady-state value of P found by control system software[2] is

$$P \approx \begin{pmatrix} 2.88 & 3.88 \\ 3.08 & 7.96 \end{pmatrix}.$$

2 Four Estimates of a State

We would like to estimate the state, i.e., the values of all variables of a dynamic system as it evolves with time. Our method is to derive equations from the system to describe how the mean and covariance of its state $\boldsymbol{x}$ propagate with time. These equations form a new dynamic system which is simulated on a computer. They are the basis for the derivation of the Kalman filter whose estimate is the mean of the original system's state $\boldsymbol{x}$, and whose covariance equals the covariance of $\boldsymbol{x}$. Every time a new measurement is taken, the mean and covariance of $\boldsymbol{x}$ are updated, in a similar way we recursively update the estimate of a constant before.

Consider a linear discrete-time system as follows:

$$\boldsymbol{x}_i = F_{i-1}\boldsymbol{x}_{i-1} + G_{i-1}\boldsymbol{u}_{i-1} + \boldsymbol{w}_{i-1}, \tag{4}$$

$$\boldsymbol{y}_i = H_i\boldsymbol{x}_i + \boldsymbol{v}_i, \tag{5}$$

where the output $\boldsymbol{y}_i$ is an $m \times 1$ vector and H_i an $m \times n$ matrix. Here, the evolution of the state $\boldsymbol{x}$ and the output $\boldsymbol{y}$ are subject to some noise processes $\boldsymbol{w}$ and $\boldsymbol{v}$, respectively. These two noise are white, zero-mean, and uncorrelated. Let Q_i and R_i be their covariance matrices, respectively.

[2]For example, the MATLAB Control System Toolbox function DLYAP(F, Q).

Then the noise characteristics are given as follows:

$$\begin{aligned}
\boldsymbol{w}_i &\sim (0, Q_i), && (6)\\
\boldsymbol{v}_i &\sim (0, R_i), && (7)\\
E(\boldsymbol{w}_j \boldsymbol{w}_k^\top) &= \begin{cases} Q_j, & \text{if } j = k; \\ 0, & \text{otherwise,} \end{cases} && (8)\\
E(\boldsymbol{v}_j \boldsymbol{v}_k^\top) &= \begin{cases} R_j, & \text{if } j = k; \\ 0, & \text{otherwise,} \end{cases} && (9)\\
E(\boldsymbol{v}_j \boldsymbol{w}_k^\top) &= 0. && (10)
\end{aligned}$$

The goal is to estimate the state $\boldsymbol{x}_i$ based on the known system dynamics (4) and (5) and from the noisy measurements $\{\boldsymbol{y}_i\}$.

In estimating the state $\boldsymbol{x}_i$ at time i, if we have all the measurements up to and including time i available, then we can form an *a posteriori* (or a posterior) estimate denoted as $\hat{\boldsymbol{x}}_i^+$. One way is to compute the expected value of $\boldsymbol{x}_i$ as

$$\hat{\boldsymbol{x}}_i^+ = E(\hat{\boldsymbol{x}}_i \mid \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_i). \tag{11}$$

If we have only the measurements before time i available, then we can form *a priori* (or a prior) estimate denoted as $\hat{\boldsymbol{x}}_i^-$. One way to form this latter estimate is to compute the expected value of $\boldsymbol{x}_i$ as

$$\hat{\boldsymbol{x}}_i^- = E(\hat{\boldsymbol{x}}_i \mid \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{i-1}). \tag{12}$$

Note that $\hat{\boldsymbol{x}}_i^-$ and $\hat{\boldsymbol{x}}_i^+$ are both estimates of $\boldsymbol{x}_i$. However, the former is our estimate before the measurement $\boldsymbol{y}_i$ is taken into account, while the latter is our estimate after the measurement $\boldsymbol{y}_i$ is taken into account. With more information, $\hat{\boldsymbol{x}}_i^+$ is expected to be a better estimate than $\hat{\boldsymbol{x}}_i^-$.

A *smoothed estimate* of $\boldsymbol{x}_i$ is formed if we also have measurements after time i available for use. One way of doing this is to compute the following expected value:

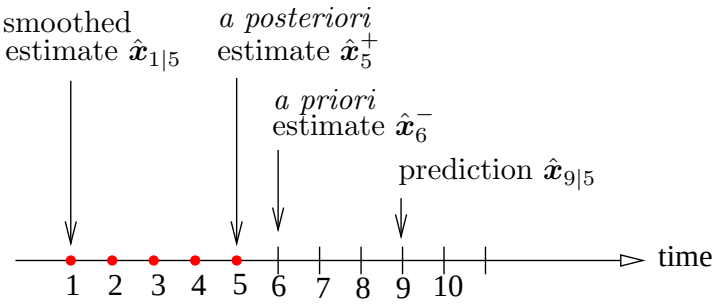
$$\hat{\boldsymbol{x}}_{i|i+s} = E(\hat{\boldsymbol{x}}_i \mid \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{i+s}), \tag{13}$$

where $s > 0$ depends on the specific problem that is being solved. A *predicted estimate* is a prediction of $\boldsymbol{x}_i$ one step or more ahead of the available measurements. This is often computed as the expected value of $\boldsymbol{x}_i$ based on the limited measurements that are available:

$$\hat{\boldsymbol{x}}_{i|i-t} = E(\hat{\boldsymbol{x}}_i \mid \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{i-t}), \tag{14}$$

where $t > 0$ is problem specific.

The figure below depicts the relationship between the posterior, prior, smoothed, and predicted state estimates. Suppose we are making estimates based on all measurements taken at time $i \leq 5$. An estimate of the state at $i = 1$ is a smoothed estimate. An estimate at $i = 5$ is the *a posteriori* estimate. An estimate at $i = 6$ is the prior estimate. And an estimate of the state at $i = 9$ is a prediction.



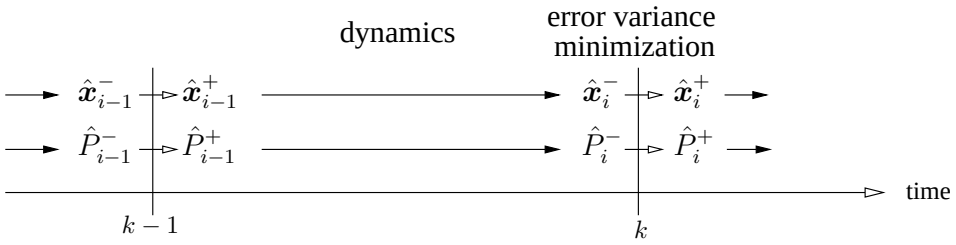
Starting from now we denote $\hat{\boldsymbol{x}}_0^+$ as our initial estimate of $\boldsymbol{x}_0$ before the first measurement is taken (at time $i = 1$). Given no measurements available, we form $\hat{\boldsymbol{x}}_0^+$ as the expected value of the initial estimate:

$$\hat{\boldsymbol{x}}_0^+ = \bar{\boldsymbol{x}}_0 = E(\boldsymbol{x}_0). \tag{15}$$

We let P_i^- be the covariance of the estimation error of $\hat{\boldsymbol{x}}_i^-$, and P_i^+ the covariance of the estimation error of $\hat{\boldsymbol{x}}_i^+$, all at time i:

$$\begin{aligned} P_i^- &= E\Big((\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i^-)(\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i^-)^\top\Big), \\ P_i^+ &= E\Big((\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i^+)(\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i^+)^\top\Big). \end{aligned}$$

The next figure shows the relationships from time $i-1$ to i. After we process the measurement at time $i-1$, we have *a posteriori* estimate $\hat{\boldsymbol{x}}_{i-1}^+$ and the covariance P_{i-1}^+. At time i, before we process the new measurement we compute a prior estimate $\hat{\boldsymbol{x}}_i^-$ and the covariance P_i^- of the estimate based on the system dynamics. Then we process the measurement $\boldsymbol{y}_i$ at time i to refine our estimate of $\boldsymbol{x}_i$. The resulting estimate that takes into account the measurement at time i is $\hat{\boldsymbol{x}}_i^+$ with covariance P_i^+.



3 From *a posteriori* Estimate to *a priori* Estimate

The estimation process starts with $\hat{\boldsymbol{x}}_0^+$, the best estimate of the initial state $\boldsymbol{x}_0$. It then computes $\hat{\boldsymbol{x}}_1^-$. Recall from (2) that the mean of $\boldsymbol{x}$ propagates with time, so we obtain

$$\hat{\boldsymbol{x}}_1^- = F_0\hat{\boldsymbol{x}}_0^+ + G_0\boldsymbol{u}_0.$$

This equation generalizes to obtain the prior estimate at time i from the posterior estimate at the previous time $i-1$:

$$\hat{\boldsymbol{x}}_i^- = F_{i-1}\hat{\boldsymbol{x}}_{i-1}^+ + G_{i-1}\boldsymbol{u}_{i-1}.$$

Namely, the state estimate propagates the same way that the mean of the state does. This update is based on the system dynamics (1).

Next, we derive the equations for updating the covariance of the state estimation error from time $i-1$ to time i but before the measurement at time i is considered. We begin with P_0^+, the covariance of our initial estimate $\hat{\boldsymbol{x}}_0^+$. The covariance represents the uncertainty in $\hat{\boldsymbol{x}}_0$:

$$P_0^+ = E\Big((\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_0)(\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_0)^\top\Big) = E\Big((\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_0^+)(\boldsymbol{x}_0 - \bar{\boldsymbol{x}}_0^+)^\top\Big). \tag{16}$$

If the initial state $\boldsymbol{x}_0$ is perfectly known, then $P_0^+ = 0$; if it is completely unknown, then $P_0^+ = \infty I$. The covariance of the state propagates with time according to (3), which is now rewritten as

$$P_i^- = F_{i-1} P_{i-1}^+ F_{i-1}^\top + Q_{i-1}. \tag{17}$$

In particular, we use

$$P_1^- = F_0 P_0^+ F_0^\top + Q_0. \tag{18}$$

for initialization.

4 From *a priori* Estimate to *a posteriori* Estimate

Now we need to update the state estimate at time i by taking into account the measurement $\boldsymbol{y}_i$ at time i. In other words, we compute $\hat{\boldsymbol{x}}_i^+$ from $\hat{\boldsymbol{x}}_i^-$, which was obtained without knowledge of $\boldsymbol{y}_i$. Freeze the time at i temporarily and imagine that a sequence of observations $\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_{l-1}$ have already been made, all at the time instant. Recall from the lecture on recursive least squares that the availability of a new observation $\boldsymbol{y}'_l$ changes the estimate as follows:

$$\begin{aligned} K'_l &= P'_{l-1} {H'_l}^\top (H'_l P'_{l-1} {H'_l}^\top + R'_l)^{-1}, \\ \hat{\boldsymbol{x}}'_l &= \hat{\boldsymbol{x}}'_{l-1} + K'_l(\boldsymbol{y}'_l - H'_l \hat{\boldsymbol{x}}'_{l-1}), \\ P'_l &= (I - K'_l H'_l) P'_{l-1}, \end{aligned}$$

where K', P', H', and R' respectively carry the same meanings of i, P, H, and R with their subscripts indicating the number of observations (hypothetically) made at the same time instant i. In particular, $\hat{\boldsymbol{x}}'_{l-1} = \hat{\boldsymbol{x}}_i^-$ is the estimate before the measurement $\boldsymbol{y}'_l = \boldsymbol{y}_i$ is processed, and $\hat{\boldsymbol{x}}'_l = \hat{\boldsymbol{x}}_i^+$ is the estimate based on $\boldsymbol{y}_i$. The full correspondences between the terms at the lth iteration of the recursive least squares and those at the (current) ith step of estimation are shown below:

$$\begin{array}{lcllcl} \hat{\boldsymbol{x}}'_{l-1} & \leftrightarrow & \hat{\boldsymbol{x}}_i^- & P'_{l-1} & \leftrightarrow & P_i^- \\ \hat{\boldsymbol{x}}'_l & \leftrightarrow & \hat{\boldsymbol{x}}_i^+ & P'_l & \leftrightarrow & P_i^+, \\ \boldsymbol{y}'_l & \leftrightarrow & \boldsymbol{y}_i & K'_l & \leftrightarrow & K_i, \\ H'_l & \leftrightarrow & H_i & R'_l & \leftrightarrow & R_i. \end{array}$$

Applying these replacements, we obtain the measurement update equations:

$$\begin{aligned} K_i &= P_i^- H_i^\top (H_i P_i^- H_i^\top + R_i)^{-1}, \\ \hat{\boldsymbol{x}}_i^+ &= \hat{\boldsymbol{x}}_i^- + K_i(\boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_i^-), \\ P_i^+ &= (I - K_i H_i) P_i^-. \end{aligned}$$

5 Kalman Filter

Now we summarize all the equations into an algorithm referred to as the *discrete-time Kalman filter*. The system dynamics are given by (4) and (5) along with noise statistics (8)–(10).

1. Initialize the Kalman filter according to (15) and (16), namely,

$$\begin{aligned} \hat{\boldsymbol{x}}_0^+ &= E(\boldsymbol{x}_0) = \bar{\boldsymbol{x}}_0, \\ P_0^+ &= E\left((\boldsymbol{x}-\bar{\boldsymbol{x}}_0)(\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0)^\top\right). \end{aligned}$$

2. Iterate the following sequentially at each time step $i = 1, 2, \ldots$:

$$P_i^- = F_{i-1}P_{i-1}^+F_{i-1}^\top + Q_{i-1}, \tag{19}$$

$$\hat{\boldsymbol{x}}_i^- = F_{i-1}\hat{\boldsymbol{x}}_{i-1}^+ + G_{i-1}\boldsymbol{u}_{i-1}; \qquad (a\ \textit{priori}\ \text{state estimate}) \tag{20}$$

$$K_i = P_i^-H_i^\top(H_iP_i^-H_i^\top + R_i)^{-1}, \tag{21}$$

$$\hat{\boldsymbol{x}}_i^+ = \hat{\boldsymbol{x}}_i^- + K_i(\boldsymbol{y}_i - H_i\hat{\boldsymbol{x}}_i^-), \qquad (a\ \textit{posteriori}\ \text{state estimate}) \tag{22}$$

$$P_i^+ = (I-K_iH_i)P_i^-(I-K_iH_i)^\top + K_iR_iK_i^\top \tag{23}$$

$$= (I-K_iH_i)P_i^-. \tag{24}$$

Updates of the state estimate $\hat{\boldsymbol{x}}_i^-$ in (20) and *a priori* covariance P_i^- in (19) are according to the system dynamics described by (1). Those of the gain matrix in (21), the *a posteriori* state estimate $\hat{\boldsymbol{x}}_i^+$ in (22), and the *a posteriori* covariance P_i^+ in (23) and (24)[3] are through recursive least-squares optimization of $E(\|\boldsymbol{x}-\boldsymbol{x}_i^-\|^2)$.

When $\boldsymbol{x}_i$ is a constant, then $F_i = I$, the covariance matrix for noise $Q_i = 0$, and the input $\boldsymbol{u}_i = 0$ for all i. So we have, by (19)–(20), (22), (24),

$$\begin{aligned} P_i^- &= P_{i-1}^+, \\ \hat{\boldsymbol{x}}_i^- &= \hat{\boldsymbol{x}}_{i-1}^+, \\ \hat{\boldsymbol{x}}_i^+ &= \hat{\boldsymbol{x}}_{i-1}^+ + K_i(\boldsymbol{y}_i - H_i\hat{\boldsymbol{x}}_i^-), \\ P_i^+ &= (I-K_iH_i)P_{i-1}^+. \end{aligned}$$

The Kalman filter consists of the last two equations above and (21). It reduces to the recursive least squares algorithm for estimating a constant vector.

Note that the calculation of P_i^-, K_i, and P_i^+ does not depend on the measurement $\boldsymbol{y}_i$. It depends on the system parameters F_i, H_i, Q_i, and R_i. For efficiency, we can precalculate the Kalman gain K_i up to large enough i and store the data beforehand. When the dynamical system starts operating, only $\hat{\boldsymbol{x}}^-$ and $\hat{\boldsymbol{x}}^+$ need to be computed in real time. This also means that the performance of the filter can be investigated and evaluated before it is actually run. The reason is that the accuracy of the filter is indicated by P_i, which does not depend on the measurements.

EXAMPLE 2. Consider a noise-free Newtonian system involving position $\boldsymbol{r}$, velocity $\boldsymbol{v}$, and constant acceleration $\boldsymbol{a}$. Here, $\boldsymbol{r}, \boldsymbol{v}, \boldsymbol{a} \in \mathbb{R}^3$. The system equation is

$$\begin{pmatrix} \dot{\boldsymbol{r}} \\ \dot{\boldsymbol{v}} \\ \dot{\boldsymbol{a}} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{v} \\ \boldsymbol{a} \end{pmatrix},$$

[3]The expression (23) for P_i^+ guarantees it to be symmetric positive definite, as long as P_i^- is so, whereas the expression (24) is simpler though it does not guarantee symmetry or positive definiteness.

or rewritten succinctly as

$$\dot{\boldsymbol{x}} = A\boldsymbol{x}.$$

The system has a solution $\boldsymbol{x} = e^{At}\boldsymbol{x}_0$, where

$$e^{At} = \sum_{j=0}^{\infty} \frac{(At)^j}{j!}.$$

Discretize the system with a sample time interval of h: $\boldsymbol{x}_{i+1} = F\boldsymbol{x}_i$, where

$$\begin{aligned} F &= e^{Ah} \\ &= I + Ah + \frac{(Ah)^2}{2!} + \cdots \\ &\approx \begin{pmatrix} 1 & h & h^2/2 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

There is no state noise or output noise. Hence $Q_i = 0$ and $R_i = 0$ for all i. Also, there is no input so $G_i = 0$. The Kalman filter for the system follows from equations (19) and (20):

$$\begin{aligned} P_i^- &= F P_{i-1}^+ F^\top, \\ \hat{\boldsymbol{x}}_i^- &= F \hat{\boldsymbol{x}}_{i-1}^+. \end{aligned} \tag{25}$$

The equations (21)–(24) are not needed since we do not make any measurements between time i^- and time i^+. For the same reason, we see that the covariance of the estimation error increases between time $(i-1)^+$ and time i^-, as indicated in the first equation above.

Now suppose that we measure the position with a variance of σ^2:

$$\begin{aligned} y_i &= H_i \boldsymbol{x}_i + v_i \\ &= (1\ 0\ 0)\boldsymbol{x}_i + v_i, \\ v_i &\sim (0, R_i), \\ R_i &= \sigma^2. \end{aligned}$$

We denote by $P_{i,jk}^-$ the (j,k)-th entry of the prior covariance. Substituting $H_i = (1,0,0)$ and $R_i = \sigma^2$ into (21), we obtain the Kalman gain:

$$K_i = \frac{1}{P_{i,11}^- + \sigma^2} \begin{pmatrix} P_{i,11}^- \\ P_{i,12}^- \\ P_{i,13}^- \end{pmatrix}.$$

Meanwhile, the posterior covariance follows from (24) as

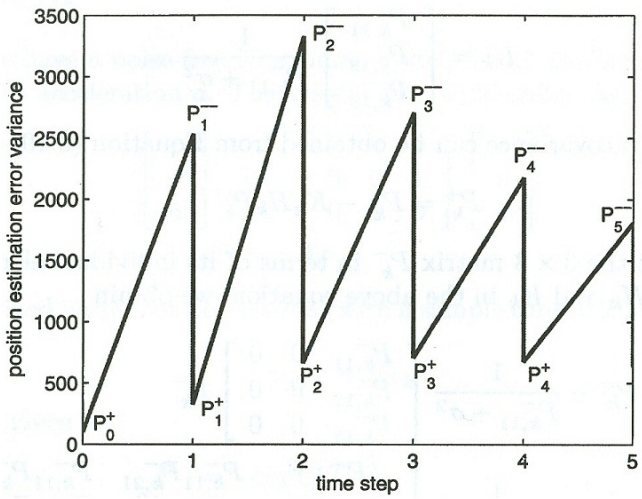
$$P_i^+ = P_i^- - K_i H_i P_i^-. \tag{26}$$

We obtain its traces:

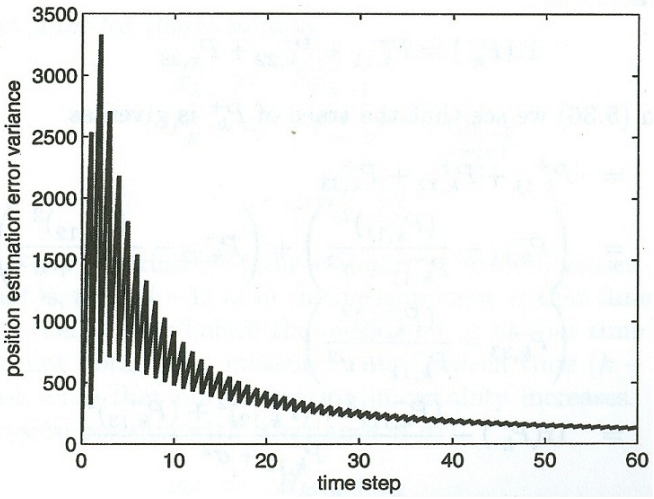
$$\operatorname{tr}(P_i^+) = \operatorname{tr}(P_i^-) - \frac{(P_{i,11}^-)^2 + (P_{i,22}^-)^2 + (P_{i,33}^-)^2}{P_{i,11}^- + \sigma^2}.$$

The above equation shows that the covariance decreases when we get a new measurement. Consequently, our state estimate improves.

The figure below[4] show the variance of the x-position estimate ($P^-_{i,11}$ and $P^+_{i,11}$) for the first five time steps of the Kalman filter. We see that the uncertainty increases from one step to the next, but then decreases as a new measurement is taken into account.



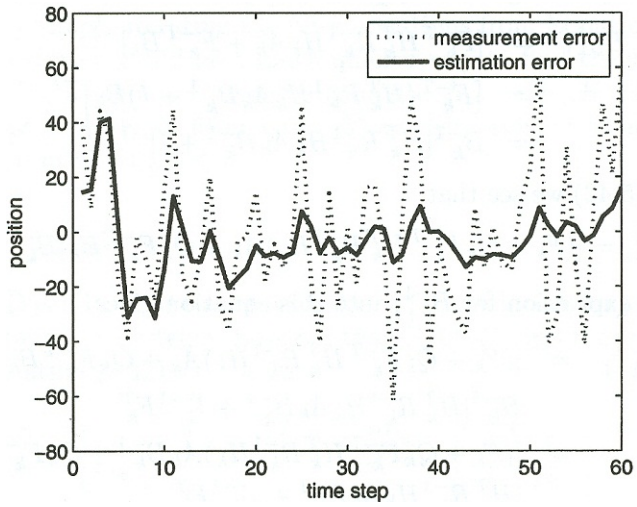
The second figure shows the variance evolution when the Kalman filter runs for the first 60 time steps. It can be seen that, despite the fluctuations before and after each new measurement, the variance converges to a steady-state value.



The next figure[5] shows the position-measurement error (with a standard deviation of 30) and the error of the posterior estimate $\hat{\boldsymbol{x}}^+$. The estimate error starts out with a standard deviation close to 30, and decreases to about 11 by the end of the simulation. The standard deviations are obtained from the covariances.

[4]This and the next figures appear as Figures 5.3 and 5.4 on p. 134 of [1].

[5]Figure 5.5 on p. 135 of [1].



6 One-Step Kalman Filter

This section combines the *a priori* and *a posteriori* Kalman filter equations into one. At time $i+1$, the prior state estimate is, by (20),

$$\hat{\boldsymbol{x}}_{i+1}^{-} = F_i \hat{\boldsymbol{x}}_i^{+} + G_i \boldsymbol{u}_i.$$

Now, we substitute (22) into the above and obtain

$$\begin{aligned} \hat{\boldsymbol{x}}_{i+1}^{-} &= F_i\Big(\hat{\boldsymbol{x}}_i^{-} + K_i(\boldsymbol{y}_i - H_i\hat{\boldsymbol{x}}_i^{-})\Big) + G_i\boldsymbol{u}_i \\ &= F_i(I - K_iH_i)\hat{\boldsymbol{x}}_i^{-} + F_iK_i\boldsymbol{y}_i + G_i\boldsymbol{u}_i. \end{aligned} \tag{27}$$

Similarly, we can obtain a one-step expression for the prior covariance. First, increase the index in (19) by one:

$$P_{i+1}^{-} = F_iP_i^{+}F_i^{\top} + Q_i.$$

Next, substitute (24) into the above:

$$\begin{aligned} P_{i+1}^{-} &= F_i(P_i^{-} - K_iH_iP_i^{-})F_i^{\top} + Q_i \\ &= F_iP_i^{-}F_i^{\top} - F_iK_iH_iP_i^{-}F_i^{\top} + Q_i. \end{aligned} \tag{28}$$

The Kalman filter at each iteration step sequentially executes (21), (27), and (28).

Similarly, we can derive one-step expressions for the *a posterior* state estimate and covariance as follows:

$$\hat{\boldsymbol{x}}_i^{+} = (I - K_iH_i)(F_{i-1}\hat{\boldsymbol{x}}_{i-1}^{+} + G_{i-1}\boldsymbol{u}_{i-1}) + K_i\boldsymbol{y}_i, \tag{29}$$

$$P_i^{+} = (I - K_iH_i)(F_{i-1}P_{i-1}^{+}F_{i-1}^{\top} + Q_{i-1}). \tag{30}$$

The Kalman filter sequentially evaluates (21), (29), and (30) at each iteration step.

7 From Discrete Time to Continuous Time

Almost all Kalman filters are implemented on digital computers and thus inherently discretized. Measurements are also taken at discrete time instants. On the other hand, the vast majority of physical processes evolve continuously in time. It is therefore important to gain some insight into the continuous-time Kalman filter before we can effectively apply the filtering technique to solve many real-world problems.

Let us first understand the relationship between white noise[6] in discrete and continuous time. We start with the following discrete-time system:

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1}, \\ \boldsymbol{x}_0 &= \boldsymbol{0}, \end{aligned} \tag{31}$$

where the noise characteristics are given in (6) and (8) with constant covariance Q. The system has the solution:

$$\boldsymbol{x}_k = \boldsymbol{w}_0 + \boldsymbol{w}_1 + \cdots + \boldsymbol{w}_{k-1}$$

with the covariance calculated as, utilizing (8),

$$E[\boldsymbol{x}_k \boldsymbol{x}_k^\top] = kQ. \tag{32}$$

Meanwhile, the continuous-time parameter is $t = t_k = kT$, where T is the sample time. We see that

$$E[\boldsymbol{x}(t)\boldsymbol{x}^\top(t)] = E[\boldsymbol{x}_k \boldsymbol{x}_k^\top] = kQ.$$

The continuous-time system corresponding to (31) is described by

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{w}(t), \tag{33}$$

where $\boldsymbol{w}(t)$ is white noise. We propose the following covariance for this noise:

$$E[\boldsymbol{w}(t)\boldsymbol{w}^\top(t)] = \frac{Q}{T}\delta(t-\tau). \tag{34}$$

Here, $\delta(t-\tau)$ is the impulse function that has a value of ∞ at $t = \tau$, a value of zero at $t \neq \tau$, and $\int_{-\infty}^{\infty} \delta(t-\tau)\, dt = 1$. Integrating (33) and then substituting the integral into (34), we can derive the covariance as

$$E\left[\boldsymbol{x}(t)\boldsymbol{x}^\top(t)\right] = \frac{t}{T}Q.$$

The above is in agreement with (32) since $k = t/T$. Discrete-time white noise with covariance Q over a sampling interval of T is equivalent to continuous-time white noise with covariance $Q_c\delta(t)$, where $Q_c = Q/T$. The continuous-time zero-mean white noise is

$$\boldsymbol{w}(t) \sim (0, Q_c) \equiv \left(0, \frac{Q}{T}\right). \tag{35}$$

Similarly, to quantify the measurement noise in continuous time, we start with a different discrete-time system:

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{x}_{k-1}, \\ \boldsymbol{y} &= \boldsymbol{x}_k + \boldsymbol{v}_k, \end{aligned}$$

[6] This is the type of noise that has values independent of each other at different time instants.

where the chacteristics of the noise $\boldsymbol{v}_k$ are given in (7) and (9) with constant covariance R. It can be shown that

$$P_k^+ = \frac{P_0 R}{kP_0 + R}.$$

We have

$$\lim_{P_0 \to \infty} P_k^+ = \frac{R}{k} = \frac{RT}{t_k}.$$

This error covariance should be independent of the sample time T, which is possible only if

$$R = \frac{R_c}{T}, \tag{36}$$

for some constant matrix R_c. As $T \to 0$, R becomes an impulse function, namely,

$$\lim_{T \to 0} R = R_c \delta(t).$$

The white measurement noise in continuous time then follows the distribution:

$$v(t) \sim (0, R_c) \equiv (0, RT). \tag{37}$$

In other words,

$$E\left[\boldsymbol{v}(t)\boldsymbol{v}^\top(\tau)\right] = R_c \delta(t - \tau).$$

8 The Continuous-Time Kalman Filter

Consider a continuous-time linear system:

$$\begin{array}{rcl} \dot{\boldsymbol{x}} & = & A\boldsymbol{x} + B\boldsymbol{u} + \boldsymbol{w}, \\ \boldsymbol{y} & = & H\boldsymbol{x} + \boldsymbol{v}, \\ \boldsymbol{w} & \sim & (0, Q_c), \\ \boldsymbol{v} & \sim & (0, R_c), \end{array} \tag{38}$$

where $\boldsymbol{u}(t)$ is some known input. We discretize time with the step size T to approximate (38) as

$$\begin{array}{rcl} \boldsymbol{x}_k & = & e^{AT}\boldsymbol{x}_{k-1} + e^{AT}\left(I - e^{-AT}\right)A^{-1}B\boldsymbol{u}_{k-1} + \boldsymbol{w}_k, \\ \boldsymbol{y}_k & = & H\boldsymbol{x}_k + \boldsymbol{v}_k, \\ \boldsymbol{w}_k & \sim & (0, Q_c T), \\ \boldsymbol{v}_k & \sim & (0, R_c/T). \end{array} \tag{39}$$

We refer to [1, pp. 27] for details on deriving the first equation above. The noise distributions above merely copy (35) and (37). For small enough T, the system dynamics in (39) can be further rewritten as

$$\boldsymbol{x}_k = F\boldsymbol{x}_{k-1} + G\boldsymbol{u}_{k-1} + \Lambda\boldsymbol{w}_{k-1}, \tag{40}$$

where

$$\begin{array}{rcl} F & \approx & I + AT, \\ G & \approx & BT, \\ \Lambda & \approx & IT \end{array}$$

With its dynamics replaced by (40), the system (39) becomes linear. We can apply the one-step Kalman filter (29)–(30). Next, we derive the change rate of the covariance P as

$$\begin{aligned}\dot{P}(t) &= \lim_{T\to 0} \frac{P_{k+1}^- - P_k^-}{T} \\ &= -PH^\top R_c^{-1} HP + AP + PA^\top + Q_c. \end{aligned} \tag{41}$$

The above is known as the differential *Riccati equation*. The state estimate changes at the rate of

$$\begin{aligned}\dot{\hat{\boldsymbol{x}}} &= \lim_{T\to 0} \frac{\hat{\boldsymbol{x}}_k^+ - \hat{\boldsymbol{x}}_{k-1}^+}{T} \\ &= A\hat{\boldsymbol{x}} + B\boldsymbol{u} + K(\boldsymbol{y} - H\hat{\boldsymbol{x}}), \end{aligned} \tag{42}$$

where

$$\begin{aligned} K &= \lim_{T\to 0} \frac{K_k}{T} \\ &= PH^\top R_c^{-1}. \end{aligned} \tag{43}$$

In the above, K_k is the gain of the discretized system given in (21), where $R_i = R = R_c/T$.

The equations (43), (42), and (41) form the continuous-time Kalman filter, which is initialized with

$$\begin{aligned}\hat{\boldsymbol{x}}(0) &= E[\boldsymbol{x}(0)], \\ P(0) &= E\left[\Big(\boldsymbol{x}(0) - \hat{\boldsymbol{x}}(0)\Big)\Big(\boldsymbol{x}(0) - \hat{\boldsymbol{x}}(0)\Big)^\top\right]. \end{aligned}$$

EXAMPLE 3. Suppose we are able to measure the velocity of an object moving in one dimension under random accelerations. The system and measurement equations are given below:

$$\begin{aligned}\dot{x} &= w, \\ y &= x + v, \\ w &\sim (0, Q_c), \\ v &\sim (0, R_c). \end{aligned}$$

Comparing with (38), we see that $A = 0$, $B = 0$, and $H = 1$. This reduces (41) to

$$\dot{P} = Q_c - \frac{P^2}{R_c}.$$

We integrate the above equation under the initial condition $P(0) = P_0$:

$$\int_{P(0)}^{P(t)} \frac{dP}{Q_c - P^2/R_c} = \int_0^t d\tau,$$

yielding

$$\frac{1}{2\sqrt{Q_c}} \ln\left(\frac{\sqrt{Q_c} + P/\sqrt{R_c}}{\sqrt{Q_c} - P/\sqrt{R_c}}\right)\Bigg|_{P(0)}^{P(t)} = t.$$

Solve the above equation for P and take its limit:

$$\lim_{t\to\infty} P = \sqrt{Q_c R_c}.$$

The Kalman gain follows from (43):

$$K = \frac{P}{R_c},$$

which has the limit

$$\lim_{t\to\infty} K = \sqrt{\frac{Q_c}{R_c}}.$$

The state estimate update is simplified from (42) to

$$\dot{\hat{x}} = K(y - \hat{x}).$$

It is intuitive that K increases if process noise Q_c increases. In this case, we change $\hat{x}$ more agressively. On the other hand, K decreases if measurement noise R_c increases. In this case, we have less confidence in the measurement y and thus change $\hat{x}$ more conservatively.

References

[1] D. Simon. *Optimal State Estimations*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[2] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. The MIT Press, Boston, MA, 2005.