CDS 270-2: Lecture 4-1
Kalman Filtering

Henrik Sandberg
17 April 2006

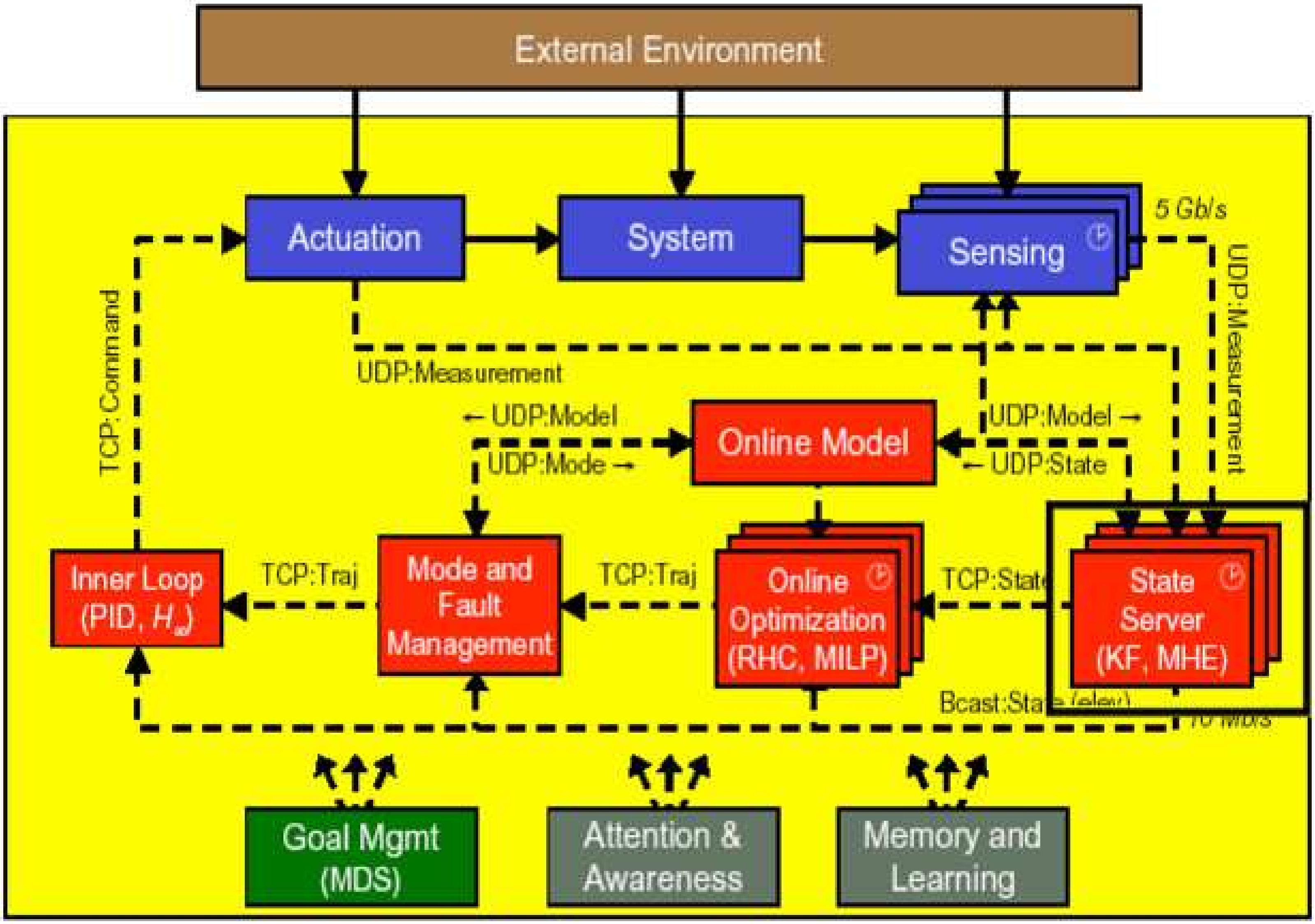
Goals:

- To understand the properties and structure of the Kalman filter.
- To derive the Kalman filter for a special case.

Reading:

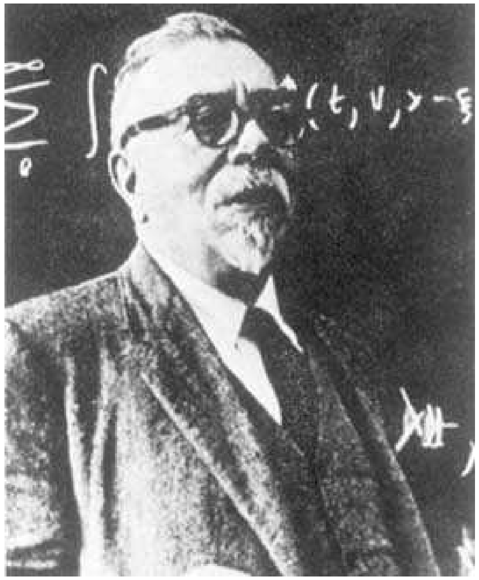
- G. Welch and G. Bishop: An Introduction to the Kalman Filter
  `http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf`
- Wikipedia: Kalman Filter
- (R.E. Kalman: "A New Approach to Linear Filtering and Prediction Problems", Transactions of the ASME, 1960
  `http://www.cs.unc.edu/~welch/kalman/kalmanPaper.html`)

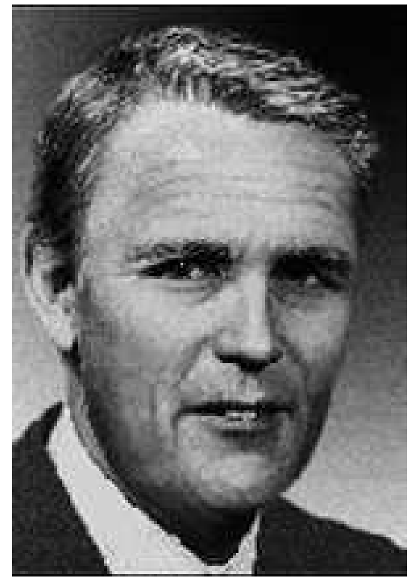
Networked Control Systems



Today: The state server with Kalman filter.

Some History





Norbert Wiener: "Father of cybernetics". Filtering, prediction, and smoothing using integral equations. Spectral factorizations. "Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications", MIT Press 1949. (Also known as "The yellow peril".)

Rudolf E. Kalman: Filtering, prediction, and smoothing using state-space formulas. Riccati equations.

Some Probability Theory

The filters are derived in a stochastic setting.

The stochastic variable x has a probability density function $p_x(x)$ such that

$$\mathbb{P}\{a \leq x \leq b\} = \int_a^b p_x(x)dx$$
$$\mathbb{E}\{x\} = \int_{-\infty}^{\infty} x \cdot p_x(x)dx.$$

If x is jointly distributed with y with conditional distribution $p_{x|y} = p_{x,y}/p_y$, then we define the conditional expectation as

$$\mathbb{E}\{x|y\} = \int_{-\infty}^{\infty} x \cdot p_{x|y}(x|y)dx.$$

Problem Formulation (Kalman)

The model:

$$x_{k+1} = A_k x_k + B_k u_k + N_k w_k$$
$$y_k = C_k x_k + D_k u_k + v_k,$$

with zero-mean stochastic process noise w_k and measurement noise v_k with covariances

$$\mathbb{E}\left\{ \begin{bmatrix} w_k \\ v_k \end{bmatrix} \begin{bmatrix} w_l^T & v_l^T \end{bmatrix} \right\} = \begin{bmatrix} Q_k & S_k \\ S_k^T & R_k \end{bmatrix} \delta_{kl} = \Sigma_k \delta_{kl}.$$

The problem: *Given* the data

$$Y_k = \{y_i, u_i : 0 \le i \le k\},$$

find the "best" (to be defined) estimate $\hat{x}_{k+m}$ of x_{k+m}.
($m = 0$ filtering, $m > 0$ prediction, and $m < 0$ smoothing.)

The Solution (Kalman)

- Assume that we know the mean and the covariance of the initial state x_0:

$$\mathbb{E}\{x_0\} = \bar{x}_0, \quad \mathbb{E}\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T\} = P_0.$$

- Denote the filter states by
  - $\hat{x}_{k|k}$ — estimate of x_k given Y_k
  - $\hat{x}_{k+1|k}$ — estimate of x_{k+1} given Y_k

The Solution (Kalman)

- Assume that we know the mean and the covariance of the initial state x_0:

$$\mathbb{E}\{x_0\} = \bar{x}_0, \quad \mathbb{E}\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T\} = P_0.$$

- Denote the filter states by
  - $\hat{x}_{k|k}$ — estimate of x_k given Y_k
  - $\hat{x}_{k+1|k}$ — estimate of x_{k+1} given Y_k

Step 0. (Initialization)

Put

$$\hat{x}_{0|-1} = \bar{x}_0$$
$$P_{0|-1} = P_0$$

The Solution (cont'd)

Step 1. (Corrector — Use the most recent measurement.)

$$
\begin{aligned}
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k e_k \\
e_k &= y_k - C_k \hat{x}_{k|k-1} - D_k u_k \quad \text{(prediction error)} \\
K_k &= P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + R_k)^{-1} \\
P_{k|k} &= P_{k|k-1} - K_k C_k P_{k|k-1}
\end{aligned}
$$

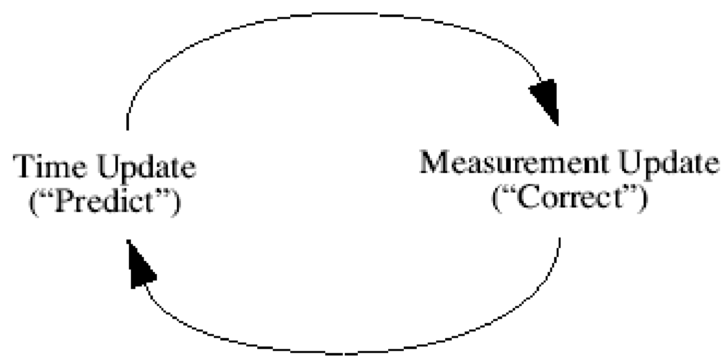
Step 2. (One-step predictor — $S_k = 0$.)

$$
\begin{aligned}
\hat{x}_{k+1|k} &= A_k \hat{x}_{k|k} + B_k u_k \\
P_{k+1|k} &= A_k P_{k|k} A_k^T + N_k Q_k N_k^T
\end{aligned}
$$

(More complicated formulas for $S_k \neq 0$. See lecture notes.)

Iterate Steps 1 and 2 and increase k.

Comments on Kalman Filters



- $P_{k|k}$ and $P_{k|k-1}$ are the covariance of the estimation error:

$$P_{k|k} = \mathbb{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T\}$$
$$P_{k|k-1} = \mathbb{E}\{(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T\}.$$

and are a measure of the uncertainty of the estimate.

- The Kalman filter gives an unbiased estimate, i.e.,

$$\mathbb{E}\{\hat{x}_{k|k}\} = \mathbb{E}\{\hat{x}_{k|k-1}\} = \mathbb{E}\{x_k\}.$$

- *Algebraic Riccati equations* in stationarity. (Exercise.)

Example

Estimate the constant scalar state x.

No process noise, but measurement noise:

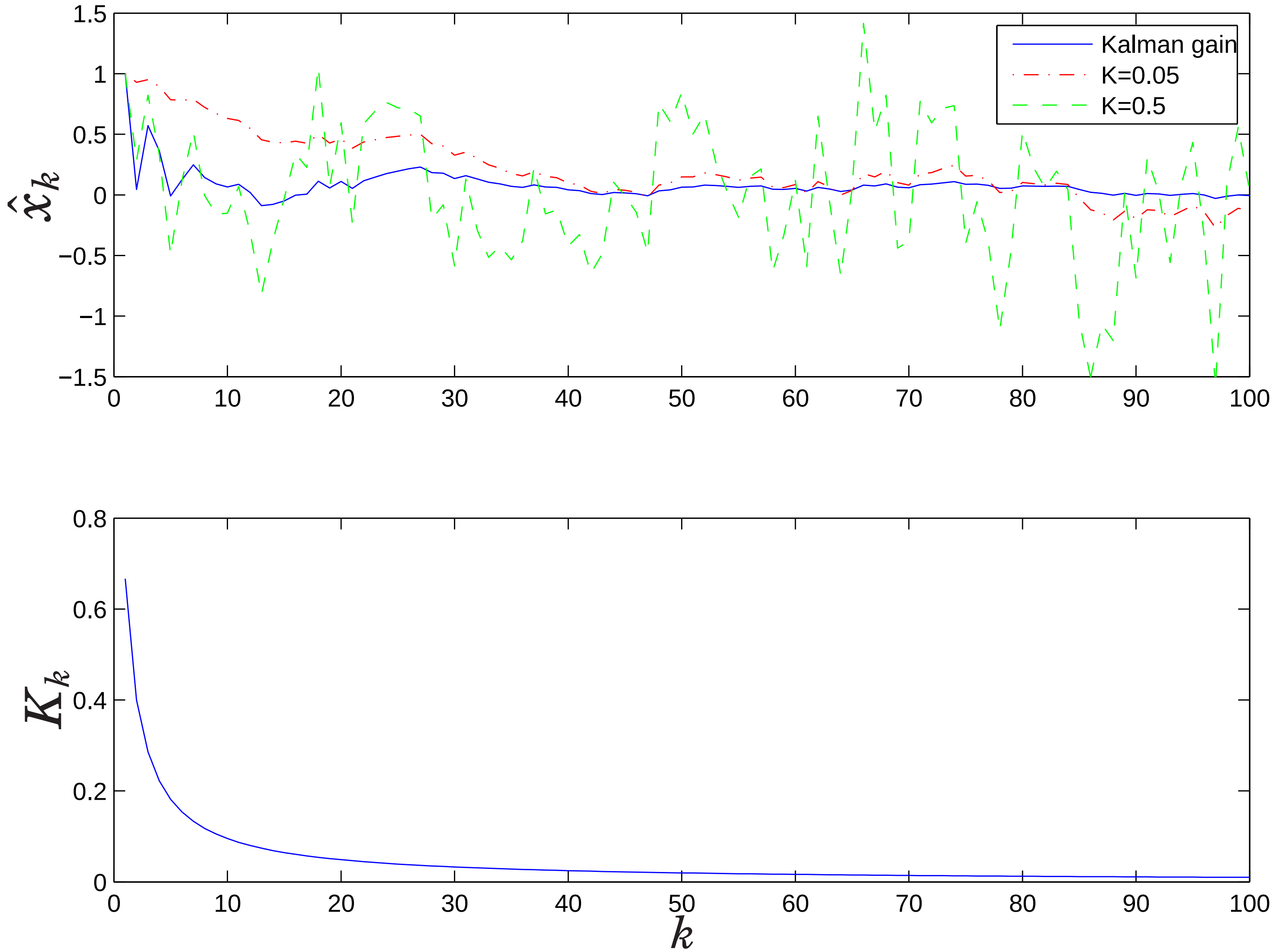
$$x_{k+1} = x_k, \quad x_0 = x = 0,$$
$$y_k = x_k + v_k, \quad \mathbb{E}\{v_k^2\} = 1.$$

Compare Kalman filter to a Luenberger-type observer

$$\hat{x}_{k+1} = \hat{x}_k + K(y_k - \hat{x}_k),$$

for some different values of K.

Example (cont'd)



To Think About

1. What was $\bar{x}_0$ here?
2. What is the problem with a large (small) K in the observer?
3. What does the Kalman filter do?
4. How would you change P_0 in the Kalman filter to get a smoother (but slower) transient in $\hat{x}_k$?
5. In practice, Q_k, R_k, and P_0 are often tuning parameters. What are their influence on the estimate?

Sensor Fusion

- Assume there are p sensors. Then the Kalman filter weights the measurements and fuses the data.
- Assume a diagonal R_k, then the gain K_k is given by

$$K_k = P_{k|k-1} C_k^T \left(C_k P_{k|k-1} C_k^T + \begin{bmatrix} R_{k,11} & & \\ & \ddots & \\ & & R_{k,pp} \end{bmatrix} \right)^{-1}.$$

- A large $R_{k,ii}$ (much measurement noise) leads to low influence on estimate.
- The covariance of estimation error is updated as

$$P_{k+1|k} = A_k P_{k|k-1} A_k^T + N_k Q_k N_k^T - A_k K_k C_k P_{k|k-1} A_k^T.$$

First two terms in the r.h.s. represent natural evolution of uncertainty. The last term shows how much uncertainty the Kalman filter removes.

Optimality 1— Gaussian Case

THEOREM 1
Assume that the white noise is Gaussian and uncorrelated with x_0, which is also Gaussian:

$$\begin{bmatrix} w_k \\ v_k \end{bmatrix} \in \mathcal{N}(0, \Sigma_k), \quad x_0 \in \mathcal{N}(\bar{x}_0, P_0).$$

Then the Kalman filter gives the minimum-variance estimate of x_k. That is, the covariances $P_{k|k}$ and $P_{k|k-1}$ are the smallest possible. We also have that the estimates are the conditional expectations

$$\hat{x}_{k|k} = \mathbb{E}\{x_k | Y_k\}$$
$$\hat{x}_{k|k-1} = \mathbb{E}\{x_k | Y_{k-1}\}.$$

□

Optimality 2— Non-Gaussian Case

THEOREM 2
Assume that the noise is uncorrelated with x_0 (and white as before). Then the Kalman filter is the optimal *linear* estimator in the sense that no other *linear* filter gives a smaller variance of the estimation error. □

Proof. See Anderson and Moore, "Optimal Filtering", Dover.

- In the non-Gaussian case, nonlinear filters can do a much better job. Use, for example, moving-horizon estimators (next time) or particle filters.

Next: We prove Theorem 1.

Covariance Inequalites

- The Kalman filter minimizes the quantity

$$\mathbb{E}\{(x_k - \hat{x}_{k|k})^T (x_k - \hat{x}_{k|k})\} = \mathbf{trace}\, \mathbb{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T\}$$
$$= \mathbf{trace}\, P_{k|k}.$$

This implies that the error covariance $\tilde{P}_k$ at time k for any other filter (Theorem 1) obeys

$$\mathbf{trace}\, P_{k|k} \leq \mathbf{trace}\, \tilde{P}_k$$

and that $\tilde{P}_k - P_{k|k}$ is positive semidefinite.

- The same is true for any positive semidefinite weight matrix W:

$$\mathbb{E}\{(x_k - \hat{x}_{k|k})^T W (x_k - \hat{x}_{k|k})\} = \mathbf{trace}\,(W P_{k|k}) \leq \mathbf{trace}\,(W \tilde{P}_k).$$

Three Lemmas from Probability Theory (1)

LEMMA 1
Assume that the stochastic variables x and y are jointly distributed. Then the minimum-variance estimate $\hat{x}$ of x, given y, is the conditional expectation

$$\hat{x} = \mathbb{E}\{x|y\}.$$

That is

$$\mathbb{E}\{\|x - \hat{x}\|^2|y\} \leq \mathbb{E}\{\|x - f(y)\|^2|y\}$$

for any other estimate $f(y)$. $\square$

Three Lemmas from Probability Theory (2)

LEMMA 2

Assume that x and y have a joint Gaussian distribution with mean and covariance

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

Then the stochastic variable x, conditioned on the information y, is Gaussian with mean and covariance

$$\bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y}) \quad \text{and} \quad \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}.$$

That is,

$$\mathbb{E}\{x|y\} = \bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y}).$$

□

Three Lemmas from Probability Theory (3)

LEMMA 3

Assume that

$$x_{k+1} = A_k x_k + N_k w_k$$

and

$$\mathbb{E}\{x_k\} = \bar{x}_k, \quad \mathbb{E}\{(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T\} = P_k,$$

and

$$\mathbb{E}\{w_k\} = 0, \quad \mathbb{E}\{w_k w_k^T\} = Q_k, \quad \mathbb{E}\{w_k x_k^T\} = 0.$$

Then

$$\bar{x}_{k+1} = \mathbb{E}\{x_{k+1}\} = A_k \bar{x}_k$$

$$P_{k+1} = \mathbb{E}\{(x_{k+1} - \bar{x}_{k+1})(x_{k+1} - \bar{x}_{k+1})^T\} = A_k P_k A_k^T + N_k Q_k N_k^T$$

$$P_{k+1,k} = \mathbb{E}\{(x_{k+1} - \bar{x}_{k+1})(x_k - \bar{x}_k)^T\} = A_k P_k.$$

□

Proof of Theorem 1

Assume $S_k = 0$ and $u_k = 0$. Iteratively update the mean and the covariance of the state and the measurement signal.

1. ("Correct") The stochastic variable $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ is Gaussian with mean and covariance

$$\begin{bmatrix} \bar{x}_0 \\ C_0\bar{x}_0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_0 & P_0C_0^T \\ C_0P_0 & C_0P_0C_0^T + R_0 \end{bmatrix}.$$

   Hence, x_0 conditioned on y_0 gives the minimum-variance estimate

$$\hat{x}_{0|0} = \mathbb{E}\{x_0|Y_0\} = \bar{x}_0 + P_0C_0^T(C_0P_0C_0^T + R_0)^{-1}(y_0 - C_0\bar{x}_0)$$
$$P_{0|0} = P_0 - P_0C_0^T(C_0P_0C_0^T + R_0)^{-1}C_0P_0.$$

Proof of Theorem 1 (cont'd)

2. ("One-step predictor for state") x_1 cond. on y_0 (use that $S_k = 0$) is a Gaussian with mean and covariance

$$\hat{x}_{1|0} = \mathbb{E}\{x_1 | Y_0\} = A_0 \hat{x}_{0|0}$$

$$P_{1|0} = A_0 P_{0|0} A_0^T + N_0 Q_0 N_0^T.$$

3. ("One-step predictor for output") y_1 cond. on y_0 is Gaussian with mean and covariance

$$\hat{y}_{1|0} = C_1 \hat{x}_{1|0} \quad \text{and} \quad C_1 P_{1|0} C_1^T + R_1$$

and

$$\mathbb{E}\{(y_1 - \hat{y}_{1|0})(x_1 - \hat{x}_{1|0})^T\} = C_1 P_{1|0}.$$

Proof of Theorem 1 (cont'd)

4. ("Collect the elements") The stochastic variable $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ conditioned on y_0 is Gaussian and has mean and covariance

$$\begin{bmatrix} \hat{x}_{1|0} \\ C_1\hat{x}_{1|0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_{1|0} & P_{1|0}C_1^T \\ C_1P_{1|0} & C_1P_{1|0}C_1^T + R_1 \end{bmatrix}.$$

5. ("Correct") Goto 1 with obvious changes of time indices.

More elegant (geometric) and powerful proofs exist. Require more advanced mathematics. See, for example, Kalman's paper.

Summary

- The Kalman filter is the optimal filter for a linear model subject to Gaussian noise.
- The Kalman filter is in state-space form and is recursive: predict, correct, predict,...
- The Kalman filter fuses measurement data.
- The Kalman filter is easy to implement. Compare with a Wiener filter that needs integral equations and stationarity.
- Can be derived by using conditional expectations.

Exercises

1. Implement a Kalman filter in Matlab and gain experience on the influence of Q_k, R_k, S_k, and P_0 on the estimate.
2. Assume the model is time invariant ($A_k = A$, $B_k = B$, $C_k = C$, and $D_k = D$). Derive the stationary Kalman filter and the algebraic Riccati equation. Find (in the literature) the assumptions under which the algebraic Riccati equation has a useful solution.
3. Derive the optimal m-step predictor $\hat{x}_{k+m|k}$ when $m > 1$.
4. Assume $S_k \neq 0$. Use Lemma 2 to derive the best estimate of w_k, given v_k. Compare the result with the Kalman filter for the case $S_k \neq 0$.