M0000.026500

학습기반 공정 동적최적화

Lecture 5: State Estimation

JONG MIN LEE

School of Chemical & Biological Engineering

Why State Estimation?

- The state is usually not fully measurable.
- Even the partial measurement of the state is usually corrupted by noise.

Linear State Estimation

$$x(k+1) = Ax(k) + Bu(k) + \omega(k)$$
$$y(k) = Cx(k) + \nu(k)$$

$\{w(k)\}$ and $\{\nu(k)\}$ are stationary random processes with known statistical properties.

Objective: To find the conditional probability density function of the state (or) posterior density of the state

$$p\left[x(k)|Y^k\right]$$

Y^k : Set of all the available measurements up to time instant k

Nonlinear State Estimation

$$x(k+1) = F\left[x(k), u(k)\right] + w(k)$$
$$y(k) = H(x(k)) + \nu(k)$$

$\{w(k)\}$ and $\{\nu(k)\}$ are stationary random processes with known statistical properties.

Objective: To find the conditional probability density function of the state (or) posterior density of the state

$$p\left[x(k)|Y^k\right]$$

Y^k : Set of all the available measurements up to time instant k

Typical Problem Setup of (Linear) State Estimation

$$x(k+1) = Ax(k) + Bu(k) + \omega(k)$$
$$y(k) = Cx(k) + \nu(k)$$

usually, $\omega(k),\ \nu(k)$: zero-mean white noise sequences with

$$\left\{ \begin{bmatrix} \omega(k) \\ \nu(k) \end{bmatrix} \begin{bmatrix} \omega(k) \\ \nu(k) \end{bmatrix}^T \right\} = \begin{bmatrix} R_\omega & R_{\omega,\nu} \\ R_{\nu,\omega} & R_\nu \end{bmatrix}$$

Given $\{y(j),\ u(j),\ j \le k\}$ estimate $x(k+i)$ for $k \ge 0$

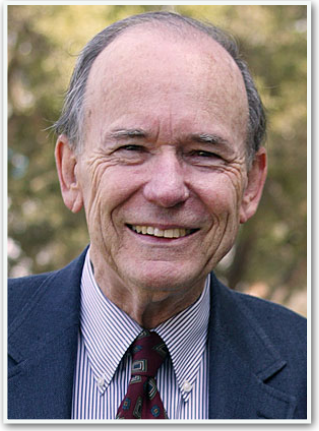
_____: Estimate $x(\,k+i\,)$ for $i = 0$

_____: Estimate $x(\,k+i\,)$ for $i > 0$

_____: Estimate $x(\,k+i\,)$ for $i < 0$

State Observer - Deterministic

"Luenberger Observer"



David. G. Luenberger (1937~)
from stanford.edu

For the deterministic system (i.e., the standard system with Rw, $Rv = 0$ uniformly) with an unknown initial condition, design an observer (of a particular structure) so that $\hat{x}(k) \rightarrow x(k)$ fast.

Linear Observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L(y(k) - \hat{y}(k))$$

$$\hat{y}(k) = C\hat{x}(k)$$

the design parameter (matrix)

Observer Stability

$\hat{x}(k) \rightarrow x(k)$ as $k \rightarrow \infty$ regardless of $x(0) - \hat{x}(0)$

Define $x_e(k) \triangleq x(k) - \hat{x}(k)$

$$x_e(k+1) = (A - LC)x_e(k)$$

All the _____ of (*A* - *LC*) need to lie strictly inside the unit disk for the observer stability

Pole Placement Design of State Observer

- If (C, A) is an observable pair, the eigenvalues of (A - LC) can be placed at arbitrary locations. The choice of L giving a particular set of eigenvalues can be found by putting the state-space system in the observer canonical form.

- Problems with pole placement design:
  - Stability is not the sole objective. Performance also matters.
  - It is not clear where the observer pole should be placed for good performance in general.
  - Systematic trade off between convergence speed and the sensitivity to noise: Fast convergence usually means higher sensitivity to noise.
  - Pole locations alone do not completely determine the response of x_e to errors introduced by a disturbance (외란에 따라 바뀜)

State Estimator - Stochastic Setting

Design with a well-defined objective (e.g., min of error var.)

The Apollo computer used 2k of magnetic core RAM and 36k wire rope [...]. The CPU was built from ICs [...]. Clock speed was under 100 kHz [...]. The fact that the MIT engineers were able to pack such good software (one of the very first applications of the Kalman filter) into such a tiny computer is truly remarkable.

— Interview with Jack Crenshaw, by Matthew Reed, TRS-80.org (2009)

R. E. Kalman (1930-2016)

Photo taken in 2008 (from zimbio.com)

R. E. KALMAN
Research Institute for Advanced Study,[2]
Baltimore, Md.

A New Approach to Linear Filtering and Prediction Problems[1]

The classical filtering and prediction problem is re-examined using the Bode-Shannon representation of random processes and the "state transition" method of analysis of dynamic systems. New results are:

(1) The formulation and methods of solution of the problem apply without modification to stationary and nonstationary statistics and to growing-memory and infinite-memory filters.

(2) A nonlinear difference (or differential) equation is derived for the covariance matrix of the optimal estimation error. From the solution of this equation the coefficients of the difference (or differential) equation of the optimal linear filter are obtained without further calculations.

(3) The filtering problem is shown to be the dual of the noise-free regulator problem.

The new method developed here is applied to two well-known problems, confirming and extending earlier results.

The discussion is largely self-contained and proceeds from first principles; basic concepts of the theory of random processes are reviewed in the Appendix.

Introduction

AN IMPORTANT class of theoretical and practical problems in communication and control is of a statistical nature. Such problems are: (i) Prediction of random signals; (ii) separation of random signals from random noise; (iii) detection of signals of known form (pulses, sinusoids) in the presence of random noise.

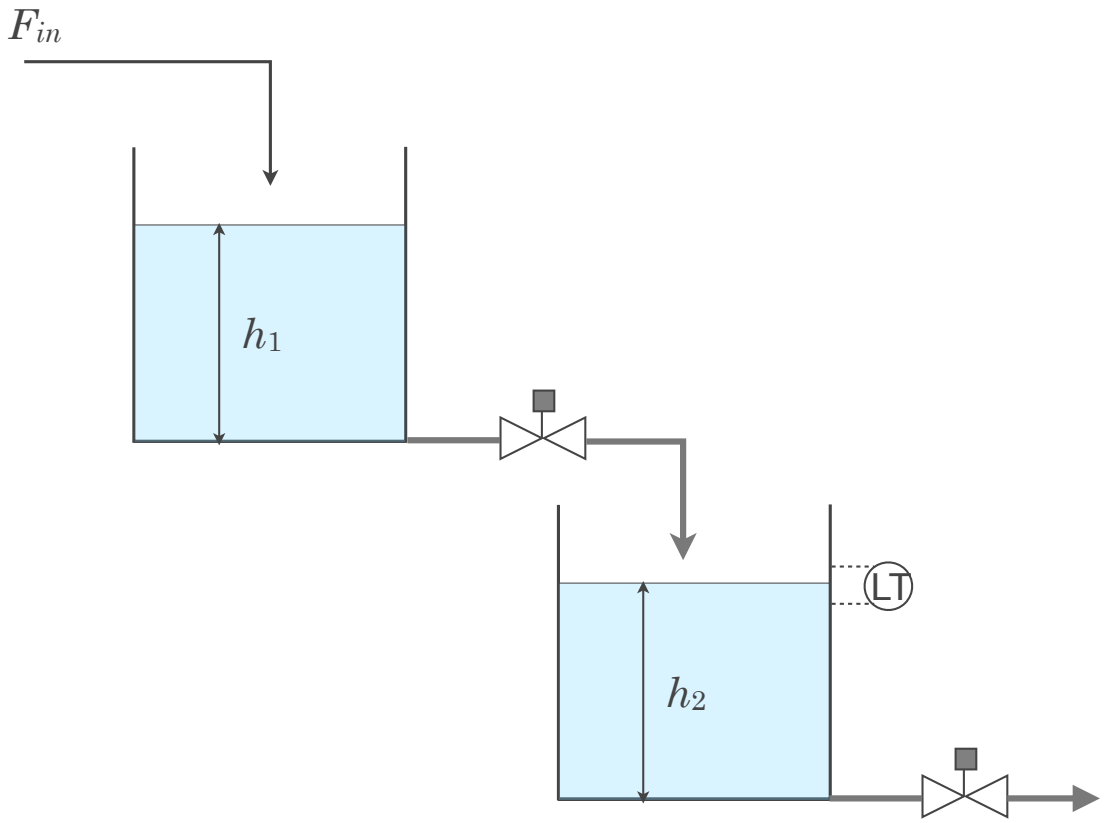
Present methods for solving the Wiener problem are subject to a number of limitations which seriously curtail their practical usefulness:

(1) The optimal filter is specified by its impulse response. It is not a simple task to synthesize the filter from such data.

(2) Numerical determination of the optimal impulse response is often quite involved and poorly suited to machine computation.

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

Why Stochastic Setting?



$$\mathbf{h}(k) = F[\mathbf{h}(k-1), F_{in}(k-1)]$$

$$\mathbf{x}(k) = F[\mathbf{x}(k-1), u(k-1)]$$

measurement eqn:

$$y(k) = \mathbf{C}\mathbf{x}(k)$$

$$\mathbf{C} = [0 \quad 1]$$

$$A\frac{dh_1}{dt} = F_{in} - F_{out,1} = F_{in} - C_{v1}\sqrt{h_1}$$

$$A\frac{dh_2}{dt} = F_{out,1} - F_{out,2} = C_{v1}\sqrt{h_1} - C_{v2}\sqrt{h_2}$$

Linearization and discretization yield:

$$\mathbf{x}'(k+1) = \mathbf{\Phi}\mathbf{x}'(k) + \mathbf{\Gamma}u'(k)$$
$$y'(k) = \mathbf{C}\mathbf{x}'(k)$$

$$\mathbf{\Phi} = \begin{bmatrix} \Phi_{11} & 0 \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \qquad \mathbf{\Gamma} = \begin{bmatrix} \Gamma_1 \\ 0 \end{bmatrix} \qquad \mathbf{C} = [0 \quad 1]$$

$A : 0.25\ [\mathrm{m}^2]$ $\qquad$ $\bar{h}_1 : 0.36\ [\mathrm{m}]$ $\qquad$ $\bar{h}_2 : 0.25\ [\mathrm{m}]$

$\bar{F}_{in} : 0.3\ [\mathrm{m}^3/\mathrm{min}]$ $\qquad$ $C_{v1} : 0.5\ [\mathrm{m}^{2.5}/\mathrm{min}]$ $\qquad$ $C_{v2} : 0.6\ [\mathrm{m}^{2.5}/\mathrm{min}]$

$$\dot{\mathbf{x}}' = \begin{bmatrix} -1.67 & 0 \\ 1.67 & -2.4 \end{bmatrix}\mathbf{x}' + \begin{bmatrix} 4 \\ 0 \end{bmatrix}u' \qquad y' = \begin{bmatrix} 0 & 1 \end{bmatrix}\mathbf{x}'$$

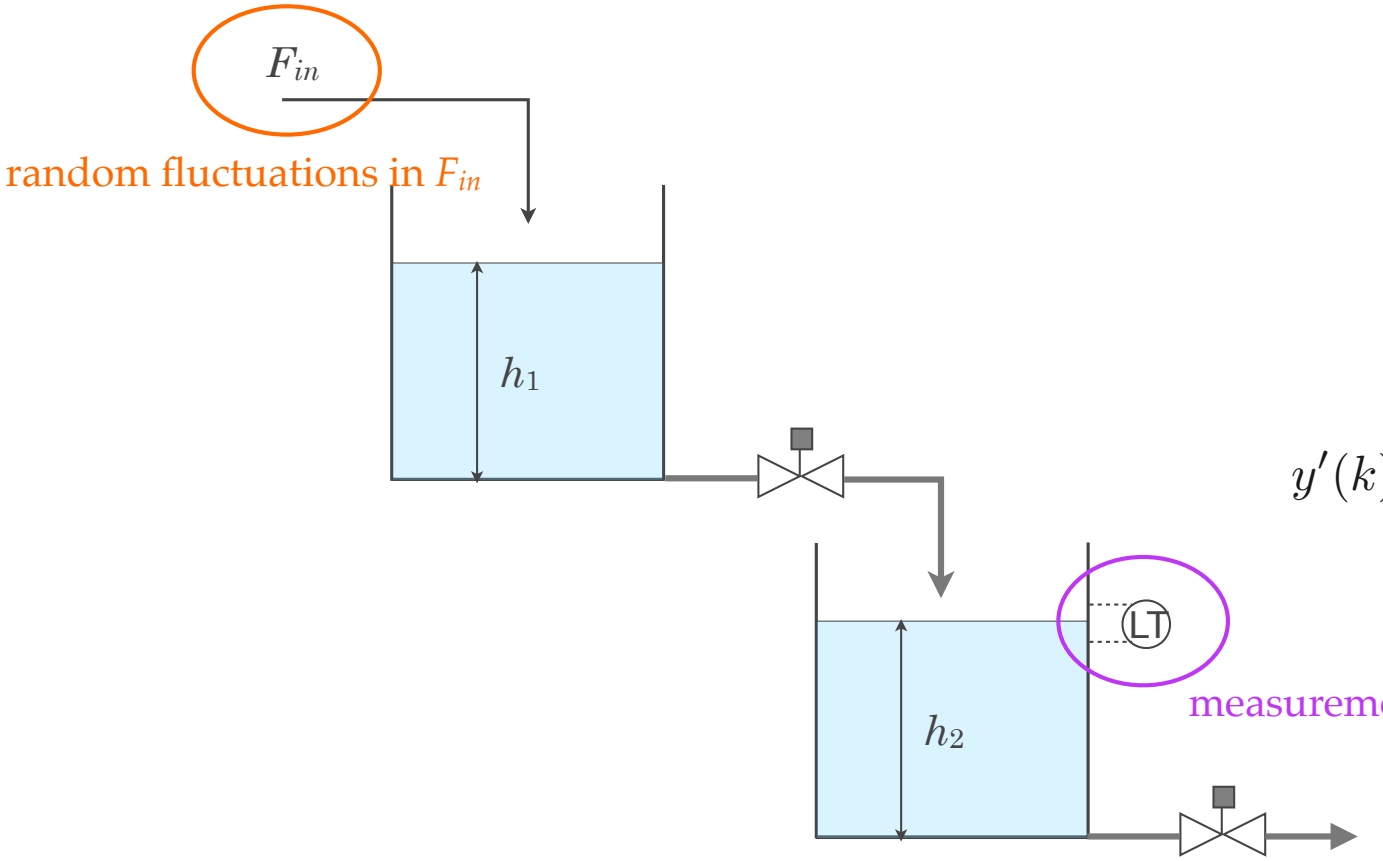
$$\xrightarrow[T_s = 0.1\ [\mathrm{min}]]{}$$

$$\mathbf{x}'(k+1) = \begin{bmatrix} 0.8462 & 0 \\ 0.1363 & 0.7866 \end{bmatrix}\mathbf{x}'(k) + \begin{bmatrix} 0.3684 \\ 0.0292 \end{bmatrix}u'(k) \qquad y'(k) = \begin{bmatrix} 0 & 1 \end{bmatrix}\mathbf{x}'(\mathbf{k})$$

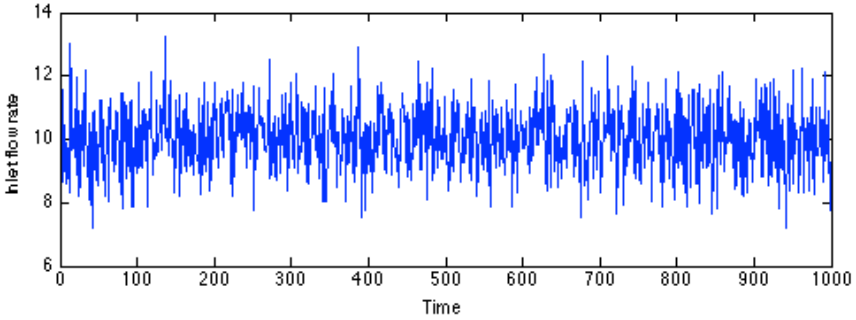
In reality,...

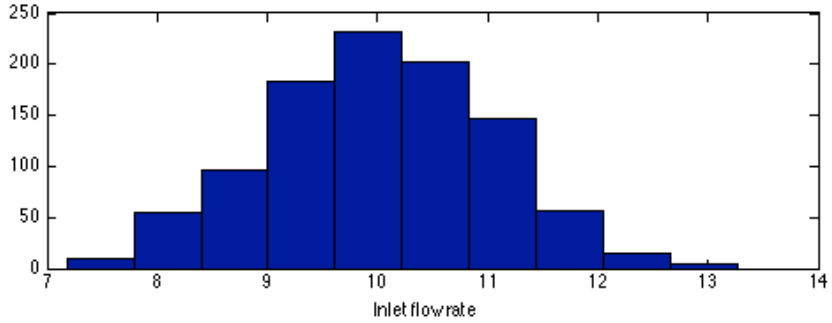
$$F_{in} = F_{in}^{d} + F_{in}^{s}$$

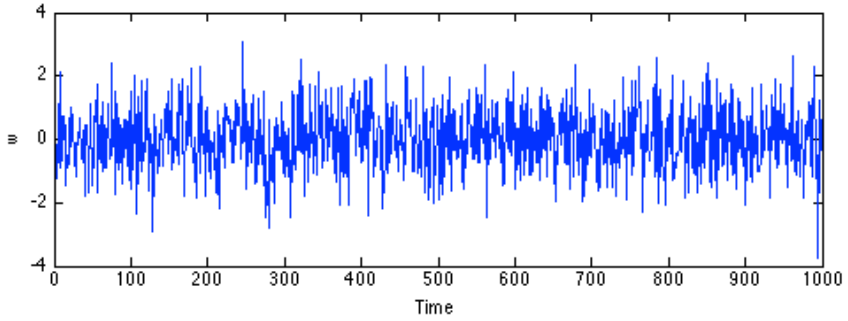
$$\mathbf{x}'(k+1) = \mathbf{\Phi}\mathbf{x}'(k) + \mathbf{\Gamma}u'(k) + \mathbf{\Gamma}w(k)$$

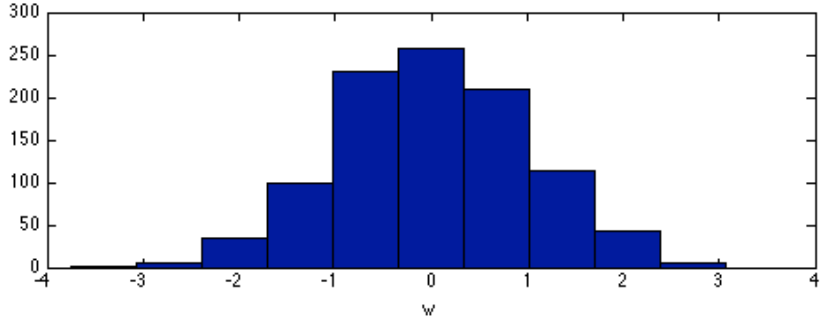


$$y'(k) = \mathbf{C}\mathbf{x}'(k) + \nu(k)$$

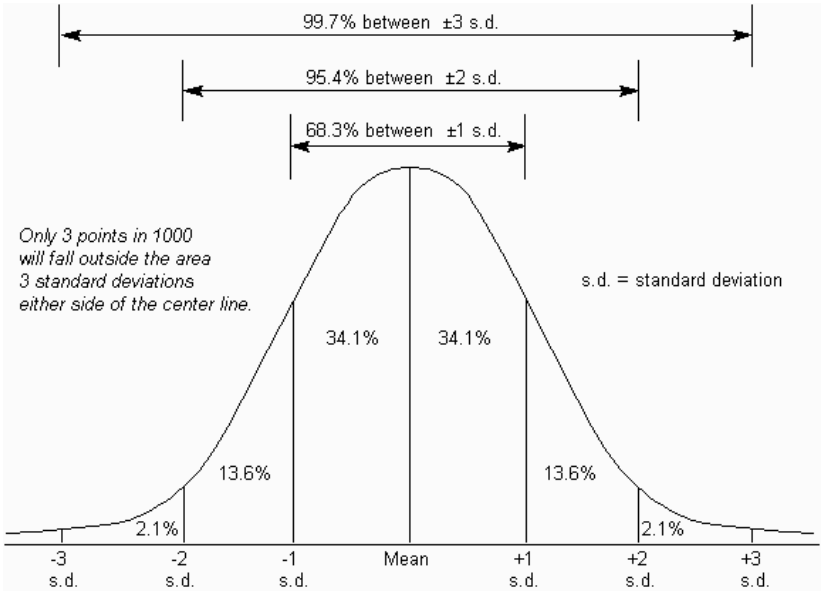




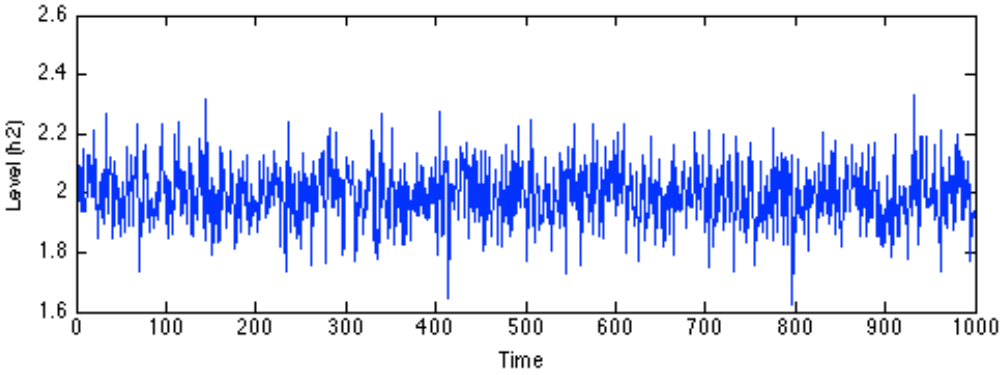


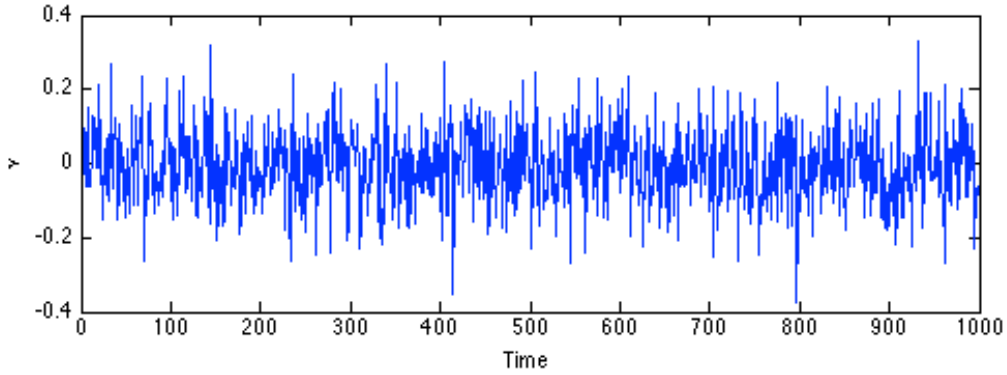


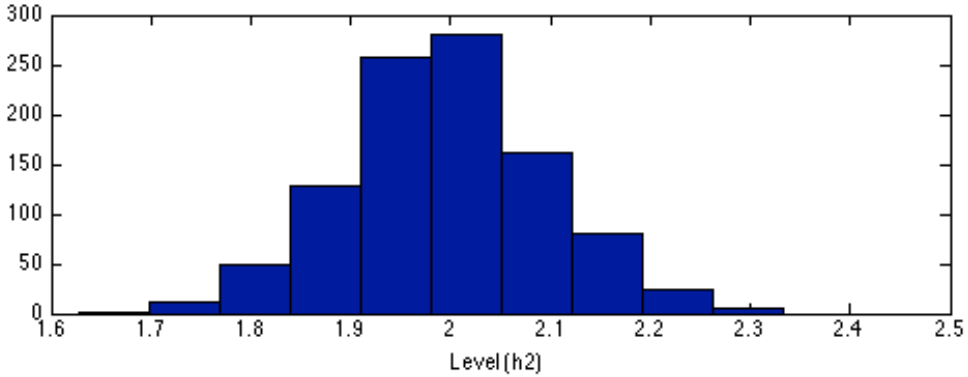
$$W \sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left[\frac{(w-\mu)^2}{2\sigma^2}\right]}$$

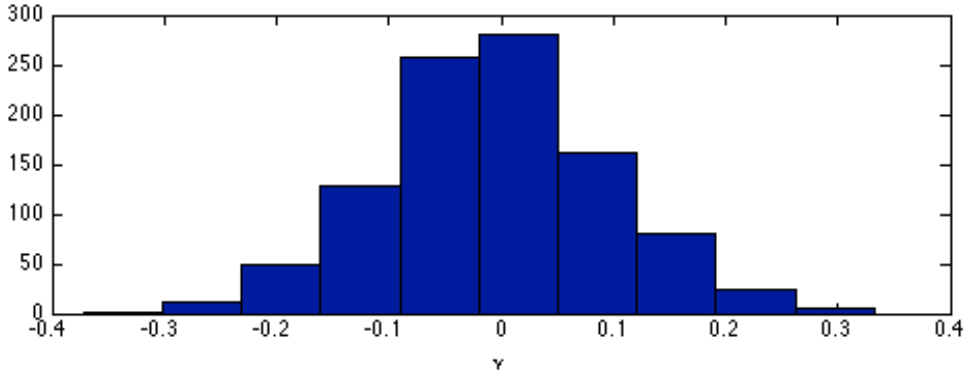


$$V \sim \mathcal{N}(\mu', \sigma'^2) = \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-\left[\frac{(\nu-\mu')^2}{2\sigma'^2}\right]}$$





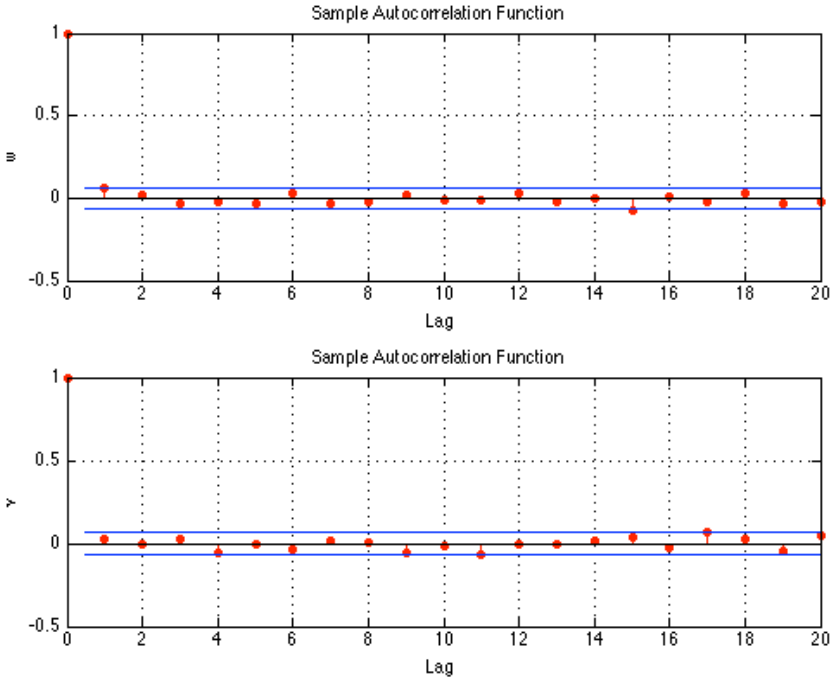


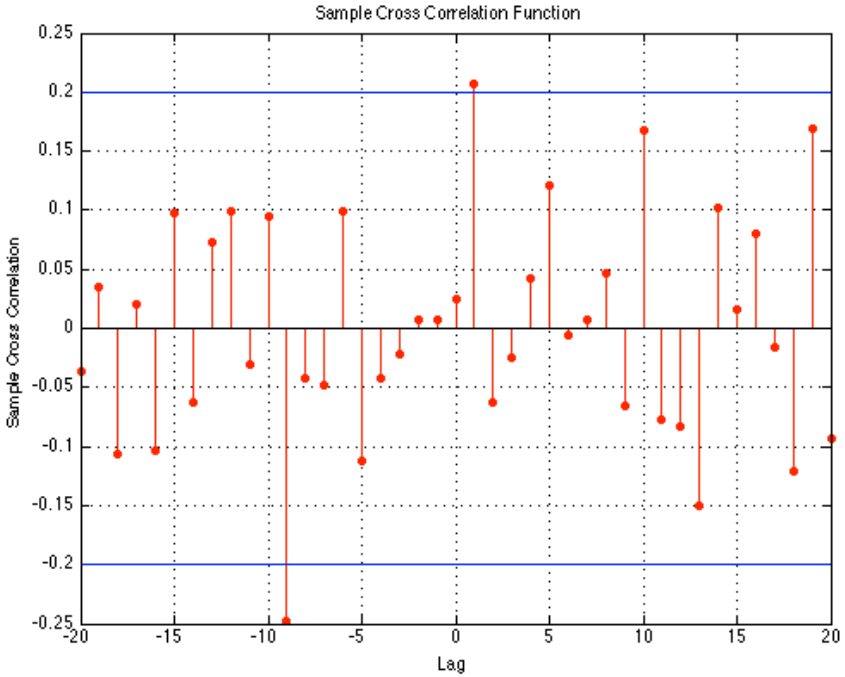


$\{w\}$ and $\{v\}$: white noise sequence with zero-mean Gaussian

$$\mathbb{E}\left[w(k)\right] = 0$$
$$\mathbb{E}\left[\nu(k)\right] = 0$$
$$\mathrm{cov}\left[w(k)\right] = Q$$
$$\mathrm{cov}\left[\nu(k)\right] = R$$

$$\mathrm{cov}\left[w(k), w(j)\right] = 0; j \neq k$$
$$\mathrm{cov}\left[\nu(k), \nu(j)\right] = 0; j \neq k$$
$$\mathrm{cov}\left[w(k), \nu(j)\right] = 0$$





Probability density function of vectors

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \mathcal{N}(\mu, P)$$

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}[z]\right]$$

$$z = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}$$

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \qquad P = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

ρ correlation coefficient

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

Linear transformation

$$\mathbf{Z} = \mathbf{T}\mathbf{X}$$

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathbf{N}(\mu, P) \qquad P = E\Lambda E^T$$

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{T}\mu,\ \mathbf{T}P\mathbf{T}^T)$$

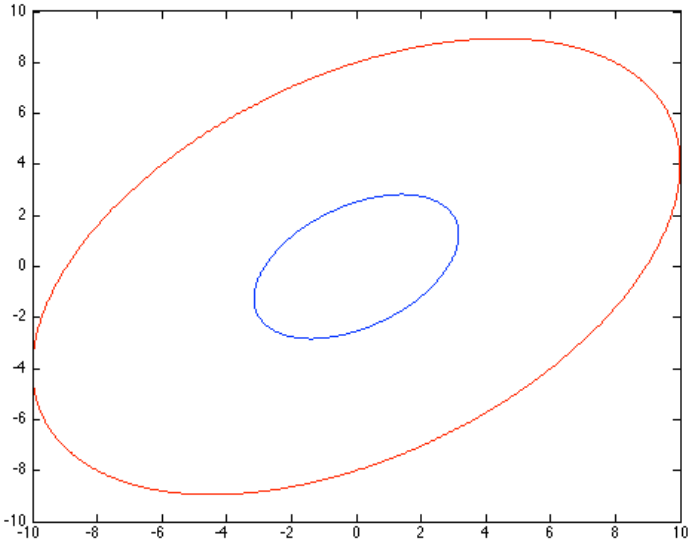
$$P = \begin{bmatrix} 100 & 40 \\ 40 & 80 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 131.21 & 0 \\ 0 & 48.7 \end{bmatrix}$$

$$P = \begin{bmatrix} 10 & 4 \\ 4 & 8 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 13.1 & 0 \\ 0 & 4.9 \end{bmatrix}$$

trace = 18 trace = 18

```
%%
% Drawing covariance ellipse
P1 = [100 40; 40 80];
P2 = [10 4; 4 8];
[E1, L1] = eig(P1);
[E2, L2] = eig(P2);

theta = 2*pi*[0:0.001:1];
for ii = 1:length(theta)
    y(:, ii) = [cos(theta(ii)); sin(theta(ii))];
    x1(:, ii) = E1*sqrt(L1)*y(:, ii);
    x2(:, ii) = E2*sqrt(L2)*y(:, ii);
end
figure
plot(x1(1, :), x1(2, :), 'r', 'LineWidth', 1);
hold on
plot(x2(1, :), x2(2, :), 'b', 'LineWidth', 1);
```



Drawing covariance ellipses

n-dim. multivariate-normal dist $p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_x)^T\mathbf{P}^{-1}(\mathbf{x}-\mu_x)\right\}$

A particularly useful contour is 1-σ bound: $(\mathbf{x}-\mu_x)^T\mathbf{P}^{-1}(\mathbf{x}-\mu_x) = 1$

Assume zero-mean. $\mathbf{x}^T\mathbf{P}^{-1}\mathbf{x} = 1$ $\mathbf{x}^T(E\Lambda E^T)^{-1}\mathbf{x} = 1$

Orthonormality

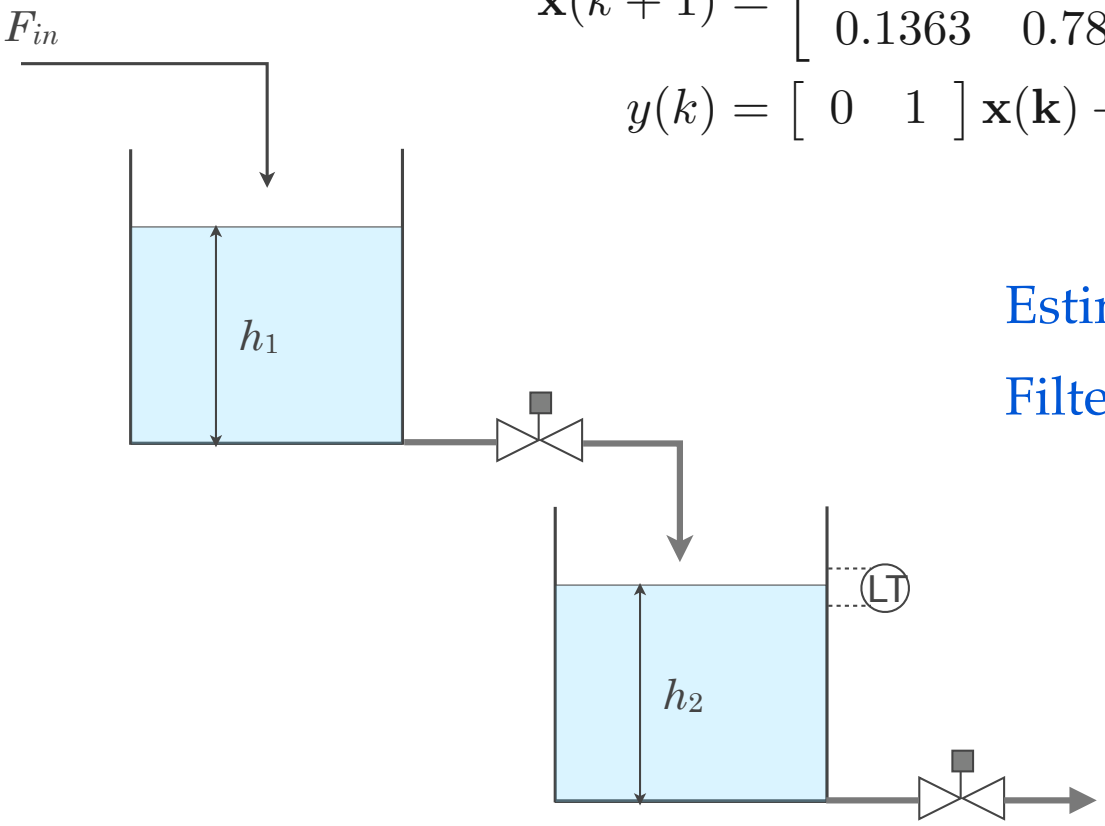
$$\mathbf{x}^T E\Lambda^{-1}E^T\mathbf{x} = 1$$
$$\mathbf{x}^T E\Lambda^{-1/2}\Lambda^{-1/2}E^T\mathbf{x} = 1$$
$$\mathbf{x}^T\mathbf{K}\mathbf{K}^T\mathbf{x} = 1 \qquad \mathbf{K} = E\Lambda^{-1/2}$$

now for any point $\mathbf{y} = [x, y]^T$ which is on the unit circle, $\mathbf{y}^T\mathbf{y} = 1$, so

$$\mathbf{x}^T\mathbf{K}\mathbf{K}^T\mathbf{x} = \mathbf{y}^T\mathbf{y}$$
$$\mathbf{K}^T\mathbf{x} = \mathbf{y}$$
$$\Rightarrow \mathbf{x} = E\Lambda^{1/2}\mathbf{y}$$

$$\mathbf{x}(k+1) = \begin{bmatrix} 0.8462 & 0 \\ 0.1363 & 0.7866 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 0.3684 \\ 0.0292 \end{bmatrix} u(k) + \mathbf{w}(k)$$

$$y(k) = \begin{bmatrix} 0 & 1 \end{bmatrix} \mathbf{x}(\mathbf{k}) + \nu(k)$$



Estimate of h_1

Filtered estimate of h_2

Can we extract information about h_1 given only h_2?

Observability

$$\text{rank of } \begin{bmatrix} \mathbf{C} \\ \mathbf{C\Phi} \\ \vdots \\ \mathbf{C\Phi}^{n-1} \end{bmatrix} = n = \text{state dimension}$$

Can we extract information about h_1 given only h_2?

$$\mathbf{C} = \begin{bmatrix} 0 & 1 \end{bmatrix}$$
$$\mathbf{\Phi} = \begin{bmatrix} \Phi_{11} & 0 \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$$
$$\mathbf{C\Phi} = \begin{bmatrix} \Phi_{21} & \Phi_{22} \end{bmatrix}$$
$$O_b = \begin{bmatrix} 0 & 1 \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$$

yes

Can we extract information about h_2 given only h_1?

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$
$$\mathbf{\Phi} = \begin{bmatrix} \Phi_{11} & 0 \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$$
$$\mathbf{C\Phi} = \begin{bmatrix} \Phi_{11} & 0 \end{bmatrix}$$
$$O_b = \begin{bmatrix} 1 & 0 \\ \Phi_{11} & 0 \end{bmatrix}$$

no

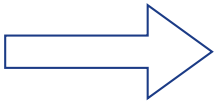
Linear Estimator Structure

Filter

$$\hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1)$$
$$\hat{x}(k|k) = \hat{x}(k|k-1) + \bar{K}(y(k) - C\hat{x}(k|k-1))$$

One-step-ahead predictor

$$\hat{x}(k|k) = \hat{x}(k|k-1) + \bar{K}(y(k) - C\hat{x}(k|k-1))$$
$$\hat{x}(k+1|k) = A\hat{x}(k|k) + Bu(k)$$



$$\hat{x}(k+1|k) = A\hat{x}(k|k-1) + Bu(k) + \underbrace{A\bar{K}}_{K(k)}(y(k) - C\hat{x}(k|k-1))$$

Estimation Error Equation

Define one-step-ahead prediction error, $x_e(k) \stackrel{\Delta}{=} x(k) - \hat{x}(k|k-1)$

Then,

$$x_e(k+1) = (A - K(k)C)x_e(k) + \omega(k) - K(k)\nu(k)$$

Define filtered estimation error, $\bar{x}_e(k) \stackrel{\Delta}{=} x(k) - \hat{x}(k|k)$

Then,

$$\bar{x}_e(k+1) = (A - \bar{K}(k+1)CA)\bar{x}_e(k) + (I - \bar{K}(k+1))\omega(k) - \bar{K}(k+1)\nu(k+1)$$

We can use either for the design since $K(k) = A\bar{K}(k)$

Kalman Filter as an Optimal Linear State Estimator

For simplicity, assume $R_{\omega,\nu} = 0$

Base the design of $K(k)$ on

$$x_e(k+1) = (A - K(k)C)x_e(k) + \omega(k) - K(k)\nu(k)$$

Let

$$\begin{aligned} P(k) &= \mathrm{Cov}\{x_e(k)\} \\ &= \mathbb{E}\left\{(x_e(k) - \mathbb{E}\{x_e(k)\})\,(x_e(k) - \mathbb{E}\{x_e(k)\})^T\right\} \end{aligned}$$

Note that $\mathbb{E}\{x_e(k)\} = 0$ for all $k \geq 0$ if $\mathbb{E}\{x_e(0)\} = 0$

From the error propagation equation,

$$\begin{aligned} P(k+1) &= \mathbb{E}\left\{x_e(k+1)x_e^T(k+1)\right\} \\ &= (A - K(k)C)P(k)(A - K(k)C)^T + R_\omega + K(k)R_\nu K^T(k) \end{aligned}$$

Choose $K(k)$ such that $\alpha^T P(k+1)\alpha$ is minimized for an arbitrary choice of α

(Note that this minimizes the sum of the variances, i.e., the mean square estimation error, and more)

From simple substitution and some algebra,

$$\begin{aligned}\alpha^T P(k+1)\alpha =& \alpha^T \left[AP(k)A^T + R_\omega - K(k)CP(k)A^T \right. \\ & \left. -AP(k)C^T K^T(k) + K(k)(R_\nu + CP(k)C^T)K^T(k)\right]\alpha\end{aligned}$$

Completing the squares,

$$\begin{aligned}\alpha^T P(k+1)\alpha = \alpha^T & \left\{\left[K(k) - AP(k)C^T\left(R_\nu + CP(k)C^T\right)^{-1}\right]\left[R_\nu + CP(k)C^T\right]\right. \\ & \left.\times \left[K(k) - AP(k)C^T\left(R_\nu + CP(k)C^T\right)^{-1}\right]^T\right\}\alpha \\ & + \alpha^T\left[AP(k)A^T + R_\omega - AP(k)C^T\left(R_\nu + CP(k)C^T\right)^{-1}CP(k)A^T\right]\alpha\end{aligned}$$

The optimal solution is independent of the choice of α

$$K(k) = AP(k)C^T(R_\nu + CP(k)C^T)^{-1}$$

and the corresponding covariance matrix is

$$P(k+1) = AP(k)A^T + R_\omega - AP(k)C^T\left(R_\nu + CP(k)C^T\right)^{-1}CP(k)A^T$$

The above equation is called Riccati Difference Equation (RDE)

The above provides a way to recursively construct the estimate $\hat{x}(k+1|k)$ (or $\hat{x}(k|k)$).

1. Start with some initial estimate $\hat{x}(0|-1)$ and the corresponding $P(0)$
2. Based on $P(0)$, compute $K(0)$ and $P(1)$
3. Also, with $K(0)$ and $y(0)$, $u(0)$, we can compute $\hat{x}(1|0)$
4. Keep repeating the calculation

Steady-State Kalman Filter

If the system is detectable, $P(k) \to P_\infty$ and therefore $K(k) \to K_\infty$ as $k \to \infty$, where

$$K_\infty = AP_\infty C^T (R_\nu + CP_\infty C^T)^{-1}$$

and

$$P_\infty = AP_\infty A^T + R_\omega - AP_\infty C^T \left(R_\nu + CP_\infty C^T\right)^{-1} CP_\infty A^T$$

The above is referred to as the Algebraic Riccati Equation (ARE) or simply Riccati Equation

If we are willing to accept some sub optimality during the initial transient period, we can compute K_∞ offline and implement the time-invariant estimator

If (C, A) is a detectable pair and $(A, R_\omega^{1/2})$ is a stabilizable pair, the steady-state Kalman filter is stable in the sense of an observer (a minimum condition needed for implementation)

Kalman Filter as the Optimal Bayesian Estimator for Gaussian Systems

Further assume that R_ω and R_ν are Gaussian white noise sequences.

Also assume that $x(0)$ is a Gaussian vector. Then, note that the whole system is Gaussian.

We are interested in computing $\mathbb{E}\{x(k)|Y(k)\}$ and $\text{cov}\{x(k)|Y(k)\}$,
where $Y(k) = \{y(0),\ y(1),\ \cdots,\ y(k)\}$

Given: $\mathbb{E}\{x(k-1)|Y(k-1)\}$ and $\text{cov}\{x(k-1)|Y(k-1)\}$

Compute: $\mathbb{E}\{x(k)|Y(k)\}$ and $\text{cov}\{x(k)|Y(k)\}$

Note that we only need to compute the first two moments.

Sequential Estimation Problem

$$P(H|D) = \frac{P(H) \times P(D|H)}{P(D)}$$

H: hypothesis, D: data

$p(H_t|H_{t-1}, D_t)$: this is what we are doing

Posterior density is estimated in two stages:

1. Prediction step: posterior density at previous time step is propagated into the next time step through the transition density

$$p\left[x(k)|Y^{k-1}\right] = \int p[x(k)|x(k-1)]p[x(k-1)|Y^{k-1}]dx(k-1)$$

2. Update step: involves application of Bayes' rule

$$p[x(k)|Y^k] \approx \underbrace{p[y(k)|x(k)]}_{\text{likelihood}} \times \underbrace{p[x(k)|Y^{k-1}]}_{\text{prior}}$$

Optimal State Estimate

- Minimum mean square error (MMSE)
  - Aimed at finding the conditional mean of x
- Maximum a posterior (MAP)
  - Aimed at finding **mode** of posterior probability density function
- Measure of accuracy of a state estimate
  - Covariance

$$x(k+1) = Ax(k) + Bu(k) + w(k)$$
$$y(k) = Cx(k) + \nu(k)$$

$$p(w) = \mathcal{N}(0,\ R_w) \qquad p(\nu) = \mathcal{N}(0,\ R_\nu)$$

$$p\left[x(k-1)|Y^{k-1}\right] = \mathcal{N}\left[\hat{x}(k-1|k-1), P(k-1|k-1)\right]$$

- initial state distribution is assumed to be normal

$$p\left[x(k)|Y^{k-1}\right] = \mathcal{N}\left[\hat{x}(k|k-1), P(k|k-1)\right]$$

- prior

$$p\left[x(k)|Y^{k}\right] = \mathcal{N}\left[\hat{x}(k|k), P(k|k)\right]$$

- posterior are Gaussian w/ system being linear and noise being Gaussian

$$p\left[x(k)|Y^k\right] \propto p[y(k)|x(k)]p\left[x(k)|Y^{k-1}\right]$$

$$p\left[x(k)|Y^{k-1}\right] = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} \exp\left[-\frac{1}{2}\left\{x(k)-\hat{x}(k|k-1)\right\}^T P(k|k-1)^{-1}\left\{x(k)-\hat{x}(k|k-1)\right\}\right]$$

$$\xi(k) \stackrel{\Delta}{=} x(k) - \hat{x}(k|k-1)$$

$$p\left[x(k)|Y^{k-1}\right] = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} \exp\left[-\frac{1}{2}\xi^T(k)P(k|k-1)^{-1}\xi(k)\right]$$

$$p\left[y(k)|x(k)\right] = \frac{1}{(2\pi)^{n/2}|R_\nu|^{1/2}} \exp\left[-\frac{1}{2}\left\{y(k)-Cx(k)\right\}^T R_\nu^{-1}\left\{y(k)-Cx(k)\right\}\right]$$

$$e(k) \stackrel{\Delta}{=} y(k) - Cx(k)$$

$$p\left[y(k)|x(k)\right] = \frac{1}{(2\pi)^{n/2}|R_\nu|^{1/2}} \exp\left[-\frac{1}{2}e^T(k)R_\nu^{-1}e(k)\right]$$

$$p\left[x(k)|Y^k\right] \propto \exp\left[-\frac{1}{2}\left\{e^T(k)R_\nu^{-1}e(k)+\xi^T(k)P(k|k-1)^{-1}\xi(k)\right\}\right]$$

$$\min_{x(k)} J = \min_{x(k)} \frac{1}{2} \left[e^T(k) R_\nu^{-1} e(k) + \xi^T(k) P(k|k-1)^{-1} \xi(k) \right]$$

$$e(k) = y(k) - Cx(k) \qquad \xi(k) = x(k) - \hat{x}(k|k-1)$$

$$\frac{\partial J}{\partial x(k)} = \left[-C^T R_\nu^{-1} e(k) + P^{-1} \xi(k) \right] = 0$$

$$\left[-C^T R_\nu^{-1} \left\{ y(k) - Cx(k) \right\} + P^{-1} \left\{ x(k) - \hat{x}(k|k-1) \right\} \right] = 0$$

$$-C^T R_\nu^{-1} y + C^T R_\nu^{-1} C x + P^{-1} x - P^{-1} \hat{x} = 0$$

$$\left[P^{-1} + C^T R_\nu^{-1} C \right] x = P^{-1} \hat{x} + C^T R_\nu^{-1} y$$

$$\left[P^{-1} + C^T R_\nu^{-1} C \right] x = P^{-1} \hat{x} + C^T R_\nu^{-1} y - C^T R_\nu^{-1} C \hat{x} + C^T R_\nu^{-1} C \hat{x}$$

$$\left[P^{-1} + C^T R_\nu^{-1} C \right] x = \left[P^{-1} + C^T R_\nu^{-1} C \right] \hat{x} + C^T R_\nu^{-1} \left[y - C\hat{x} \right]$$

$$x = \hat{x} + \left[P^{-1} + C^T R_\nu^{-1} C \right]^{-1} C^T R_\nu^{-1} \left[y - C\hat{x} \right]$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K \left\{ y - C\hat{x}(k|k-1) \right\}$$

$$K = \left[P^{-1} + C^T R_\nu^{-1} C \right]^{-1} C^T R_\nu^{-1}$$

$$K = \left[P^{-1} + C^T R_\nu^{-1} C\right]^{-1} C^T R_\nu^{-1}$$

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B\left[C^{-1} + DA^{-1}B\right]^{-1} DA^{-1}$$

$$\left[P^{-1} + C^T R_\nu^{-1} C\right]^{-1} = P - PC^T\left[R_\nu + CPC^T\right]^{-1} CP$$

$$K = \left[P - PC^T\left[R_\nu + CPC^T\right]^{-1} CP\right] C^T R_\nu^{-1}$$

$$K = \underline{P(k|k-1)}C^T\left[CP(k|k-1)C^T + R_\nu\right]^{-1}$$

second moment

Note $z = \Phi X$ $X \sim \mathcal{N}(\mu, P)$ $z \sim \mathcal{N}(\Phi\mu, \Phi P \Phi^T)$

$$x(k) = Ax(k-1) + Bu(k-1) + w(k)$$

$$\mathcal{N}\left(\hat{x}(k-1|k-1), P(k-1|k-1)\right)$$

Prediction:

$$\therefore \quad \hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1) + 0$$

$$P(k|k-1) = AP(k-1|k-1)A^T + R_w$$

Update: posterior ≃ prior x likelihood

$$x(k+1) = Ax(k) + Bu(k) + w(k)$$
$$y(k) = Cx(k) + \nu(k)$$

$$p\left[x(k)|Y^{k-1}\right] = \mathcal{N}\left[\hat{x}(k|k-1), P(k|k-1)\right]$$

$$\begin{aligned}\mathbb{E}\left[x(k)|Y^{k-1}\right] &= \hat{x}(k|k-1) \\ &= \mathbb{E}\left[Ax(k-1) + Bu(k-1) + w(k-1)|\, Y^{k-1}\right]\end{aligned}$$

$$\hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1)$$

$$P(k|k-1) = \mathrm{cov}\left[\xi(k|k-1)\right] = \mathbb{E}\left[\xi(k|k-1)\xi(k|k-1)^T\right]$$

$$\begin{aligned}\xi(k|k-1) &= x(k) - \hat{x}(k|k-1) \\ &= \left[Ax(k-1) + Bu(k-1) + w(k-1)\right] - \left[A\hat{x}(k-1|k-1) + Bu(k-1)\right] \\ &= \left[A\xi(k-1|k-1) + w(k-1)\right]\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\xi(k|k-1)\xi(k|k-1)^T\right] &= \mathbb{E}\left[\left\{A\xi(k-1|k-1)+w(k-1)\right\}\left\{A\xi(k-1|k-1)+w(k-1)\right\}^T\right] \\ &= \mathbb{E}\left[A\xi(k-1|k-1)\xi^T(k-1|k-1)A^T\right]+\mathbb{E}\left[w(k-1)w(k-1)^T\right]\end{aligned}$$

$$P(k|k-1) = AP(k-1|k-1)A^T + R_w$$

Why is K a fcn of two covariance matrices?

$$\mathbb{E}[y(k)|Y^{k-1}] = \hat{y}(k|k-1) = CA\hat{x}(k-1|k-1) + CBu(k-1)$$

$$\begin{aligned} e(k|k-1) &= y(k) - \hat{y}(k|k-1) \\ &= [CAx(k-1) + CBu(k-1) + Cw(k-1) + \nu(k)] - [CA\hat{x}(k-1|k-1) + CBu(k-1)] \end{aligned}$$

$$e(k|k-1) = [CA\xi(k-1|k-1) + Cw(k-1) + \nu(k)]$$

$$e(k|k-1) = [C\xi(k|k-1) + \nu(k)]$$

$$\begin{aligned} \text{cov}\left[e(k|k-1)\right] &= \mathbb{E}\left[\{C\xi(k|k-1) + \nu(k)\}\{C\xi(k|k-1) + \nu(k)\}^T\right] \\ &= CP(k|k-1)C^T + R_\nu \end{aligned}$$

$$\begin{aligned} \text{cov}\left[\xi(k|k-1)e^T(k|k-1)\right] &= \mathbb{E}\left[\xi(k|k-1)\{C\xi(k|k-1) + \nu(k)\}^T\right] \\ &= P(k|k-1)C^T \end{aligned}$$

$$\begin{aligned} K &= \text{cov}\left[\xi(k|k-1)e^T(k|k-1)\right]\left[e(k|k-1)\right]^{-1} \\ &= P(k|k-1)C^T\left[CP(k|k-1)C^T + R_\nu\right]^{-1} \end{aligned}$$

$$p[x(k)|Y^k] = \mathcal{N}\left[\hat{x}(k|k), P(k|k)\right]$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K\left[y(k) - C\hat{x}(k|k-1)\right]$$

$$P(k|k) = \mathrm{cov}\left[\xi(k|k)\right] = \mathbb{E}\left[\xi(k|k)\xi(k|k)^T\right]$$

$$\begin{aligned}
\xi(k|k) &= x(k) - \hat{x}(k|k) \\
&= \left[Ax(k-1) + Bu(k-1) + w(k-1)\right] - \left[A\hat{x}(k-1|k-1) + Bu(k-1) + Ke(k|k-1)\right] \\
&= A\xi(k-1|k-1) + w(k-1) - Ke(k|k-1) \\
&= \xi(k|k-1) - Ke(k|k-1)
\end{aligned}$$

$$\begin{aligned}
P(k|k) &= \mathbb{E}\left[\xi(k|k)\xi(k|k)^T\right] \\
&= \mathbb{E}\left[\left\{\xi(k|k-1) - Ke(k|k-1)\right\}\left\{\xi(k|k-1) - Ke(k|k-1)\right\}^T\right] \\
&= P(k|k-1) + K\mathrm{cov}\left[e(k|k-1)\right]K^T - PC^TK^T - KCP \\
&= P(k|k-1) + P(k|k-1)C^T\left[CP(k|k-1)C^T + R_\nu\right]^{-1}\left[CP(k|k-1)C^T + R_\nu\right]K^T - P(k|k-1)C^TK^T - KCP(k|k-1) \\
&= P(k|k-1) - KCP(k|k-1) = [I - KC]P(k|k-1)
\end{aligned}$$

Kalman filter

Prediction step $\mathcal{N}\left[\hat{x}(k|k-1), P(k|k-1)\right]$

$$\hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1)$$

$$P(k|k-1) = AP(k-1|k-1)A^T + R_w$$

Correction step $\mathcal{N}\left[\hat{x}(k|k), P(k|k)\right]$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + Ke(k|k-1)$$

$$P(k|k) = [I - KC]P(k|k-1)$$

$$K = P(k|k-1)C^T\left[CP(k|k-1)C^T + R_\nu\right]^{-1}$$

Kalman Gain Computation

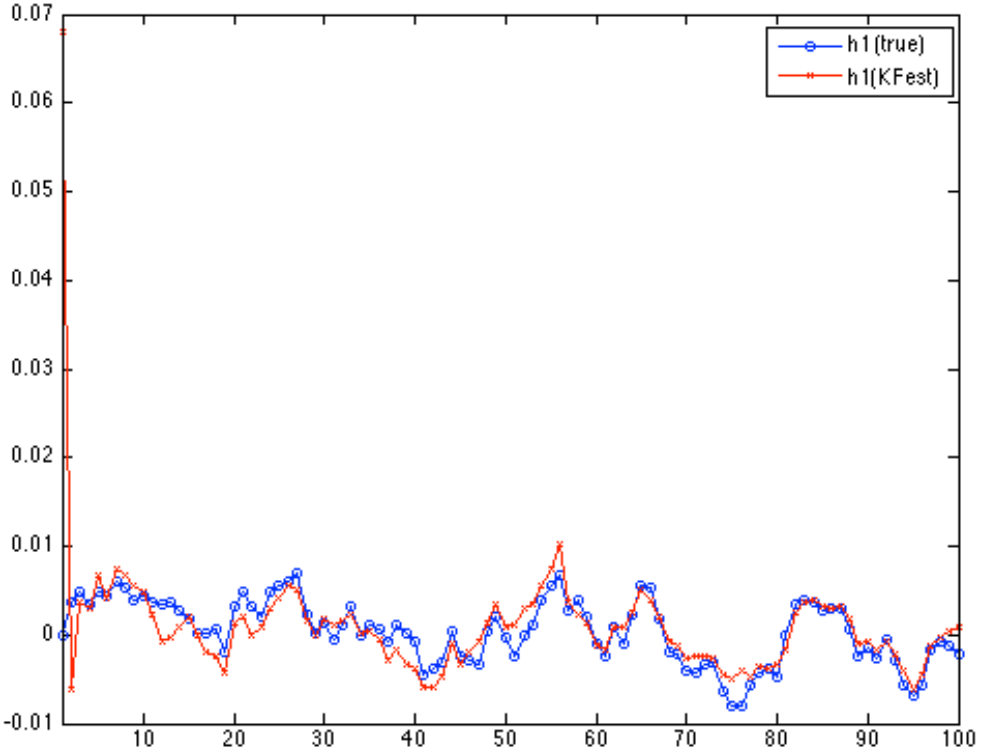
$e(k|k-1) = [y(k) - C\hat{x}(k|k-1)]$ (innovations)

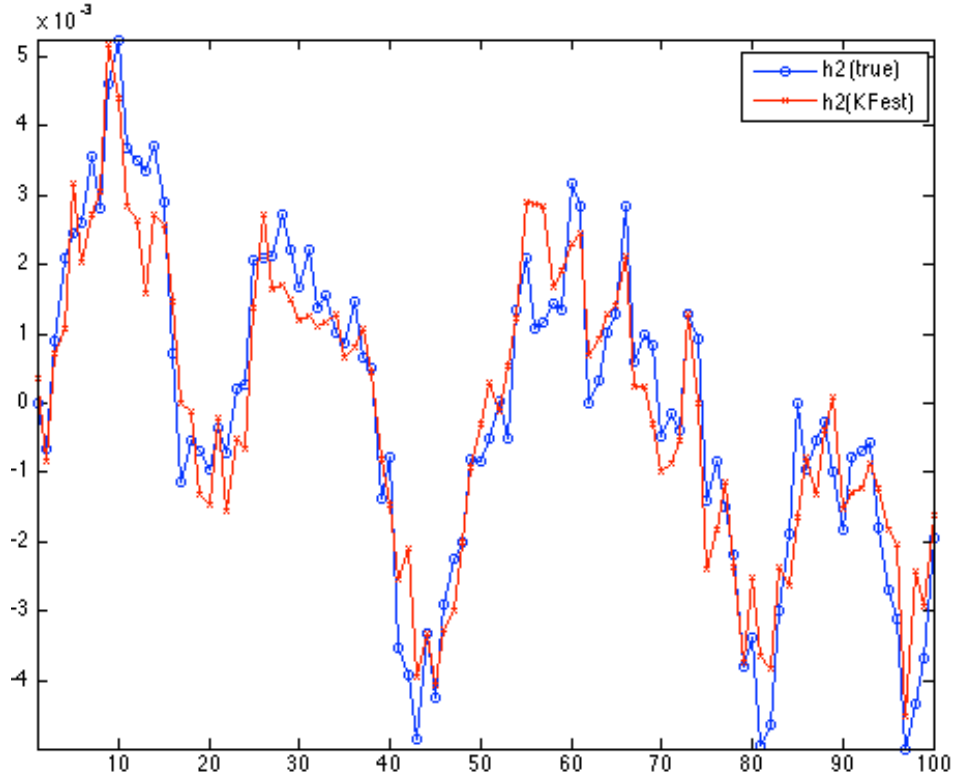
$\xi(k) = x(k) - \hat{x}(k|k-1)$ (estimation error)

$$K(k) = P_{\xi,e}\left[P_{e,e}(k)\right]^{-1}$$

(covariance matrices)

Simulation results





Issues in state estimation

- Generates the maximum likelihood (or minimum variance) estimates of the states when noises are Gaussian and the system is linear.

- Model accuracy is critical to state estimation.

- Noise model parameters: measurement and state noise covariance matrices (R_ω and R_ν) are difficult to estimate. These matrices are typically used as tuning parameters.

Difficulties with Kalman filter

- Linear perturbation model: range of applicability limited to a narrow region around an operating point
- Cannot be used for
  - strongly nonlinear systems or operation over wide operating range
  - batch/semi-batch processes
- State estimation using nonlinear dynamic model:
  - Extended Kalman filter (EKF)
  - Ensemble Kalman filter (EnKF)
  - Unscented Kalman filter (UKF)