Kalman Predictor/Filter

Stochastic Processes

Optimal Estimation and Prediction

Let us assume now that our plant is:

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

with initial condition x_0 given, and where w_k and v_k are independent zero-mean, white stochastic processes (independent of x_0).

$$E(w_k) = 0, E(v_k) = 0, E\left(\begin{bmatrix} w_k \\ v_k \end{bmatrix}\begin{bmatrix} w_l^T & v_l^T \end{bmatrix}\right) = \begin{bmatrix} Q_k & 0 \\ 0 & R_k \end{bmatrix}\delta_{k,l}, \qquad Q_k, R_k \geq 0 \ \ \forall k$$

Prediction Problem: How can we estimate x_k given measurements of $\{u_l, y_l: l \leq k-d\}$? We denote the d-step-ahead prediction as

$$\hat{x}_{k|k-d}$$

Filtering Problem: How do we compute the filtered estimate $\hat{x}_{k|k}$?

Stochastic Processes

How do we take two independent, unbiased measurements, x_1 and x_2, with variances σ_1^2 and σ_2^2, of the same quantity x and combine them to get a better, unbiased estimate?

$$\hat{x} = a x_1 + (1-a) x_2 \qquad \text{unbiased}$$

Covariance $$E\left(\left(\hat{x} - E(\hat{x})\right)^2\right) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$$

$$\frac{\partial E\left(\left(\hat{x} - E(\hat{x})\right)^2\right)}{\partial a} = 2a\sigma_1^2 - 2(1-a)\sigma_2^2 = 0 \Rightarrow a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$$

Stochastic Processes

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$$

Linear, variance-weighted sum. Nothing mysterious.

$$E\left(\hat{x}^2\right) = \frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \left[\left(\sigma_1^2\right)^{-1} + \left(\sigma_2^2\right)^{-1}\right]^{-1} \leq \min\left\{\sigma_1^2, \sigma_2^2\right\}$$

The quality of the estimation is improved. Can we do similar things when x satisfies a state equation? We just need to propagate means and variances.

• taking into account the effect of state equation (easy, deterministic)

• taking into account the new measurements (a bit harder, conditional probability)

Stochastic Processes

Minimum Variance Estimates

We want to find the estimate $\hat{x}_k$ which minimizes

$$E\left(\left\|x_k - \hat{x}_k\right\|^2 \middle| y_{k-d}, y_{k-d-1}, \ldots\right)$$

The minimum variance estimate is given by the conditional mean

$$\boxed{\hat{x}_k = E\left(x_k \middle| y_{k-d}, y_{k-d-1}, \ldots\right)}$$

Most important property in estimation

Proof: $E\left(\left(x_k - \hat{x}_k\right)^2 \middle| Y\right) = E\left(x_k^T x_k \middle| Y\right) - 2E\left(x_k^T \hat{x}_k \middle| Y\right) + E\left(\hat{x}_k{}^T \hat{x}_k \middle| Y\right)$

$$= \left(\hat{x}_k - E\left(x_k \middle| Y\right)\right)^2 + E\left(x_k^T x_k \middle| Y\right) - \left(E\left(x_k \middle| Y\right)\right)^2$$

is minimized when $\hat{x}_k = E\left(x_k \middle| Y\right)$

The *conditional mean* is the *least square estimate*.

Stochastic Processes

Jointly Gaussian Random variables

x is a gaussian random n-vector if its probability density function is of the form

$$p(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-m)^T \Sigma^{-1} (x-m)}$$

where m is the mean value of x, $E(x)$, and non-singular matrix Σ is the covariance matrix, $E[(x-m)(x-m)^T]$.

If x is gaussian $N(m, \Sigma)$ then $y=Ax+b$ is gaussian $N(Am+b, A\Sigma A^T)$.

If x and y are commensurate gaussian random n-vectors, then $x+y$ is a gaussian random n-vector $N(m_x+m_y, \Sigma_{x+y})$.

$$\Sigma_{x+y} = \Sigma_{xx} + \Sigma_{xy} + \Sigma_{yx} + \Sigma_{yy}$$

Stochastic Processes

If x and y are jointly gaussian, i.e, $[x^T\ y^T]^T$ is a gaussian process $N([m_x^T\ m_y^T], \Sigma)$ with

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

Then,

$$E\left[x \middle| y\right] = m_x - \Sigma_{xy}\Sigma_{yy}^{-1} m_y + \Sigma_{xy}\Sigma_{yy}^{-1} y \qquad (1)$$

also a gaussian random n-vector.

$$E\left[\left(x - E\left[x|y\right]\right)^2 \middle| y\right] = E\left[\left(x - E\left[x|y\right]\right)^2\right] = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \qquad (2)$$

The gaussian conditional-mean estimate has the conditional error variance equal to the unconditional error variance (not true if not gaussian).

$$E\left[x - E\left[x|y\right] \middle| y\right] = E\left[x|y\right] - E\left[x|y\right] = 0$$

The conditional-mean estimate is unbiased.

Stochastic Processes

Kalman Filter

State equations:

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

Assumptions:

- x_0 is gaussian with mean $\bar{x}_0$ and covariance P_0.
- w_k is gaussian, zero-mean, white process ($E\left(w_k w_l^T\right) = Q_k \delta_{k,l}$), independent of x_0 and v_k.
- v_k is gaussian, zero-mean, white process ($E\left(v_k v_l^T\right) = R_k \delta_{k,l}$), independent of x_0 and w_k.

Everything is gaussian --- linear operations preserve gaussian properties

Stochastic Processes

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad \text{is gaussian} \quad N\left(\begin{bmatrix} \bar{x}_0 \\ C_0 \bar{x}_0 \end{bmatrix}, \begin{bmatrix} P_0 & P_0 C_0^T \\ C_0 P_0 & C_0 P_0 C_0^T + R_0 \end{bmatrix} \right)$$

From (1): $\hat{x}_{0|0} = \bar{x}_0 - P_0 C_0^T \left(C_0 P_0 C_0^T + R_0 \right)^{-1} C_0 \bar{x}_0 + P_0 C_0^T \left(C_0 P_0 C_0^T + R_0 \right)^{-1} y_0$

$= \bar{x}_0 + P_0 C_0^T \left(C_0 P_0 C_0^T + R_0 \right)^{-1} \left(y_0 - C_0 \bar{x}_0 \right)$ Filter

From (2): $\sum_{0|0} = P_0 - P_0 C_0^T \left(C_0 P_0 C_0^T + R_0 \right)^{-1} C_0 P_0$ Filter covariance

Time update: $\hat{x}_{1|0} = A_0 \hat{x}_{0|0} + B_0 u_0$ One-step predictor

$\sum_{1|0} = A_0 \sum_{0|0} A_0^T + Q_0$ Predictor covariance

$\hat{y}_{1|0}$ is gaussian $N\left(C_1 \hat{x}_{1|0}, C_1 \sum_{1|0} C_1^T + R_1 \right)$

Stochastic Processes

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \Bigg| y_0 \quad \text{is gaussian} \quad N\left(\begin{bmatrix} \hat{x}_{1|0} \\ C_1 \hat{x}_{1|0} \end{bmatrix}, \begin{bmatrix} \Sigma_{1|0} & \Sigma_{1|0} C_1^T \\ C_1 \Sigma_{1|0} & C_1 \Sigma_{1|0} C_1^T + R_1 \end{bmatrix} \right)$$

From (1): $$\hat{x}_{1|1} = \hat{x}_{1|0} - \Sigma_{1|0} C_1^T \left(C_1 \Sigma_{1|0} C_1^T + R_1 \right)^{-1} C_1 \hat{x}_{1|0} + \Sigma_{1|0} C_1^T \left(C_1 \Sigma_{1|0} C_1^T + R_1 \right)^{-1} y_1$$

$$= \hat{x}_{1|0} + \Sigma_{1|0} C_1^T \left(C_1 \Sigma_{1|0} C_1^T + R_1 \right)^{-1} \left(y_1 - C_1 \hat{x}_{1|0} \right)$$ Filter

From (2): $$\Sigma_{1|1} = \Sigma_{1|0} - \Sigma_{1|0} C_1^T \left(C_1 \Sigma_{1|0} C_1^T + R_1 \right)^{-1} C_1 \Sigma_{1|0}$$ Filter covariance

Time update: $$\hat{x}_{2|1} = A_1 \hat{x}_{1|1} + B_1 u_1$$ One-step predictor

$$\Sigma_{2|1} = A_1 \Sigma_{1|1} A_1^T + Q_1$$ Predictor covariance

$\hat{y}_{2|1}$ is gaussian $N\left(C_2 \hat{x}_{2|1}, C_2 \Sigma_{2|1} C_2^T + R_2 \right)$

Stochastic Processes

Generalizing, for the plant

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

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \Sigma_{k|k-1} C_k^T \left(C_k \Sigma_{k|k-1} C_k^T + R_k \right)^{-1} \left(y_k - C_k \hat{x}_{k|k-1} \right)$$ Filter

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + B_k u_k$$ Predictor

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1} C_k^T \left(C_k \Sigma_{k|k-1} C_k^T + R_k \right)^{-1} C_k \Sigma_{k|k-1}$$

$$\Sigma_{k+1|k} = A_k \Sigma_{k|k} A_k^T + Q_k$$

These are the discrete-time Kalman filtering equations
They consist of time update and measurement update parts
One-step-ahead predictions and filtered estimates are given
The prediction error is used to update the state estimate

Stochastic Processes

Kalman Filter properties

$$x_{k+1} = A_k x_k + B_k u_k + w_k$$

$$y_k = C_k x_k + v_k$$

Kalman Predictor:

$$\hat{x}_{k+1|k} = \left(A_k - M_k C_k\right)\hat{x}_{k|k-1} + B_k u_k + M_k y_k$$

$$M_k = A_k \sum_{k|k-1} C_k^T \left(C_k \sum_{k|k-1} C_k^T + R_k\right)^{-1}$$

$$\sum_{k+1|k} = A_k \sum_{k|k-1} A_k^T - A_k \sum_{k|k-1} C_k^T \left(C_k \sum_{k|k-1} C_k^T + R_k\right)^{-1} C_k \sum_{k|k-1} A_k^T + Q_k$$

Kalman Filter:

$$\hat{x}_{k|k} = \left(I - L_k C_k\right) A_k \hat{x}_{k-1|k-1} + \left(I - L_k C_k\right) B_{k-1} u_{k-1} + L_k y_k$$

$$L_k = \sum_{k|k-1} C_k^T \left(C_k \sum_{k|k-1} C_k^T + R_k\right)^{-1}$$

$$\sum_{k|k} = \left(I - L_k C_k\right) \sum_{k|k-1}$$

- The Kalman filter and the Kalman predictor are state observers!
- $\sum_{k|k-1}$ satisfies a Riccati Difference Equation (RDE)

Stochastic Processes

Kalman Filter properties

Kalman Filter RDE:

$$\Sigma_{k+1|k} = A_k \Sigma_{k|k-1} A_k^T - A_k \Sigma_{k|k-1} C_k^T \left(C_k \Sigma_{k|k-1} C_k^T + R_k \right)^{-1} C_k \Sigma_{k|k-1} A_k^T + Q_k$$

LQR RDE:

$$S_k = A_k^T S_{k+1} A_k - A_k^T S_{k+1} B_k \left(B_k^T S_{k+1} B_k + R_k \right)^{-1} B_k^T S_{k+1} A_k + Q_k$$

Covariances:

$$\Sigma_{k|k-1} = E\left[\left(x_k - \hat{x}_{k|k-1} \right) \left(x_k - \hat{x}_{k|k-1} \right)^T \right]$$

$$\Sigma_{k|k} = E\left[\left(x_k - \hat{x}_{k|k} \right) \left(x_k - \hat{x}_{k|k} \right)^T \right]$$

Stochastic Processes

Kalman Filter properties

Observer error equations:

Kalman Predictor:

$$x_{k+1} - \hat{x}_{k+1|k} = A_k x_k + B_k u_k + w_k - \left(A_k - M_k C_k\right)\hat{x}_{k|k-1} - B_k u_k - M_k\left(C_k x_k + v_k\right)$$

$$\tilde{x}_{k+1|k} = \left(A_k - M_k C_k\right)\tilde{x}_{k|k-1} + w_k - M_k v_k$$

Kalman Filter:

$$x_{k+1} - \hat{x}_{k+1|k+1} = A_k x_k + B_k u_k + w_k - \left(I - L_{k+1}C_{k+1}\right)\left(A_k \hat{x}_{k|k} + B_k u_k\right) + L_{k+1}\left(C_{k+1}x_{k+1} + v_{k+1}\right)$$

$$\tilde{x}_{k+1|k+1} = \left(I - L_{k+1}C_{k+1}\right)A_k \tilde{x}_{k|k} - L_{k+1}v_{k+1} + \left(I - L_{k+1}C_{k+1}\right)w_k$$

Stochastic Processes

Kalman Filter properties

$$\tilde{x}_{k+1|k} = \left(A_k - M_k C_k\right)\tilde{x}_{k|k-1} + w_k - M_k v_k$$

Use this equation to compute the covariance propagation

$$E\left(\tilde{x}_{k+1|k}\tilde{x}_{k+1|k}^T\right) = X_{k+1} = \left(A_k - M_k C_k\right)X_k\left(A_k - M_k C_k\right)^T + M_k R_k M_k^T + Q_k$$

The Kalman filter is an observer in which we balance

- the stabilizing properties of big gain M_k

- the role of this stability in reducing the Q_k effect effect in X_k

- the amplification of the measurement noise variance R_k effect in X_k

The Kalman filter is an optimal observer which balances the smoothing of measurements noise v_k against the tracking of the state knocked around by w_k.

Stochastic Processes

Kalman Filter properties

• The Kalman filter is a linear, discrete-time, finite-dimensional system.

• The input of the filter is y_k and the output is $\hat{x}_{k|k-1}$.

• The gains M_k, L_k and covariance matrix $\sum_{k|k-1}$ do not depend on the data and can be pre-computed before the filter is run.

• The matrix $\sum_{k|k-1}$ is the error covariance matrix of the state estimator

$$\sum_{k|k-1} = E\left[\left(x_k - \hat{x}_{k|k-1}\right)\left(x_k - \hat{x}_{k|k-1}\right)^T\right]$$

•Even if the underlying signal model $[A, B, C, Q, R]$ is time-invariant, the resulting Kalman filter will be time-varying unless

$$\sum_{0|-1} = \sum_{\infty|\infty-1}$$

Stochastic Processes

Kalman Filter properties

- The Kalman filter equations are dual of the LQR equations.

$$A \to A^T, \quad B \to C^T, \quad Q^C \to Q^O, \quad R^C \to R^O$$

- The LQR RDE is replaced by the KF RDE evolving forwards in time.

Stochastic Processes

Extended Kalman Filter

For the nonlinear signal model

$$x_{k+1} = f\left(x_k, u_k\right) + w_k$$

$$y_k = h\left(x_k\right) + v_k$$

with the usual assumptions holding and all functions sufficiently continuously differentiable. The extended Kalman filter is given by

$$\hat{x}_{k+1|k} = f\left(\hat{x}_{k|k}, u_k\right)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k\left(y_k - h\left(\hat{x}_{k|k-1}\right)\right)$$

$$L_k = \sum_{k|k-1} C_k^T \left(C_k \sum_{k|k-1} C_k^T + R_k\right)^{-1}$$

$$\sum_{k+1|k} = A_k \sum_{k|k-1} A_k^T - A_k \sum_{k|k-1} C_k^T \left(C_k \sum_{k|k-1} C_k^T + R_k\right)^{-1} C_k \sum_{k|k-1} A_k^T + Q_k$$

$$A_k = \left.\frac{\partial f}{\partial x}\right|_{\hat{x}_{k|k}}, C_k = \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_{k|k}}$$

The Extended Kalman Filter is a non-linear filter! The linearization is only in the gain calculation.