Kalman Filter as the best linear estimator

Notation: - $y(k)$ instead y_k

- System starts $k=0$ measurements $y(1), \dots y(k)$
- Most important: 2 sources of uncertainty
  - $\to v(k)$ "process noise"
  - $\to w(k)$ "measurement noise"

  $v(k) = F_k W_k$

  $w(k) = E_k w_k$
- Variances: RV q. $\underline{Var(q)}$ Σ_q Σ^q
  - $\to Var(w(k)) = R(k)$
  - $\to Var(v(k)) = Q(k)$
  - $\to$ Estimation error: ε $Var(\varepsilon(k)) = P(k)$
  - $\to$ Initial state $Var(x(0)) = P_0$
- Estimates $\hat{x}_m(k)$ $\hat{x}_p(k)$

  $\hat{x}_{k|k}$ $\hat{x}_{k|k-1}$

$x(k) = x(k-1)$

RLS : Problem: Static x (no dynamics)

Observation: $y(k) = C(k)x + w(k)$; $y(k) \in \mathbb{R}^m$ $w(k) \in \mathbb{R}^m$

$x \in \mathbb{R}^n$

$\to$ Know: prior of x: mean $\bar{x} = E[x]$

$P_x = E[(x-\bar{x})(x-x)^T]$

$\to$ Meas. noise $w(k)$: zero-mean, known variance $E[w(k)] = 0$

$R(k) := Var(w(k))$

$\to \{x, w(1), w(2), \dots\}$ are independent.

Standard Weighted least squares

Collect . $\{y(1), y(2) \dots y(k)\}$

- Pseudo measurement: $y(0) = \bar{x} = I \cdot x + (\bar{x} - x)$

$$y(0{:}k) = C(0{:}k)x + \omega(0{:}k)$$

$$\begin{bmatrix} y(0) \\ \vdots \\ y(k) \end{bmatrix} = \begin{bmatrix} I \\ C(1) \\ \vdots \\ C(k) \end{bmatrix} x + \begin{bmatrix} \bar{x} - x \\ \omega(1) \\ \vdots \\ \omega(k) \end{bmatrix} \qquad R(0{:}k) = \begin{bmatrix} P_x & 0 & \cdots & 0 \\ 0 & R(1) & & \\ \vdots & & \ddots & \\ 0 & & & R(k) \end{bmatrix}$$

$$\hat{x}^{WLS}(k) = \arg\min_{\hat{x}} \ \varepsilon^T(\hat{x}, 0{:}k)\, R(0{:}k)\, \varepsilon(\hat{x}, 0{:}k)$$

$$\varepsilon(\hat{x}, 0{:}k) = y(0{:}k) - C(0{:}k)\hat{x}$$

$$\hat{x}^{WLS}(k) = \left(C(0{:}k)^T R(0{:}k)^{-1} C(0{:}k)\right)^{-1} C(0{:}k)^T R(0{:}k)^{-1} y(0{:}k)$$

Recursive update

Restrict to: $\hat{x}(k) = \hat{x}(k-1) + K(k)\left(y(k) - C(k)\,\hat{x}(k-1)\right)$

$\rightarrow \hat{x}(k)$ estimate of x at time k, using measurements $y(1), \dots y(k)$
& prior knowledge $\bar{x}, P_0$

$\rightarrow K(k)$: gain matrix, design variable

$\rightarrow \varepsilon(k) := x - \hat{x}(k)$ estimation

$$= x - \hat{x}(k-1) - K(k)\left(\cancel{y(k)} - C(k)\,\hat{x}(k-1)\right) \qquad C(k)x + \omega(k)$$

$$= \varepsilon(k-1) - K(k)C(k)\varepsilon(k-1) - K(k)\cdot\omega(k)$$

$$= \left(I - K(k)\cdot C(k)\right)\varepsilon(k-1) - K(k)\,\omega(k).$$

Expectation: $E[\varepsilon(k)] = E\left[\left(I - K(k)\,C(k)\right)\varepsilon(k-1) - K(k)\,\omega(k)\right]$

$$= \left(I - K(k)\cdot C(k)\right)E\left[\varepsilon(k-1)\right] - \cancel{K(k)\,E\left[\omega(k)\right]}$$

If we set $\hat{x}(0) = \bar{x} \Rightarrow E[\varepsilon(0)] = 0$ $\qquad \varepsilon(0) = x - \hat{x}(0)$

$$\Rightarrow E[\varepsilon(k)] = 0 \quad \forall k$$

Goal choose $K(k)$ intelligently

Specifically: minimize mean squared error

Cost: $J(k) := E\left[\varepsilon^T(k)\,\varepsilon(k)\right] = E\left[\mathrm{trace}\left(\varepsilon(k)\,\varepsilon(k)^T\right)\right]$

$= \mathrm{trace}\left(E\left[\varepsilon(k)\,\varepsilon(k)^T\right]\right)$

$= \mathrm{trace}\left(P(k)\right)$

$P(k) := E\left[\varepsilon(k)\,\varepsilon(k)^T\right]$

$= \mathrm{Var}\left(\varepsilon(k)\right)$

$$\varepsilon(k) = \left(I - K(k)\cdot C(k)\right)\varepsilon(k-1) - K(k)\cdot w(k)$$

$$P(k) = E\left[\varepsilon(k)\,\varepsilon(k)^T\right]$$

$$= E\left[\left(\left(I - K(k)\cdot C(k)\right)\varepsilon(k-1) - K(k)\,w(k)\right]\left[\text{---}\right]^T\right)$$

$$= E\left[\left(I - K(k)\cdot C(k)\right)\varepsilon(k-1)\varepsilon(k-1)^T\left(I - K(k)C(k)\right)^T\right]$$

$$+ E\left[K(k)\,w(k)\,w(k)^T K(k)^T\right]$$

$$- E\left[\left(I - K(k)\cdot C(k)\right)\varepsilon(k-1)w(k)^T K(k)^T\right]$$

$$- E\left[\quad\right]^T$$

$$= \left(I - K(k)\cdot C(k)\right)P(k-1)\left(I - K(k)\,C(k)\right)^T$$

$$+ K(k)\,R(k)\,K(k)^T$$

$$- \left(I - K(k)\,C(k)\right)\cancel{E\left[\varepsilon(k-1)\,w(k)^T\right]}^{\,0}K(k)$$

$$- \left(\quad\right)^T \quad 0$$

$$P(k) = \left(I - K(k)C(k)\right)P(k-1)\left(\text{---}\right)^T + K(k)\,R(k)\,K(k)^T$$

Goal: $\min J(k) = \min \mathrm{trace}\, P(k)$

Need: Results, state without proof

$\rightarrow \frac{\partial}{\partial A}\, trace(ABA^T) = 2AB \quad \text{if } B = B^T$

where: for a scalar function $g(A)$ for $A \in \mathbb{R}^{m \times n}$

$$\frac{\partial g}{\partial A} = \begin{bmatrix} \frac{\partial g(A)}{\partial A_{11}} & \cdots & \frac{\partial g(A)}{\partial A_{1,n}} \\ \vdots & & \vdots \\ \frac{\partial g(A)}{\partial A_{m,1}} & & \frac{\partial g(A)}{\partial A_{m,n}} \end{bmatrix}$$

$\rightarrow \frac{\partial}{\partial A}\, trace(AB) = B^T$

Omitting the index k

$$\rightarrow J = trace((I-KC)P(I-KC)^T + KRK^T)$$
$$= trace((I-KC)P(\quad)^T) + trace(KRK^T)$$

At optimum: $\frac{\partial J}{\partial K} = 0$

$$\Rightarrow 0 = 2(I-KC)P(-C) + 2K \cdot R$$

$$K(k) = P(k-1)C^T(k)\left(C(k)P(k-1)C(k)^T + R(k)\right)^{-1}$$

Kalman filter gain

Packard Notes Part 1, page 33
Part 3, Fact 16

RLS: init. $\hat{x}(0) = \bar{x}$ $\quad P(0) = P_x = Var(x)$

Observe $y(k)$

Update:
$$K(k) = P(k-1)C^T(k)\left(C(k)P(k-1)C(k)^T + R(k)\right)^{-1}$$
$$\hat{x}(k) = \hat{x}(k-1) + K(k)\left(y(k) - C(k)\hat{x}(k)\right)$$
$$P(k) = (I - K(k)C(k))P(k-1)(\quad)^T + K(k)R(k)K(k)^T$$
