The Kalman Filter

Moritz Diehl

Overview

- Recursive Least Squares Revisited
- Discrete Time Kalman Filter
- Continuous Time Kalman Filter
- Extended Kalman Filter

Comments

Recursive Least Squares (RLS) Revisited

For linear models $y_k = \phi_k^\top \theta + \epsilon_k$ and i.i.d. Gaussian noise, we had the following recursion.

- start with some a-priori knowledge on θ in form of a mean $\hat{\theta}_0$ and inverse covariance Q_0. Start loop at $k = 1$
- at time k, when measurement y_k is known, compute new inverse covariance k as

$$\boxed{Q_k = Q_{k-1} + \phi_k \phi_k^\top}$$

(adding measurements increases "information")
- compute new estimate for the mean ("innovation update")

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \underbrace{Q_k^{-1}\phi_k(y_k - \phi_k^\top \hat{\theta}_{k-1})}_{\text{"innovation"}}$$

$y(1) = y_1$

$$\underbrace{\overbrace{y_1, y_2, y_3, \ldots y_{k-1},}^{\hat{\theta}_{k-1}} y_k}_{\rightarrow \hat{\theta}_k}$$

Recursive Least Squares (RLS) Revisited

For linear models $y_k = \phi_k^\top \theta + \epsilon_k$ and i.i.d. Gaussian noise, we had the following recursion.

- start with some a-priori knowledge on θ in form of a mean $\hat{\theta}_0$ and inverse covariance Q_0. Start loop at $k = 1$
- at time k, when measurement y_k is known, compute new inverse covariance k as

$$Q_k = \alpha Q_{k-1} + \phi_k \phi_k^\top \qquad \alpha \in (0,1)$$

  (adding measurements increases "information")
- compute new estimate for the mean ("innovation update")

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \underbrace{Q_k^{-1} \phi_k (y_k - \phi_k^\top \hat{\theta}_{k-1})}_{\text{"innovation"}}$$

- delivers recursive solution to LLS problem

$$\boxed{\hat{\theta}_k = \arg\min_{\theta} \quad (\theta - \hat{\theta}_0)^\top Q_0 \, (\theta - \hat{\theta}_0) \; + \; \sum_{i=1}^{k} (y_i - \phi_k^\top \theta)^2}$$

Comments

Recursive Least Squares for State Estimation (1)

? x_0 x_1 x_k

y_1 y_2 y_k

$\hat{\theta}_k$

- Assume now that we have a known deterministic linear system

$$\boxed{x_i = A_{i-1} x_{i-1}}$$

with linear measurement equation (for $i = 1, \ldots, k$)

$$\boxed{y_i = C_i x_i + v_i}$$

(where v_i is i.i.d. Gaussian zero mean noise), and initial knowledge on initial state in form of mean $\hat{x}_0$ and inverse covariance Q_0

- Clearly, $\boxed{x_k = A_{k-1} \cdots A_0 x_0}$ and thus

$$\boxed{y_k = C_k A_{k-1} \cdots A_0 x_0 + v_k}$$

- Using $\theta \equiv x_0$, $\epsilon_k \equiv v_k$ and $\phi_k^\top = C_k A_{k-1} \cdots A_0$, this can be cast into the standard RLS framework from before

Comments

Recursive Least Squares for State Estimation (2)

- With $\theta \equiv x_0$ and

$$\phi_k^\top = C_k A_{k-1} \cdots A_0$$

our recursion becomes:

$$Q_k = Q_{k-1} + (C_k A_{k-1} \cdots A_0)^\top C_k A_{k-1} \cdots A_0$$

and

$$\hat{\theta}_k = \hat{\theta}_{k-1} + Q_k^{-1} (C_k A_{k-1} \cdots A_0)^\top (y_k - C_k A_{k-1} \cdots A_0 \hat{\theta}_{k-1})$$

- often, we are most interested in current state x_k. Let us denote its estimate given the data $(y_1, \ldots, y_m)$ by $\hat{x}_{[k|m]}$. Clearly,

$$\boxed{\hat{x}_{[k|m]} = \underline{A_{k-1} \cdots A_0} \hat{\theta}_m}$$

$\hat{x}_{[k|k]}$

and

$$\underbrace{\mathrm{Cov}\{x_{[k|m]}\}}_{=:P_{[k|m]}} = A_{k-1} \cdots A_0 \cdot Q_m^{-1} \cdot A_0^\top \cdots A_{k-1}^\top$$

Comments

Recursive Least Squares for State Estimation (3)

- the innovation update of the mean then simplifies to

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + P_{[k|k]} \cdot C_k^\top (y_k - C_k \hat{x}_{[k|k-1]})$$ 3

- the covariance update becomes (assuming for simplicity invertibility of $A_{k-1} \cdots A_0$)

$$P^{-1}_{[k|k]} = P^{-1}_{[k|k-1]} + C_k^\top C_k$$ 2

- for the propagation in time (prediction), we can use

$$\hat{x}_{[k|k-1]} = A_{k-1} \cdot \hat{x}_{[k-1|k-1]}$$

and

$$P_{[k|k-1]} = A_{k-1} \cdot P_{[k-1|k-1]} \cdot A_{k-1}^\top$$

1

Comments

Summary of Recursive Least Squares

Two steps to compute estimates $x_{[k|k]}$ and covariances $P_{[k|k]}$:

- Prediction Step (before measurement):

$$\hat{x}_{[k|k-1]} = A_{k-1} \cdot \hat{x}_{[k-1|k-1]} \tag{1}$$

$$P_{[k|k-1]} = A_{k-1} \cdot P_{[k-1|k-1]} \cdot A_{k-1}^{\top} \tag{2}$$

- Innovation Update Step (after measurement):

$$P_{[k|k]} = \left(P_{[k|k-1]}^{-1} + C_k^{\top} C_k \right)^{-1} \tag{3}$$

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + P_{[k|k]} \cdot C_k^{\top} (y_k - C_k \hat{x}_{[k|k-1]}) \tag{4}$$

- Can interpret $\hat{x}_{[k|k-1]}$ and $P_{[k|k-1]}$ as a-priori information on x_k, based on prediction model $\boxed{x_k = A_{k-1} x_{k-1}}$

Summary of Recursive Least Squares

Two steps to compute estimates $x_{[k|k]}$ and covariances $P_{[k|k]}$:

- Prediction Step (before measurement):

$$\hat{x}_{[k|k-1]} = A_{k-1} \cdot \hat{x}_{[k-1|k-1]} \tag{1}$$

$$P_{[k|k-1]} = A_{k-1} \cdot P_{[k-1|k-1]} \cdot A_{k-1}^\top \tag{2}$$

- Innovation Update Step (after measurement):

$$P_{[k|k]} = \left(P_{[k|k-1]}^{-1} + C_k^\top C_k\right)^{-1} \tag{3}$$

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + P_{[k|k]} \cdot C_k^\top (y_k - C_k \hat{x}_{[k|k-1]}) \tag{4}$$

- Can interpret $\hat{x}_{[k|k-1]}$ and $P_{[k|k-1]}$ as a-priori information on x_k, based on prediction model $\boxed{x_k = A_{k-1} x_{k-1}}$
- Could we also incorporate i.i.d. state noise w_{k-1} with zero mean and covariance W_{k-1}?

Summary of Recursive Least Squares

Two steps to compute estimates $x_{[k|k]}$ and covariances $P_{[k|k]}$:

- Prediction Step (before measurement):

$$\hat{x}_{[k|k-1]} = A_{k-1} \cdot \hat{x}_{[k-1|k-1]} \tag{1}$$

$$P_{[k|k-1]} = A_{k-1} \cdot P_{[k-1|k-1]} \cdot A_{k-1}^\top \tag{2}$$

- Innovation Update Step (after measurement):

$$P_{[k|k]} = \left(P_{[k|k-1]}^{-1} + C_k^\top C_k \right)^{-1} \tag{3}$$

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + P_{[k|k]} \cdot C_k^\top (y_k - C_k \hat{x}_{[k|k-1]}) \tag{4}$$

- Can interpret $\hat{x}_{[k|k-1]}$ and $P_{[k|k-1]}$ as a-priori information on x_k, based on prediction model $\boxed{x_k = A_{k-1} x_{k-1} + w_{k-1}}$
- Could we also incorporate i.i.d. state noise w_{k-1} with zero mean and covariance W_{k-1}?

Summary of Recursive Least Squares (and Extension)

Two steps to compute estimates $x_{[k|k]}$ and covariances $P_{[k|k]}$:

- Prediction Step (before measurement):

$$\hat{x}_{[k|k-1]} = A_{k-1} \cdot \hat{x}_{[k-1|k-1]} \tag{1}$$
$$P_{[k|k-1]} = A_{k-1} \cdot P_{[k-1|k-1]} \cdot A_{k-1}^\top + W_{k-1} \tag{2}$$

- Innovation Update Step (after measurement):

$$P_{[k|k]} = \left(P_{[k|k-1]}^{-1} + C_k^\top C_k\right)^{-1} \tag{3}$$
$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + P_{[k|k]} \cdot C_k^\top (y_k - C_k \hat{x}_{[k|k-1]}) \tag{4}$$

- Can interpret $\hat{x}_{[k|k-1]}$ and $P_{[k|k-1]}$ as a-priori information on x_k, based on prediction model $\boxed{x_k = A_{k-1}x_{k-1} + w_{k-1}}$
- Could we also incorporate i.i.d. state noise w_{k-1} with zero mean and covariance W_{k-1}?

Comments

The Kalman Filter

We have basically obtained the Kalman Filter. More generally, can assume measurement noise v_k with covariance V, such that measurements are weighted with V^{-1}.

- Full model: $\boxed{x_{k+1} = A_k x_k + w_k}$ and $\boxed{y_k = C_k x_k + v_k}$ with i.i.d. zero mean noises with covariances W, V.
- The steps of the Kalman Filter are:
  - Prediction Step:

$$\hat{x}_{[k|k-1]} = A_{k-1} \cdot \hat{x}_{[k-1|k-1]} \tag{5}$$

$$P_{[k|k-1]} = A_{k-1} \cdot P_{[k-1|k-1]} \cdot A_{k-1}^\top + W_{\cancel{k-1}} \tag{6}$$

  - Innovation Update Step:

$$P_{[k|k]} = \left(P_{[k|k-1]}^{-1} + C_k^\top \underline{V^{-1}} C_k \right)^{-1} \tag{7}$$

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + P_{[k|k]} \cdot C_k^\top \underline{V^{-1}} (y_k - C_k \hat{x}_{[k|k-1]}) \tag{8}$$

Comments

Optimization Problem solved by Kalman Filter

$$\underset{x_0, \dots x_N}{\text{minimize}} \quad (x_0 - \hat{x}_0)^T P_0^{-1} (x_0 - \hat{x}_0) + \sum_{i=1}^{N} (y_k - C_k \cdot x_k)^T V^{-1} (y_k - C_k x_k) + \sum_{k=0}^{N-1} (x_{k+1} - A_k x_k)^T W^{-1} (x_{k+1} - A_k x_k)$$

$x_0^*, \dots, x_{N-1}^*, \; x_N^* \equiv x_{[N|N]}$

Comments

The Kalman Filter in Continuous Time (1)

- Continuous time model with noises w^{c}, v^{c}:

$$\dot{x}(t) = A^{\mathrm{c}}(t)x(t) + w^{\mathrm{c}}(t)$$
$$y(t) = C^{\mathrm{c}}(t)x(t) + v^{\mathrm{c}}(t)$$

- Note that $w^{\mathrm{c}}(t)$ has unit $[x]/[t]$ and $v^{\mathrm{c}}(t)$ has unit $[y]$
- Assume white noises:
  - $\mathrm{Cov}(w^{\mathrm{c}}(t_1), w^{\mathrm{c}}(t_2)) = \delta(t_1 - t_2) \cdot W^{\mathrm{c}}$ where $[W^{\mathrm{c}}] = [x]^2/[t]$
  - $\mathrm{Cov}(v^{\mathrm{c}}(t_1), v^{\mathrm{c}}(t_2)) \ = \delta(t_1 - t_2) \cdot V^{\mathrm{c}}$ where $[V^{\mathrm{c}}] = [y]^2 \cdot [t]$
- Transfer to discrete time: use small time step Δt and time points $t_k = k \cdot \Delta t$, identify $x_k = x(t_k)$.
- Identify $w_k = \int_{t_k}^{t_{k+1}} w^{\mathrm{c}}(t)\,\mathrm{d}t$. Random walk theory says: $W = \Delta t \cdot W^{\mathrm{c}}$ (the longer we wait, the more uncertain we become)
- Identify $C_k = C(t_k)$, $y_k = \frac{1}{\Delta t}\int_{t_k}^{t_{k+1}} y(t)\,\mathrm{d}t$, and thus $v_k = \frac{1}{\Delta t}\int_{t_k}^{t_{k+1}} v^{\mathrm{c}}(t)\,\mathrm{d}t$. Due to averaging, covariance matrix shrinks with longer time intervals, i.e. $V = V^{\mathrm{c}}/Deltat$ (we give more weight to each measurement when taking fewer measurements).

Comments

The Kalman Filter in Continuous Time (2)

- Discrete time model:

$$x_{k+1} = \underbrace{[I + \Delta t \cdot A^{\mathrm{c}}(t_k)]}_{=:A_k} x_k + w_k$$

$$y_k = \underbrace{C^{\mathrm{c}}(t_k)}_{=C_k} x_k + v_k$$

- Covariances:
  $\mathrm{Cov}(w_k) = W = \Delta t \cdot W^{\mathrm{c}}$ and
  $\mathrm{Cov}(v_k) = V = \Delta t^{-1} \cdot V^{\mathrm{c}}$

Comments

The Kalman Filter in Continuous Time (3)

- Prediction step (up to first order):

$$\begin{aligned}
\hat{x}_{[k+1|k]} &= \hat{x}_{[k|k]} + \Delta t \cdot A^{\mathrm{c}}(t_k)\hat{x}_{[k|k]} \\
P_{[k+1|k]} &= P_{[k|k]} + \Delta t \left[A^{\mathrm{c}}(t_k)P_{[k|k]} + P_{[k|k]}A^{\mathrm{c}}(t_k)^\top + W^{\mathrm{c}}\right]
\end{aligned}$$

- Innovation Update Step:

$$\begin{aligned}
P_{[k|k]} &= \left(P_{[k|k-1]}^{-1} + \Delta t \cdot C_k^\top (V^{\mathrm{c}})^{-1} C_k\right)^{-1} \qquad (9) \\
\hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + \Delta t \cdot P_{[k|k]} \cdot C_k^\top (V^{\mathrm{c}})^{-1}(y_k - C_k\hat{x}_{[k|k-1]})
\end{aligned}$$

- By Taylor expansion, Eq. (9) becomes

$$P_{[k|k]} = P_{[k|k-1]} - \Delta t \cdot P_{[k|k-1]} C_k^\top (V^{\mathrm{c}})^{-1} C_k P_{[k|k-1]}$$

Comments

The Kalman Filter in Continuous Time (4)

- Adding both steps together, identifying $\hat{x}(t_k) \equiv \hat{x}_{[k+1|k]}$ and $P(t_k) \equiv P_{[k+1|k]}$, and taking the limit $\Delta t \to 0$ yields the two differential equations:

$$
\begin{aligned}
\dot{\hat{x}}(t) = &\ A^{\mathrm{c}}(t_k)\hat{x}(t) + P(t)C^{\mathrm{c}}(t)^{\top}(V^{\mathrm{c}})^{-1}\left(y(t) - C^{\mathrm{c}}(t)\hat{x}(t)\right) \\
\dot{P}(t) = &\ A^{\mathrm{c}}(t)P(t) + P(t)A^{\mathrm{c}}(t)^{\top} \\
&+ W^{\mathrm{c}} - P(t)C^{\mathrm{c}}(t)^{\top}(V^{\mathrm{c}})^{-1}C^{\mathrm{c}}(t)P(t)
\end{aligned}
$$

- These continuous time equations are called the "Kalman-Bucy-Filter"

Comments

The Extended Kalman Filter in Continuous Time

Comments

The Extended Kalman Filter in Discrete Time

Comments

Comments

Comments

Comments

Comments

Comments

Comments