Kalman Filter

Kalman Filter

$\underline{y}$: observation

$\underline{x}$: state

$\underline{n}$: random noise (measurement noise

$\underline{n} \sim \mathcal{N}(0,\underline{R})$

$$\underline{y} = h(\underline{x}) + \underline{n}$$

The estimation problem $\hat{x} = \mathscr{L}(y)$

$\hat{x}$: estimate

$\mathscr{L}$: estimator

- If the observation function $h(\cdot)$ is linear, then
- $$\underline{y} = \underline{H}\underline{x} + \underline{n}$$

  where $\underline{y}$ is a linear function of h but with some added noise.

  Then $\underline{\hat{x}} = \underline{L}\underline{y}$

  where $\underline{L} = \underline{H}^{-1}$

Over constrained System

- If $\underline{H}$ has more rows than the number of columns then there is no inverse of $\underline{H}$.
- But we can find the best $\underline{\hat{x}}$ such that the residue $|\underline{y} - \underline{H}\hat{x}|$ is as small as possible.
- The best estimator for a linear observation function happens to be a linear function.
- We denote the residue as $\underline{n} = \underline{y} - \underline{H}\underline{\hat{x}}$

- Different equations can have noise terms of different variance.
  So some equations may be more reliable.
- We formulate minimisation of the norm of $\underline{W}\underline{n}$ where $\underline{W}$ is a matrix as

$$||\underline{W}\underline{n}||^2 = \underline{n}^\top \underline{W}^\top \underline{W}\underline{n}$$

  which is equivalent to solving the system

  $\underline{W}\underline{y} = \underline{W}\underline{H}\underline{x}$

  because $\underline{W}\underline{n} = \underline{W}(\underline{y} - \underline{H}\underline{x}) \quad = \underline{W}\underline{y} - \underline{W}\underline{H}\underline{x}$

- That is we want to solve $\underline{W}\underline{y} = \underline{W}\underline{H}\underline{x}$ subject to minimising $||\underline{W}\underline{y} - \underline{W}\underline{H}\underline{x}||^2$

- Consider the standard notation for a system of equations
  $\underline{A}\underline{x} = \underline{b}$ to be solved for $\hat{\underline{x}}$ so as to minimise $||\underline{b} - \underline{A}\underline{x}||^2$
- The objective to be minimised can be re-written as $||\underline{A}\underline{x} - \underline{b}||^2$ or

$$(\underline{A}\underline{x} - \underline{b})^\top(\underline{A}\underline{x} - \underline{b}) \qquad \text{or} \qquad (\underline{x}^\top \underline{A}^\top - \underline{b}^T)(\underline{A}\underline{x} - \underline{b})$$

- Or $f \equiv (\underline{x}^\top \underline{A}^T \underline{A}\underline{x} - \underline{x}^T \underline{A}^\top \underline{b} - \underline{b}^T \underline{A}\underline{x} + \underline{b}^T \underline{b})$

$$\frac{df}{d\underline{x}} = 2\underline{A}^\top \underline{A}\underline{x} - \underline{A}^\top \underline{b} - \underline{A}^\top \underline{b} + 0$$

- $\dfrac{df}{d\underline{x}} = 2\underline{A}^\top \underline{A} * \underline{x} - 2A^\top b = 0$
- Therefore $\underline{A}^\top \underline{A}\underline{x} = \underline{A}^\top \underline{b}$ or $\quad \hat{\underline{x}} = (\underline{A}^\top A)^{-1} \underline{A}^T \underline{b}$
- The given system of equations $\underline{W}\underline{H}\underline{x} = \underline{W}\underline{y}$

  Substituting $\underline{A} \equiv \underline{W}\underline{H}$ and $\underline{b} \equiv \underline{W}\underline{y}$ in $\hat{\underline{x}} = (\underline{A}^\top A)^{-1} \underline{A}^T \underline{b}$

- $\hat{\underline{x}} = \left((\underline{W}\underline{H})^\top \underline{W}\underline{H}\right)^{-1} (\underline{W}\underline{H})^\top \underline{W}\underline{y} \quad = (\underline{H}^\top \underline{W}^\top \underline{W}\underline{H})^{-1} \underline{H}^\top \underline{W}^\top \underline{W}\underline{y}$

  $= (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1} (\underline{H}^\top \underline{R}^{-1}) \underline{y}$

- $\underline{R}$ is the covariance matrix of the measurement noise.
- Entries in $\underline{R}$ which correspond to a large variance or covariance will translate to smaller entries in the $\underline{W}$ matrix.

  $\underline{R}^{-1} = \underline{W}^\top \underline{W}$

Quality of the estimator

- The size of the covariance matrix should be small.
- Covariance matrix $\underline{P} = \mathbb{E}\left[(x-\hat{x})(x-\hat{x})^T\right]$
- A smaller norm of $\underline{P}$ implies reduced uncertainty or fluctuations
- The estimate $\hat{x}$ should be unbiased

  $\mathbb{E}\left[\underline{x}-\underline{\hat{x}}\right] = 0$
- For a linear estimator $\underline{\hat{x}} = \underline{L}\underline{y}$ corresponding to the system $\underline{y} = \underline{H}\underline{x} + \underline{n}$

  $\mathbb{E}\left[\underline{x}-\underline{\hat{x}}\right] = \mathbb{E}\left[\underline{x}-\underline{L}\underline{y}\right] = \mathbb{E}\left[\underline{x}-L(\underline{H}\underline{x}+\underline{n})\right] = \mathbb{E}\left[\left(\underline{I}-\underline{L}\underline{H}\right)\underline{x}\right] - \mathbb{E}\left[\underline{L}\underline{n}\right] = 0$

- $\mathbb{E}\left[\left(\underline{I}-\underline{L}\underline{H}\right)\underline{x}\right] = 0$

  or $\left(\underline{I}-\underline{L}\underline{H}\right)\mathbb{E}\left[\underline{x}\right] = 0$

  $\therefore \underline{I} = \underline{L}\underline{H}$ or $\underline{L}\underline{H} = \underline{I}$
- $\mathbb{E}\left[\underline{\hat{x}}\right] = \underline{x}$

- The covariance matrix of our estimate

  $\underline{P} = \mathbb{E}\left[(\underline{x}-\underline{\hat{x}})(\underline{x}-\underline{\hat{x}})^\top\right]$

  $= \mathbb{E}\left[(\underline{x}-\underline{L}\underline{y})(\underline{x}-\underline{L}\underline{y})^\top\right]$
- Substituting $\underline{y} = \underline{H}\underline{x} + \underline{n}$ we have

  $= \mathbb{E}\left[(\underline{x}-\underline{L}\underline{H}\underline{x}-\underline{L}\underline{n})\ (\underline{x}-\underline{L}\underline{H}\underline{x}-\underline{L}\underline{n})^\top\right]$

  $= \mathbb{E}\left[\left[(\underline{I}-\underline{L}\underline{H})\underline{x}-\underline{L}\underline{n}\right]\ \left[(\underline{I}-\underline{L}\underline{H})\underline{x}-\underline{L}\underline{n}\right]^\top\right] = \mathbb{E}\left[(\underline{L}\underline{n})\ (\underline{L}\underline{n}]^\top\right]$

- $= \mathbb{E}\left[\underline{L}\underline{n}\underline{n}^\top\underline{L}\right]$

  $= \underline{L}\ \mathbb{E}\left[\underline{n}\underline{n}^\top\right]\ \underline{L}$

  $= \underline{L}\ \underline{R}\ \underline{L}^\top$

  We define $P \equiv \underline{L}\underline{R}\underline{L}^\top$
- Here $\underline{R}$ is the covariance matrix of the measurement noise.
- Substituting for the estimator $\underline{L} = (\underline{H}^\top\underline{R}^{-1}\underline{H})^{-1}\ \underline{H}^\top\underline{R}^{-1}$ we get

- $P \equiv \underline{L}\,\underline{R}\,\underline{L}^\top$
- $\underline{L} = (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1}\, \underline{H}^\top \underline{R}^{-1}$

- $\underline{P} = (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1}\, \underline{H}^\top \underline{R}^{-1} \;\; \underline{R} \;\; \underline{R}^{-1} \underline{H}\, (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1}$

  $= (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1} \;\; \underline{H}^\top \underline{R}^{-1} \underline{H} \;\; (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1}$

  $= (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1}$

- Here we have used the result that the inverse of a symmetric and invertible matrix will also be symmetric.

  That is, given $A^\top = A$ and A^{-1} exists we have

  $A^{-1}A = I$

  Taking transpose of both sides $(A^{-1}A)^\top = A^\top (A^{-1})^\top = I^\top = I$

  Premultiplying both sides by A^{-1}

  $A^{-1}A\,(A^{-1})^\top = A^{-1}\,I$

  $I(A^{-1})^\top = A^{-1}I$ or $(A^{-1})^\top = A^{-1}$ Therefore A^{-1} is symmetric.

Claim

- The estimator $\underline{L} = (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1}\, \underline{H}^\top \underline{R}^{-1}$ has the minimum norm covariance, given by $P \equiv \underline{L}\,\underline{R}\,\underline{L}^\top$. We denote this optimal estimator as L_0

  $\underline{L}_0 = \underline{P}\,\underline{H}^\top \underline{R}^{-1}$

  Proof:

  We know that $\underline{L}_0 = (\underline{H}^\top \underline{R}^{-1} \underline{H})^{-1}\, \underline{H}^\top \underline{R}^{-1}$ is the solution to the normal equations $\underline{W}\,\underline{H}\,\underline{x} = \underline{W}\,\underline{y}$

- The covariance matrix for estimator $\underline{L}_0$ is $\underline{L}_0 \underline{R}\, \underline{L}_0^\top \equiv \underline{P}$
- $\underline{L}_0$ is an unbiased estimator.

  So $\underline{L}_0 \underline{H} = \underline{I}$
- Now consider an alternative unbiased estimator $\underline{L}$
- We trivially write $\underline{L} = \underline{L}_0 + (\underline{L} - \underline{L}_0)$
- The covariance matrix for $\underline{L}$ is given by

  $\underline{P} = \underline{L}\,\underline{R}\,\underline{L}^\top$

- $\underline{P} = \underline{L}\,\underline{R}\,\underline{L}^\top$

  $= (\underline{L}_0 + (\underline{L} - \underline{L}_0))\ \underline{R}\ (\underline{L}_0 + (\underline{L} - \underline{L}_0))^\top$

  $= \underline{L}_0\underline{R}\underline{L}_0^\top + (\underline{L} - \underline{L}_0)\underline{R}\underline{L}_0^\top + \underline{L}_0\underline{R}(\underline{L} - \underline{L}_0)^\top + (\underline{L} - \underline{L}_0)\underline{R}(\underline{L} - \underline{L}_0)^\top$

- $\underline{L}_0^\top = (\underline{H}^\top R^{-1})^\top\ (\underline{H}^\top R^{-1}\underline{H})^{-1^\top}$

  $= R^{-1^\top}\underline{H}\ (\underline{H}^\top R^{-1}\underline{H})^{-1}$

  $= R^{-1}\underline{H}\ (\underline{H}^\top R^{-1}\underline{H})^{-1}$

- $\underline{R}\underline{L}_0^\top = \underline{R}R^{-1}\underline{H}\ (\underline{H}^\top R^{-1}\underline{H})^{-1} = \underline{H}\ (\underline{H}^\top R^{-1}\underline{H})^{-1}$

- $(\underline{L} - \underline{L}_0)\,\underline{R}\underline{L}_0^\top = (\underline{L} - \underline{L}_0)\ \underline{H}\ (\underline{H}^\top R^{-1}\underline{H})^{-1}$

  $= (\underline{L}\underline{H} - \underline{L}_0\underline{H})\ \ (\underline{H}^\top R^{-1}\underline{H})^{-1}$

- Since L and L_0 are unbiased estimators

  $\underline{L}\underline{H} - \underline{L}_0\underline{H} = \underline{I}$

- $\therefore\ \ (\underline{L} - \underline{L}_0)\,\underline{R}\underline{L}_0^\top = 0$

- Since $(\underline{L} - \underline{L}_0)\underline{R}\underline{L}_0^\top = 0$

  $\left((\underline{L} - \underline{L}_0)\underline{R}\underline{L}_0^\top\right)^\top = 0$

  or $\left((\underline{L} - \underline{L}_0)\underline{R}\underline{L}_0^\top\right)^\top = 0$

  $\underline{L}_0\underline{R}\,(\underline{L} - \underline{L}_0)^\top = 0$

- $\underline{P} = \underline{L}_0\underline{R}\underline{L}_0^\top + (\underline{L} - \underline{L}_0)\underline{R}\underline{L}_0^\top + \underline{L}_0\underline{R}(\underline{L} - \underline{L}_0)^\top + (\underline{L} - \underline{L}_0)\underline{R}(\underline{L} - \underline{L}_0)^\top$

  $= \underline{L}_0\underline{R}\underline{L}_0^\top + (\underline{L} - \underline{L}_0)\underline{R}(\underline{L} - \underline{L}_0)^\top$

- $\underline{P} = \underline{L}_0\underline{R}\underline{L}_0^\top + (\underline{L} - \underline{L}_0)\underline{R}(\underline{L} - \underline{L}_0)^\top$
- This is the sum of two positive semi-definite matrices.
- For such matrices, the norm of the sum is greater than or equal to either norm.
- The norm of $\underline{P}$ will be minimised when the 2nd term on the RHS vanishes when $L = L_0$
- Therefore, the estimator $\underline{L}_0 = (\underline{H}^\top \underline{R}^{-1}\underline{H})^{-1}\ \underline{H}^\top\ \underline{R}^{-1}$ has the minimum covariance, given by $\underline{P} = (\underline{H}^\top R^{-1}\underline{H})^{-1}$

- In fact, $\underline{L}_0 = \underline{P}\underline{H}^\top \underline{R}^{-1}$ in terms of $\underline{P}$
- So far we have seen how to estimate the state $\hat{\underline{x}}$ given the measurement equation.

$$\underline{y} = \underline{H}\underline{x} + \underline{n} \qquad \underline{n} \sim \mathcal{N}(0,\mathbb{R}))$$

$$\hat{\underline{x}} = \underline{P}\underline{H}^\top \underline{R}^{-1}\underline{y}$$

$$= \underline{L}\underline{y}$$

Dynamic System

- In a linear dynamic system, the states update from one time tick to another.

State Update: $\underline{x}_{k+1} = \underline{F}_k \underline{x}_k + \underline{G}_k \underline{u}_k + \underline{\eta}_k$

Measurement: $\underline{y}_k = \underline{H}_k \underline{x}_k + \underline{\xi}_k$

- $\underline{\eta}_k \sim \mathcal{N}(0, \underline{Q}_k)$ is the system noise

$\underline{\xi}_k \sim \mathcal{N}(0, \underline{R}_k)$ is the measurement noise

- At every time tick there are 2 evidences which determine the current state.
  - Contribution (evidence) from the previous state, through state update
  - Evidence from the measurement made at the current time tick. (state update from observing the evidence)
- Notation

$\hat{\underline{x}}_{k|k-1}$ state estimate at time k given the measurements up to time $k-1$

$\hat{\underline{x}}_{k|k}$ state estimate after taking account the measurement up to time tick k

$\hat{\underline{y}}_k$ predicted observation at time tick k

Update Step

- There are two sources that provide estimate to the state $\underline{\hat{x}}_k$
    1. Estimate from the state propagation $\underline{\hat{x}}_{k|k-1}$

       $$\underline{\hat{x}}_{k|k-1} = \underline{x}_k + \underline{e}_k$$

       The estimate $\underline{\hat{x}}_{k|k-1}$ is considered to have a random deviation or error from the true state $\underline{x}_k$

       The covariance $P_{k|k-1}$ of the error term summarises the uncertainty of prediction $\underline{\hat{x}}_{k|k-1}$ given the past history of measurements.

       $$P_{k|k-1} = \mathbb{E}[\underline{e}_k \underline{e}_k^\top]$$

- There are two sources that provide estimate to the state $\underline{\hat{x}}_k$
    2. Evidence from the new measurement

       $$\underline{y}_k = \underline{H}_k \underline{x}_k + \underline{\xi}_k$$

       $$R_k = \mathbb{E}[\underline{\xi}_k \underline{\xi}_k^\top]$$

- Given the evidences about the unknown true state $\underline{x}_k$ our task is to compute an estimate $\underline{\hat{x}}_k$ of $\underline{x}_k$
- We collect the two evidences

  $$\underline{\hat{x}}_{k|k-1} = \underline{x}_k + \underline{e}_k$$

  $$\underline{y}_k = \underline{H}_k \underline{x}_k + \underline{\xi}_k$$

  to form a system of equations

  $$\underline{y} = \underline{H} \underline{x}_k + \underline{n}$$

- $\underline{y} = \underline{H}\underline{x}_k + \underline{n}$

$$\underline{y} = \begin{bmatrix} \underline{\hat{x}}_{k|k-1} \\ \underline{y}_k \end{bmatrix} \qquad \underline{H} = \begin{bmatrix} \underline{I} \\ \underline{H}_k \end{bmatrix} \qquad \underline{n} = \begin{bmatrix} \underline{e}_k \\ \underline{n}_k \end{bmatrix}$$

- $$\hat{x}_{k|k-1} = \underline{I}\underline{x}_k + \underline{e}_k$$
$$\underline{y}_k = \underline{H}_k\underline{x}_k + \underline{\xi}_k$$

- The covariance matrix of the vector $\underline{n} = \begin{bmatrix} \underline{e}_k \\ \underline{\xi}_k \end{bmatrix}$ is formulated as

$$R = \begin{bmatrix} \underline{P}_{k|k-1} & 0 \\ 0 & \underline{R}_k \end{bmatrix}$$

Here we assume that the two noise vectors are independent.

- The formulation $\underline{y} = \underline{H}\underline{x}_k + \underline{n}$ is a classical estimation problem.

- The solution to the estimation problem $\underline{y} = \underline{H}_k\underline{x}_k + \underline{n}$ is

$$\underline{\hat{x}}_{k|k} = \underline{P}_{k|k}\underline{H}^\top \underline{R}^{-1}\underline{y}$$
$$\underline{P}_{k|k} = (\underline{H}^\top \underline{R}^{-1}\underline{H})^{-1}$$

This is the update stage of the Kalman Filter

- Simplification

$$\underline{P}_{k|k}^{-1} = \underline{H}^\top \underline{R}^{-1}\underline{H}$$
$$= \begin{bmatrix} \underline{I} & \underline{H}_k^\top \end{bmatrix} \begin{bmatrix} \underline{P}_{k|k-1}^{-1} & 0 \\ 0 & \underline{R}_k^{-1} \end{bmatrix} \begin{bmatrix} \underline{I} \\ \underline{H}_k \end{bmatrix}$$
$$= \underline{P}_{k|k-1}^{-1} + \underline{H}_k^\top \underline{R}_k^{-1}\underline{H}_k$$

Simplifying the posterior estimate

- $\underline{\hat{x}}_{k|k} = \underline{P}_{k|k}\underline{H}^\top \underline{R}^{-1}\underline{y}$

$$= \underline{P}_{k|k} \begin{bmatrix} \underline{I} & \underline{H}_k^\top \end{bmatrix} \begin{bmatrix} \underline{P}_{k|k-1}^{-1} & 0 \\ 0 & \underline{R}_k^{-1} \end{bmatrix} \begin{bmatrix} \underline{\hat{x}}_{k|k-1} \\ \underline{y}_k \end{bmatrix}$$

$$= \underline{P}_{k|k} \begin{bmatrix} \underline{P}_{k|k-1}^{-1} & \underline{H}_k^\top \underline{R}_k^{-1} \end{bmatrix} \begin{bmatrix} \underline{\hat{x}}_{k|k-1} \\ \underline{y}_k \end{bmatrix}$$

$$= \underline{P}_{k|k} \left(\underline{P}_{k|k-1}^{-1}\underline{\hat{x}}_{k|k-1} + \underline{H}_k^\top \underline{R}_k^{-1}\underline{y}_k \right)$$

- $$\underline{\hat{x}}_{k|k} = \underline{P}_{k|k}\underline{H}^\top \underline{R}^{-1}\underline{y}$$

$$= \underline{P}_{k|k}\left(\left(\underline{P}_{k|k}^{-1} - \underline{H}_k^\top \underline{R}_k^{-1}\underline{H}_k\right)\underline{\hat{x}}_{k|k-1} + \underline{H}_k^\top \underline{R}_k^{-1}\underline{y}_k\right)$$

$$= \left(\underline{I} - \underline{P}_{k|k}\underline{H}_k^\top \underline{R}_k^{-1}\underline{H}^k\right)\underline{\hat{x}}_{k|k-1} + \underline{P}_{k|k}\underline{H}_k^\top \underline{R}_k^{-1}\underline{y}_k$$

$$= \underline{\hat{x}}_{k|k-1} - \underline{P}_{k|k}\underline{H}_k^\top \underline{R}_k^{-1}\underline{H}^k\underline{\hat{x}}_{k|k-1} + \underline{P}_{k|k}\underline{H}_k^\top \underline{R}_k^{-1}\underline{y}_k$$

$$= \underline{\hat{x}}_{k|k-1} + \left(\underline{P}_{k|k}\underline{H}_k^\top \underline{R}_k^{-1}\right)\left(\underline{y}_k - \underline{H}_k\underline{\hat{x}}_{k|k-1}\right)$$