Lecture 8 - Kalman Filter & LQG

- Minimum Mean Square Estimation (MMSE)
- Relation to Least Squares
- Kalman Filter (KF)
- LQG = LQR + KF

Gaussian Random Vectors

A random vector $x \in \mathbb{R}^n$ is Gaussian distributed, denoted by $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$

- where $\bar{x} \in \mathbb{R}^n$ is mean of x or expected value of x.

$$\bar{x} = \mathbb{E}[x] = \int v \, p_x(v) \, dv$$

$\Sigma_x = \Sigma_x^T > 0$ is covariance of x

$$\Sigma_x = \mathbb{E}\left[(x-\bar{x})(x-\bar{x})^T\right] = \mathbb{E}[xx^T] - \bar{x}\bar{x}^T = \int (v-\bar{x})(v-\bar{x})^T p_x(v) \, dv$$

$p_x(v)$ is the probability density function of x

Normal distribution density $\Big\}$ $p_x(v) = (2\pi)^{-\frac{n}{2}} (\det \Sigma_x)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v-\bar{x})^T \Sigma_x^{-1} (v-\bar{x})\right)$.

FACT: $w \sim \mathcal{N}(0, I) \Rightarrow x = \Sigma_x^{1/2} w + \bar{x}$ is $\mathcal{N}(\bar{x}, \Sigma_x)$. (useful for simulating vectors with given mean $\bar{x}$ and covariance Σ_x)

If $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$, then $z = \Sigma_x^{-\frac{1}{2}}(x-\bar{x})$ is $\mathcal{N}(0, I)$ called Whitening or normalizing.

Affine Transformation: Suppose that $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$. Consider the affine transformation $\boxed{z = Ax+b}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

$$\Rightarrow \boxed{z \sim \mathcal{N}(A\bar{x}+b,\ A\Sigma_x A^T)}.$$

Proof: $\bar{z} = \mathbb{E}[z] = \mathbb{E}[Ax+b] = A\,\mathbb{E}[x] + b = A\bar{x}+b.$

$$\Sigma_z = \mathbb{E}\left[(z-\bar{z})(z-\bar{z})^T\right] = \mathbb{E}\left[A(x-\bar{x})(x-\bar{x})^T A^T\right] = A\Sigma_x A^T.$$

Linear Measurements: Consider $y = Ax + v$

$x \in \mathbb{R}^n$ – what we want to estimate, $x \sim \mathcal{N}(\bar{x}, \Sigma_x) \Rightarrow$ initial uncertainty about x

$y \in \mathbb{R}^p$ – measurement from sensor $\boxed{x \text{ and } v \text{ are independent}}$

$A \in \mathbb{R}^{p \times n}$ – sensor selection matrix

$v \in \mathbb{R}^p$ – sensor noise, $v \sim \mathcal{N}(\bar{v}, \Sigma_v)$

$$\Rightarrow \begin{bmatrix} x \\ v \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{x} \\ \bar{v} \end{bmatrix}, \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_v \end{bmatrix}\right)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} I & 0 \\ A & I \end{bmatrix}\begin{bmatrix} x \\ v \end{bmatrix} \Rightarrow \mathbb{E}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \bar{x} \\ A\bar{x}+\bar{v} \end{bmatrix} \Rightarrow \mathbb{E}\left[\begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}\begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}^T\right] = \begin{bmatrix} \Sigma_x & \Sigma_x A^T \\ A\Sigma_x & \underbrace{A\Sigma_x A^T}_{\text{signal cov.}} + \underbrace{\Sigma_v}_{\text{noise cov}} \end{bmatrix}$$

MMSE Problem: Suppose $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^p$ are random vectors (not necessarily Gaussian). Estimate x given y. That is, seek a function $\phi: \mathbb{R}^p \to \mathbb{R}^n$ such that $\boxed{\hat{x} = \phi(y)}$ and $\hat{x}$ should be near x.

(MSE)
Mean Square Error: A common measure of nearness $\mathbb{E}\left[\|\phi(y)-x\|^2\right]$

MMSE: ϕ_{mmse} minimizes MSE, and $\phi_{mmse}(y) = \mathbb{E}[x|y]$ = conditional expectation of x given y.

Suppose that $\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}\right)$, then conditional density is

$$p_{x|y}(v|y) = (2\pi)^{-\frac{n}{2}} (\det \Lambda)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v-w)^T \Lambda^{-1}(v-w)\right), \text{ where}$$

$$w = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y-\bar{y})$$

$$\Lambda = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx} = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T$$

$\mathbb{E}[\hat{x}] = \mathbb{E}[x]$

Best linear unbiased estimator

$$\Rightarrow \hat{x} = \phi_{mmse}(y) = \mathbb{E}(x|y) = w = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y-\bar{y}).$$ (affine function of y output)

$$\Rightarrow \text{MMSE Estimation Error} = \underbrace{\hat{x}-x}_{:=e} \sim \mathcal{N}(0, \underbrace{\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T}_{\Sigma_e})$$

Fact: $\Sigma_e = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T \leq \Sigma_x$

MMSE with Linear Measurements: $y = Ax + v$, $x \sim \mathcal{N}(\bar{x}, \Sigma_x)$, $v \sim \mathcal{N}(\bar{v}, \Sigma_v)$
$\Rightarrow \bar{y} = A\bar{x} + \bar{v}$
x, v independent.

then, $\hat{x} = \phi_{mmse}(y) = \bar{x} + B(y-\bar{y})$, where $B = \Sigma_x A^T (A\Sigma_x A^T + \Sigma_v)^{-1}$

$\bar{x}$: best prior guess of x before measurement y.
$y-\bar{y}$: residual = actual o/p − expected o/p

B: Gain that modifies $\bar{x}$ given y to give us $\hat{x}$.
Σ_v large $\Rightarrow$ very noisy o/p $\Rightarrow$ B is small $\Rightarrow$ $\hat{x} \approx \bar{x}$ + small

MMSE Estimation Error : $\tilde{x} = \hat{x} - x$ and $\tilde{x} \sim N\left(0, \underbrace{\Sigma_x - \Sigma_x A^T (A\Sigma_x A^T + \Sigma_v)^{-1} A\Sigma_x}_{\Sigma_e}\right)$.

Facts: $\Sigma_e \leq \Sigma_x \Rightarrow$ measurement always decreases uncertainty about x.

$\Sigma_x - \Sigma_e$: Value of measurement y in estimating x.

Recall: $\Sigma_e = \Sigma_x - \Sigma_x A^T (A\Sigma_x A^T + \Sigma_v)^{-1} A\Sigma_x$

$\Rightarrow \Sigma_e$ can be estimated even before measurement y is made, as we have all information available beforehand $\{A, \Sigma_x, \Sigma_v\}$

Linear System driven by Stochastic Noise

Consider DT-LTI system $x_{t+1} = Ax_t + Gw_t$, x_0 is random. we know $\bar{x}_0$ and Σ_{x_0}.

$w_t \rightarrow$ called process noise or disturbance.

We know $w_0, w_1, \ldots$ are random as well and we know $\bar{w}_t$ and Σ_{w_t}.

Denote: $\bar{x}_t = \mathbb{E}[x_t]$, $\Sigma_{x_t} = \mathbb{E}[(x_t - \bar{x}_t)(x_t - \bar{x}_t)^T]$.

$\Rightarrow x_{t+1} = Ax_t + Gw_t \Rightarrow \bar{x}_{t+1} = A\bar{x}_t + G\bar{w}_t$

Mean propagates via the same linear dynamical system (A, G).

$\Rightarrow x_{t+1} - \bar{x}_{t+1} = A(x_t - \bar{x}_t) + G(w_t - \bar{w}_t)$

$\Rightarrow \mathbb{E}\left[(x_{t+1} - \bar{x}_{t+1})(x_{t+1} - \bar{x}_{t+1})^T\right] = \Sigma_{x_{t+1}} = A\Sigma_{x_t} A^T + G\Sigma_{w_t} G^T + A\Sigma_{xw_t} G^T + G\Sigma_{wx_t} A^T$.

"Lyapunov-like dyn. system driven by Σ_{xw} and Σ_w"

If x and w are uncorrelated, i.e. $\Sigma_{xw_t} = \Sigma^T_{wx_t} = 0$. Then

$$\Sigma_{x_{t+1}} = A\Sigma_{x_t}A^T + G\Sigma_{w_t}G^T,$$ and $\Sigma_{x_{t+1}}$ stable iff A is stable.

If A is stable and $\Sigma_{w_t} = \Sigma_w, \forall t \in \mathbb{N}$, then $\Sigma_{x_t} \to \Sigma_x$ steady state covariance

and Σ_x satisfies Lyapunov eqn $\Sigma_x = A\Sigma_x A^T + G\Sigma_w G^T$.

Linear Gauss - Markov Model

Consider
$$x_{t+1} = Ax_t + w_t, \quad x \in \mathbb{R}^n \to \text{state}, \; w_t \in \mathbb{R}^n \to \text{process noise}$$
$$y_t = Cx_t + v_t, \quad y \in \mathbb{R}^p \to \text{output}, \; v_t \in \mathbb{R}^p \to \text{sensor noise}.$$

Assumptions: $x_0 \sim \mathcal{N}(\bar{x}_0, \Sigma_{x_0})$, $w_t \sim \mathcal{N}(0, \Sigma_w)$, $v_t \sim \mathcal{N}(0, \Sigma_v)$

$\left.\begin{matrix}(x_0, w_t) \\ (x_0, v_t) \\ (w_t, v_t) \\ (w_t, y_t)\end{matrix}\right\}$ independent.

Stacked Notations: $X_t := (x_0, \dots, x_t)$, $W_t := (w_0, \dots, w_t)$

$Y_t := (y_0, \dots, y_t)$, $V_t := (v_0, \dots, v_t)$.

Markov Property: We say a process x is Markov if $x_t \mid x_0, \dots, x_{t-1} = x_t \mid x_{t-1}$

If you know x_{t-1}, then knowledge of $x_{t-2}, \dots, x_0$ don't give more info about x_t.

Uncertainty Propagation: $\bar{x}_{t+1} = A\bar{x}_t \Rightarrow \bar{x}_t = A^t \bar{x}_0 \Rightarrow \bar{x}_{t+1} = A^{t+1}\bar{x}_0$.

Lyap. eqn

$\Sigma_{x_{t+1}} = A\Sigma_{x_t}A^T + \Sigma_w$. If A stable $\Rightarrow \Sigma_{x_t} \to \Sigma_x$ & $\boxed{\Sigma_x = A\Sigma_x A^T + \Sigma_w}$

Conditioning on Observed Output

Conditioning Notation: $\hat{x}_{t|s} = \mathbb{E}[x_t \mid y_0, \ldots, y_s]$ ⇒ MMSE estimate of x_t based on $y_0, \ldots, y_s$

$$\Sigma_{x_{t|s}} = \mathbb{E}[(x_t - \hat{x}_{t|s})(x_t - \hat{x}_{t|s})]$$

The random variable $x_t \mid y_0, \ldots, y_s \sim \mathcal{N}(\hat{x}_{t|s}, \Sigma_{x_{t|s}})$

State Estimation: 1) find $\hat{x}_{t|t}$
2.) find $\hat{x}_{t+1|t}$

Measurement Update: 1.) Find $\hat{x}_{t|t}$ in terms of $\hat{x}_{t|t-1}$
2.) Find $\Sigma_{x_{t|t}}$ in terms of $\Sigma_{x_{t|t-1}}$

Since x_t, Y_t are jointly Gaussian, we see

$$\hat{x}_{t|t} = \bar{x}_t + \Sigma_{x_t Y_t} \Sigma_{Y_t}^{-1} (Y_t - \bar{Y}_t). \longrightarrow (A)$$

Problem: $\Sigma_{Y_t}^{-1} \in \mathbb{R}^{pt \times pt}$ grows with time t.

Solution: KF computes $\hat{x}_{t|t}$ recursively ☺

Take linear measurement & condition it on Y_{t-1}.

$$\Rightarrow y_t = Cx_t + v_t \Rightarrow \underbrace{y_t \mid Y_{t-1}}_{\Rightarrow \text{also Gaussian}} = C\underbrace{x_t \mid Y_{t-1}}_{\text{Gaussian}} + v_t \mid Y_{t-1} = Cx_t \mid Y_{t-1} + v_t$$

$\rightarrow$ indep of Y_{t-1}

$$\Rightarrow \begin{bmatrix} x_t \mid Y_{t-1} \\ y_t \mid Y_{t-1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \hat{x}_{t|t-1} \\ C\hat{x}_{t|t-1} \end{bmatrix}, \begin{bmatrix} \Sigma_{x_{t|t-1}} & \Sigma_{x_{t|t-1}} C^T \\ C\Sigma_{x_{t|t-1}} & C\Sigma_{x_{t|t-1}} C^T + \Sigma_v \end{bmatrix} \right)$$

Recall (A) to get mean & covariance of random variable $(x_t \mid Y_{t-1}) \mid (y_t \mid Y_{t-1})$ which is exactly the same as $x_t \mid Y_t$.

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \underbrace{\Sigma_{x_{t|t-1}} C^T \left(C\Sigma_{x_{t|t-1}} C^T + \Sigma_v\right)^{-1}}_{:= \text{Kalman Gain (Almost) not yet!!!}} (y_t - C\hat{x}_{t|t-1})$$

can be computed before any observations are made.

$$\Sigma_{x_{t|t}} = \Sigma_{x_{t|t-1}} - \Sigma_{x_{t|t-1}} C^T \left(C\Sigma_{x_{t|t-1}} C^T + \Sigma_v\right)^{-1} C\Sigma_{x_{t|t-1}}$$

(★) measurement update ⇒ because it updates $\hat{x}_{t|t-1}$ & $\Sigma_{x_{t|t-1}}$ based on measurement y_t.

Time Update:

$x_{t+1} = Ax_t + w_t$. Increment the time from $t-1$ to t and condition on Y_t.

$\Rightarrow x_{t+1}|Y_t = A\, x_t|Y_t + w_t|Y_t = A\, x_t|Y_t + w_t$. → ind. of Y_t

(★★)

$$\boxed{\Rightarrow \hat{x}_{t+1|t} = A\hat{x}_{t|t} \text{ and } \Sigma_{x_{t+1|t}} = \mathbb{E}\left[(\hat{x}_{t+1|t} - x_{t+1})(\hat{x}_{t+1|t} - x_{t+1})^T\right] = A\Sigma_{x_{t|t}}A^T + \Sigma_w}$$

Kalman Filter

Apply Measurement update & Time update together to get a recursive solution.

Start: Given prior mean & covariance $\hat{x}_{0|-1} = \bar{x}_0$, $\Sigma_{x_{0|-1}} = \Sigma_0$

Step 1: Find $\hat{x}_{0|0}$ and $\Sigma_{x_{0|0}}$ by applying (★) on prior given info.

Step 2: Apply time update (★★) to get $\hat{x}_{1|0}$ and $\Sigma_{x_{1|0}}$. ☺

Step 3: Go to step 1 .....

Fuse measurement & time update into 1 step to get Filter ARE (FARE) → runs fwd in time.

Riccati Recursion given $\Sigma_{x_{0|-1}} = \Sigma_0$:

$$\Sigma_{x_{t+1|t}} = A\Sigma_{x_{t|t-1}}A^T + \Sigma_w - A\Sigma_{x_{t|t-1}} C^T\left(C\Sigma_{x_{t|t-1}}C^T + \Sigma_v\right)^{-1} C\Sigma_{x_{t|t-1}} A^T$$

Kalman Filter as an Observer

One Step KF: $$\hat{x}_{t+1|t} = \underbrace{A\hat{x}_{t|t-1}}_{\substack{\text{prediction}\\ \text{of } x_{t+1} \text{ based}\\ \text{on } Y_{t-1}}} + \underbrace{A\Sigma_{x_{t|t-1}} C^T\left(C\Sigma_{x_{t|t-1}}C^T + \Sigma_v\right)^{-1}}_{:= L_t \text{ called as Kalman Gain.}}\left(y_t - \overbrace{C\hat{x}_{t|t-1}}^{\hat{y}_{t|t-1}}\right)$$

Steady-State Kalman Filter: Just like LQR, $\underbrace{\Sigma_{x_{t|t-1}}}_{\substack{\text{converges even}\\ \text{if } A \text{ is unstable}}} \to \hat{\Sigma}_x$ if (A,C) observable & (A,Σ_w) controllable.

$\hat{\Sigma}_x$ satisfies steady state FARE, $\boxed{\hat{\Sigma}_x = A\hat{\Sigma}_x A^T + \Sigma_w - A\hat{\Sigma}_x C^T(C\hat{\Sigma}_x C^T + \Sigma_v)^{-1}C\hat{\Sigma}_x A^T}$

$\Rightarrow$ Steady-state Kalman Gain $L_{ss} = A\hat{\Sigma}_x C^T(C\hat{\Sigma}_x C^T + \Sigma_v)^{-1}$

$\Rightarrow$ Estimation error is $\tilde{x}_{t|t-1} = x_t - \hat{x}_{t|t-1}$

$$\Rightarrow y_t - \hat{y}_{t|t-1} = Cx_t + v_t - C\hat{x}_{t|t-1} = C\tilde{x}_{t|t-1} + v_t.$$

$\Rightarrow$ Estimation error evolves as

$$\tilde{x}_{t+1|t} = x_{t+1} - \hat{x}_{t+1|t} = Ax_t + w_t - A\hat{x}_{t|t-1} - L(C\tilde{x}_{t|t-1} + v_t)$$

$$= \underbrace{(A-LC)}_{\substack{\text{will be stable}\\ \text{if } (A,W) \text{ controllable.}\\ (A,C) \text{ observable.}}}\tilde{x}_{t|t-1} + \underbrace{w_t - Lv_t}_{:= \bar{v}_t}.$$

$\Rightarrow \mathbb{E}[\bar{v}_t] = 0,\ \mathbb{E}[\bar{v}_t\bar{v}_t^T] = \Sigma_w + L\Sigma_v L^T$

FACTS:

→ KF Riccati Recursion is same as LQR recursion with $(A^T, C^T, \Sigma_w, \Sigma_v)$ replacing (A, B, Q, R) in LQR recursion.

→ This is why KF is the dual of LQR.
- ↗ LQR Riccati Recursion goes backward in time
- ↘ KF Riccati Recursion goes forward in time.

→ KF Riccati Recursion is independent of input u_t.
⇒ Gave Rise to separation principle in optimal control.

LQG: (LQR + KF) $\min J = \mathbb{E}\left[\sum_{t=0}^{T-1} (x_t^T Q x_t + u_t^T R u_t) + x_T^T Q x_T\right]$ can also be extended to ∞ horizon.

control policy u_t is allowed to depend on Y_t and not on x_t.

$$\mathbb{E}[x_t^T Q x_t] = Tr[Q\, \mathbb{E}[x_t x_t^T]] = Tr(Q\, \mathbb{E}[\mathbb{E}[x_t x_t^T \mid Y_t]]) = Tr(Q\, \mathbb{E}[\Sigma_{x_{t|t}} + \hat{x}_{t|t} \hat{x}_{t|t}^T])$$
$$= Tr(Q\, \mathbb{E}[\Sigma_{x_{t|t}}]) + \mathbb{E}[\hat{x}_{t|t}^T Q \hat{x}_{t|t}].$$

$$\Rightarrow J = Tr\left(Q\, \mathbb{E}\left[\sum_{t=1}^{T} \underbrace{\Sigma_{x_{t|t}}}_{\substack{\text{not a fn. of } u_t \\ \Rightarrow \text{same for all } u_t}}\right]\right) + \mathbb{E}\left[\sum_{t=1}^{T} \hat{x}_{t|t}^T Q \hat{x}_{t|t} + u_t^T R u_t + \hat{x}_{T|T}^T Q \hat{x}_{T|T}\right]$$

Recall: $\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t \Rightarrow$ use $u_t = K_t \hat{x}_{t|t}$, where $K_t = lqr(A, B, Q, R)$.

$$LQG = \underbrace{LQR}_{\substack{\text{use LQR} \\ \text{Gain } k \text{ to} \\ \text{control}}} + \underbrace{KF}_{\substack{\text{use KF} \\ \text{Gain } L \text{ to} \\ \text{estimate}}}.$$

FACTS:

→ LQR → robust ☺

→ KF → dual of LQR ⇒ also robust ☺

→ LQG = LQR + KF → not robust ☹
See Famous Paper by John Doyle.

→ In some special cases, can recover robustness properties → LOOP TRANSFER RECOVERY (LTR).
(study for yourself !!!)