10-708 12/06/2020

L10-Review

(also ½ S.2019 L11) - clear up doubts, ambiguities.

- Kalman filtering (Jordan 2003) Ch15 15.4

- model: SSM
- inference algo: KF

(*) inference problem for SSM is exactly the same as it was for HMMs.
- i.e. calculate posterior probability of latent states given output sequence

(*) we wish to calculate $p(x_t | y_0, y_1, \ldots, y_t)$ (lecture slides use $p(x_t | y_1, \ldots, y_t)$)

(*) inference in SSM → find a recursion linking these conditional means
- & variances at neighbouring moments in time.

(*) As Gaussians; charact. by mean, cov.

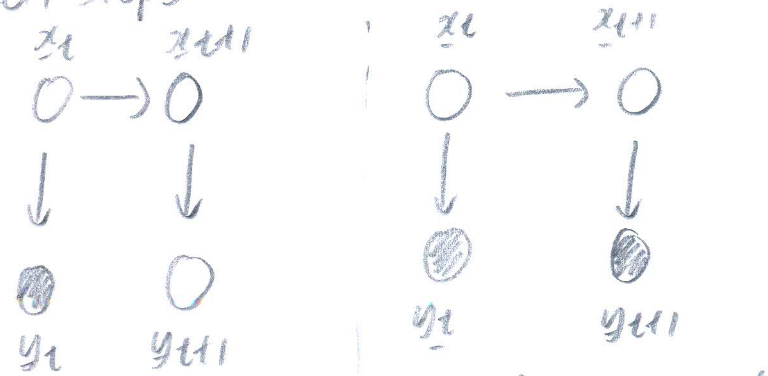
(*) Important notation (P3):-
- To indicate condit. mean/covariances with emphasis on part. output sequence being conditioned on
- $\hat{x}_{t|t}$ - mean of x_t conditioned on partial sequence $y_0, \ldots, y_t$
- $P_{t|t}$ - covariance matrix of x_t ——— " ———

$$\hat{x}_{t|t} := \mathbb{E}[x_t | y_0, \ldots, y_t]$$

$$P_{t|t} := \mathbb{E}[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T | y_0, \ldots, y_t]$$

- Also: we will require distn of x_t conditioned on $y_0, \ldots, y_{t-1}$ (mean/var)
i.e. $\hat{x}_{t|t-1}$ and $P_{t|t-1}$

(*) Clarity on time update, measurement steps
- LHS: conditioned on $y_0, \ldots, y_t$
- Assume already calc. $p(x_t | y_0, \ldots, y_t)$ (i.e. $\hat{x}_{t|t}, P_{t|t}$)
- We want to carry this distn forward into fragment on RHS (conditioning on $y_0, \ldots, y_{t+1}$)



before measurement after measurement

(*) Decompose transformation (ii):-

<u>Time update</u> $p(x_t|y_0,\dots,y_t) \longrightarrow p(x_{t+1}|y_0,\dots,y_t)$

<u>Measurement update</u> $p(x_{t+1}|y_0,\dots,y_t) \longrightarrow p(x_{t+1}|y_0,\dots,y_{t+1})$

<u>Time update</u>: propagate distri forward in time, evaluating new mean and covariance $(\hat{x}_{t+1|t}, P_{t+1|t})$ from old mean and cov $(\hat{x}_{t|t}, P_{t|t})$; but with no new measurements ✓

<u>Measurement update</u>: incorporate new measurement y_{t+1}, update p.d. for x_{t+1}

- <u>Overall result</u>: transformation from $\hat{x}_{t|t}$ and $\hat{P}_{t|t} \rightarrow \hat{x}_{t+1|t+1}$ and $P_{t+1|t+1}$

(*) Time update equations:- (dynamical model)

- <u>SSM equations</u>:-

$$x_{t+1} = Ax_t + Gw_t \qquad w_t \sim N(0,Q) \qquad (i) \qquad (15.9)$$

$$y_t = Cx_t + v_t \qquad v_t \sim N(0,R) \qquad (ii) \qquad (15.2)$$

- note:- $y_t|x_t \sim N(Cx_t, R)$ and $x_{t+1}|x_t \sim N(Ax_t, GQG^T)$

(*) Take conditional expectations on both sides of (i)
- w_t is independent of conditioning variables $y_0,\dots,y_t$ (process noise is independent of all observations)

$$\hat{x}_{t+1|t} = \mathbb{E}[(Ax_t + Gw_t)|y_0,\dots,y_t] = A\mathbb{E}[x_t|y_{0:t}] + G\,\mathbb{E}[\underbrace{w_t|y_{0:t}}_{=0}]$$

$$\Rightarrow \hat{x}_{t+1|t} = A\hat{x}_{t|t} \qquad (15.10)$$

(*) Take condit. covariance of both sides of (ii)
- $\hat{x}_{t+1|t}$ - constant in condit. distri
- w_t - zero mean
- w_t and x_t are independent (process noise and state are independent)
  - $\mathbb{E}[x_t w_t] = \mathbb{E}[x_t]\mathbb{E}[w_t]$

$$P_{t+1|t} = \mathbb{E}\left[(x_{t+1} - \hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T \,|\, y_{0:t}\right]$$

subs. (15.9), (15.10)

$$= \mathbb{E}\left[\underbrace{(Ax_t + Gw_t - A\hat{x}_{t|t})(Ax_t + Gw_t - A\hat{x}_{t|t})^T}\,|\, y_{0:t}\right]$$

· expanding and collecting terms:-

$$P_{t+1|t} = \mathbb{E}\left[Ax_t x_t^T A^T - Ax_t \hat{x}_{t|t}^T A^T - A\hat{x}_{t|t} x_t^T A^T + A\hat{x}_{t|t}\hat{x}_{t|t}^T A^T \,|\, y_{0:t}\right]$$

$$+ G\,\mathbb{E}\left[w_t w_t^T \,|\, y_{0:t}\right] G^T$$

$$+ A\,\underbrace{\mathbb{E}[x_t w_t^T | y_{0:t}]}_{(i)} G^T + G\,\underbrace{\mathbb{E}[w_t x_t^T | y_{0:t}]}_{(i)} A^T - G\,\underbrace{\mathbb{E}[w_t | y_{0:t}]}_{(ii)} \hat{x}_{t|t}^T A^T$$

$$- A\hat{x}_{t|t}\,\underbrace{\mathbb{E}[w_t^T | y_{0:t}]}_{(ii)} G^T$$

note: (i):- $\mathbb{E}[x_t w_t^T | y_{0:t}] \underset{\text{via independence}}{=} \mathbb{E}[x_t | y_{0:t}]\,\underbrace{\mathbb{E}[w_t^T | y_{0:t}]}_{=0} = 0$

(ii):- $\mathbb{E}[w_t | y_{0:t}] = 0$

$$\Rightarrow P_{t+1|t} = A\,\mathbb{E}\left[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T | y_{0:t}\right] A^T + G\,\underbrace{\mathbb{E}\left[w_t w_t^T | y_{0:t}\right]}_{=Q \text{ as } w_t \text{ has 0 mean}} G^T$$

$$= AP_{t|t}A + GQG^T \qquad (15.13)$$

(*) we have established the conditional distn of x_{t+1} given $y_0, \dots, y_t$

(*) now we establish conditional distn of y_{t+1} —— " ——

(*) use this to establish joint conditional of x_{t+1}, y_{t+1} —— " ——

- SSM equations:- $\quad y_{t+1} = Cx_{t+1} + v_{t+1} \quad v_{t+1} \sim N(0, R)$
- (observation model)
- Similar to previous, take conditional expectations wrt $y_0, \dots, y_t$

- $\mathbb{E}[y_{t+1} | y_0, \dots, y_t] = \mathbb{E}[Cx_{t+1} + v_{t+1} | y_0, \dots, y_t]$

$$= C\mathbb{E}[x_{t+1} | y_{0:t}] + \underbrace{\mathbb{E}[v_{t+1}]}_{=0}$$

$$= C\hat{x}_{t+1|t}$$

- $\mathbb{E}[(y_{t+1} - \hat{y}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T | y_{0:t}]$

$= \mathbb{E}[(Cx_{t+1} + v_{t+1} - \hat{y}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T | y_{0:t}]$ - suppressing conditioning - brevity

$= C\mathbb{E}[x_{t+1} x_{t+1}^T] - C\mathbb{E}[x_{t+1}\hat{x}_{t+1|t}^T] + \mathbb{E}[v_{t+1} x_{t+1}^T] - \mathbb{E}[v_{t+1}\hat{x}_{t+1|t}^T]$

$- \mathbb{E}[\hat{y}_{t+1|t} x_{t+1}^T] + \mathbb{E}[\hat{y}_{t+1|t}\hat{x}_{t+1|t}^T]$

?

0/S1

=

=

- $\mathbb{E}[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})^T | y_0, \dots, y_t]$ - suppressing condit.

$= \mathbb{E}[(Cx_{t+1} + v_{t+1} - C\hat{x}_{t+1|t})(Cx_{t+1} + v_{t+1} - C\hat{x}_{t+1|t})^T]$

- similar to earlier

$= \mathbb{E}[(Cx_{t+1}x_{t+1}^T C^T - Cx_{t+1}\hat{x}_{t+1|t}^T C^T - C\hat{x}_{t+1|t} x_{t+1}^T C^T + C\hat{x}_{t+1|t}\hat{x}_{t+1|t}^T C^T]$

$+ \mathbb{E}[v_{t+1} v_{t+1}^T]$

$+ C\mathbb{E}[x_{t+1} v_{t+1}^T] + \mathbb{E}[v_{t+1} x_{t+1}^T]C^T - \mathbb{E}[v_{t+1}\hat{x}_{t+1|t}^T]C^T - C\mathbb{E}[\hat{x}_{t+1|t} v_{t+1}^T]$

all 0 terms (see before)

$= C\mathbb{E}[(x_{t+1} - \hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T] C^T + R$

$= CP_{t+1|t}C^T + R$

(*) Joint distribution (cond. on past outputs $y_{0:t}$)

- $p(x_{t+1}, y_{t+1} | y_0, \ldots, y_t) = N(m_{t+1}, V_{t+1})$

$$m_{t+1} = \begin{pmatrix} \hat{x}_{t+1|t} \\ C\hat{x}_{t+1|t} \end{pmatrix} \qquad V_{t+1} = \begin{pmatrix} P_{t+1|t} & P_{t+1|t}C^T \\ CP_{t+1|t} & CP_{t+1|t}C^T + R \end{pmatrix}$$

(*) similar to FA;
reverse arrow → compute condit. distri of x_{t+1} given y_{t+1} (i.e. find posterior)
where x_{t+1} and y_{t+1} have a joint Gaussian distri.

(*) difference: joint distri is itself also a conditional.

(*) conditional distri :- (measurement update)

$p(x_{t+1} | y_0, \ldots, y_{t+1}) = N(x_{t+1} | m_{1|2}, V_{1|2})$

$m_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2)$

$V_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$$\hat{x}_{t+1|t+1} = m_{1|2} = \hat{x}_{t+1|t} + P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}(y_{t+1} - C\hat{x}_{t+1|t})$$

$$P_{t+1|t+1} = V_{1|2} = P_{t+1|t} - P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}CP_{t+1|t}$$

(*) Don't be put off by dense symbology → a straightforward application of condit. Gaussians.

(*) Kalman gain matrix

$K_{t+1} = P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}$

(*) Filtering equations:

- At time t, assume we have the mean estimate $\hat{x}_{t|t}$ and covariance est $P_{t|t}$.

- using this, we recursively calculate $\hat{x}_{t+1|t+1}$ and $P_{t+1|t+1}$:-

Time update :- $\hat{x}_{t+1|t} = A\hat{x}_{t|t}$

$P_{t+1|t} = AP_{t|t}A^T + GQG^T$

- Measurement update :- $\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + K_{t+1}(y_{t+1} - C\hat{x}_{t+1|t})$

$P_{t+1|t+1} = P_{t+1|t} - KCP_{t+1|t}$ (from slides)

(?) Q/S 2 : lecture slides state that K_t can be precomputed, as independent of data. Referring to?
- slides use different notation for K_t and K_{t+1}; while in Jordan, they are the same expression.
- what am I missing here?

(*) Omitted review from Jordan:-

i) LMS interp.

ii) information filter

iii) RTS smoothing

(*) Following things to cement understanding:-

i) Probabilistic story (via appl. of Gaussians)

ii) Graphical models story

iii) physical model e.g airplane position

iv) other intuitive anchors.