L5, Friday

8.4 Bayesian formulation of state estimation

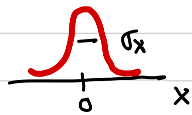
Kalman: track mean + variance of state estimates

Bayes: look at entire pdf (in principle ...)

Toy model: One meas. of scalar $y = x + \zeta$

– observe y, want to infer x.

"prior" on x: $x \sim N(0, \sigma_x^2)$



Meas. noise ζ: $\zeta \sim N(0, \sigma_\zeta^2)$

Bayes Thm: $p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x)$

likelihood

posterior $$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$ prior

evidence (norm.)

$$p(x|y) \sim p(y|x)\,p(x)$$ neglecting normalization

$$= N(\overbrace{y-x}^{\zeta}, \sigma_\zeta^2) \cdot N(0, \sigma_x^2)$$

$$\sim \exp\left[-\frac{(y-x)^2}{2\sigma_\zeta^2}\right] \cdot \exp\left[-\frac{x^2}{2\sigma_x^2}\right]$$

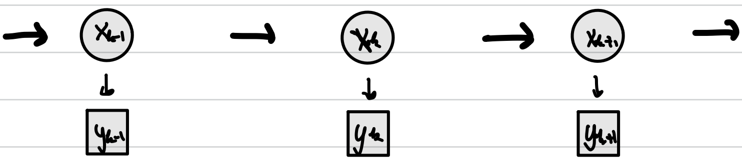
$$\sim \exp\left[-\frac{\left(x - \frac{\sigma_x^2}{\sigma_x^2+\sigma_\zeta^2}\,y\right)^2}{2\sigma_0^2}\right],$$

$$\frac{1}{\sigma_0^2} = \frac{1}{\sigma_\zeta^2} + \frac{1}{\sigma_x^2}$$

So the posterior is another Gaussian, with

$$\hat{x} = \left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_\xi^2}\right) y$$ and standard dev. $\sigma_0 < \mathrm{Min}[\sigma_x, \sigma_\xi]$

between true pos. x & obs. y (closer to prior mean, 0)



Probabilistic state space model ("hidden" Markov proc.)

$x_1 \sim p(x_1)$ initial state

$x_{k+1} \sim p(x_{k+1} | x_k)$ dynamics

$y_k \sim p(y_k | x_k)$ observation

Markov dynamics: $p(x_{k+1} | x^k, y^k) = p(x_{k+1} | x_k)$

$x^k = \{x_1, x_2, \dots, x_k\}$

$y^k = \{y_1, y_2, \dots, y_k\}$

Conditional independence: knowing $x_k \Rightarrow$ all other info irrelevant

Observations a "memoryless" function of state alone.

$$p(y_k | x^k, y^{k-1}) = p(y_k | x_k)$$

Note that $\dot{x} = f(x, u)$ $y = h(x, u)$ $\rightarrow$ the cond. prob. dist.

General scheme: $p(x_k|y^k) \xrightarrow[\text{predict}]{} p(x_{k+1}|y^k) \xrightarrow[\text{update}]{} p(x_{k+1}|y^{k+1})$

Predict using **Chapman-Kolmogorov** for Markov dyn.

$$p(x_{k+1}|y^k) = \int dx_k \overbrace{p(x_{k+1}|x_k)\, p(x_k|y^k)}^{p(x_{k+1},x_k|y^k)}$$

→ process noise (not nec. additive)

Dynamics: $x_{k+1} = f(x_k, u_k, \nu_k)$

where $p(x_{k+1}|x_k) = \int d\nu_k\, p(x_{k+1}, \nu_k | x_k)$ marginalization

$= \int d\nu_k\, p(x_{k+1}|x_k,\nu_k)\, p(\nu_k|x_k)$ cond. prob.

$= \int d\nu_k\, \delta[x_{k+1} - f(x_k,u_k,\nu_k)]\, p(\nu_k)$ noise ind. of x_k

$= p(\nu_k^*)$ ν_k^* solves $x_{k+1} - f(\cdot\cdot\cdot) = 0$

Update (Bayes): $p(x_{k+1}|y^{k+1}) = \frac{1}{Z}\, \underset{\downarrow \atop p(y_{k+1}|x_{k+1})}{p(y_{k+1}|x_{k+1}, y^k)}\, p(x_{k+1}|y^k)$

Normalization $Z = \int dx_{k+1}\, p(y_{k+1}|x_{k+1})\, p(x_{k+1}|y^k) = p(y_{k+1}|y^k)$

Bayesian filter eqs.

$$p(x_{k+1} \mid y^k) = \int dx_k \; p(x_{k+1} \mid x_k)\, p(x_k \mid y^k)$$ predict

$$p(x_{k+1} \mid y^{k+1}) = \frac{1}{Z}\, p(y_{k+1} \mid x_{k+1})\, p(x_{k+1} \mid y^k)$$ update

$$Z = p(y_{k+1} \mid y^k) = \int dx_{k+1} \; p(y_{k+1} \mid x_{k+1})\, p(x_{k+1} \mid y^k)$$

with
$$x_{k+1} = f(x_k, u_k, \nu_k)$$ dynamics
$$y_k = h(x_k, \xi_k)$$ measurement

Hybrid dynamics

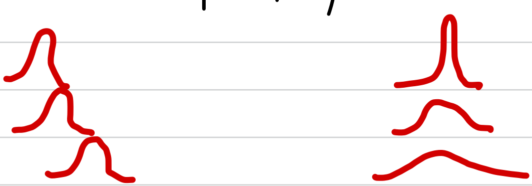
$$\dot{x} = f(x) + g(x)\nu$$ slight specialization

Fokker-Planck:

$$\partial_t\, p(x,t) = -\nabla \cdot [f(x)\, p] + \nabla_x^2 (D p)$$ Laplacian

$p(x,t)$

$$D = \tfrac{1}{2} g g^T$$



drift

diffusion

From Bayes to Kalman

- linear dynamics, Gaussian noise + initial condition

$$x_{k+1} = Ax_k + Bu_k + v_k, \qquad y_k = Cx_k + \xi_k$$

$$p(x_{k+1} \mid y^k) \sim N(x_{k+1};\ \hat{x}^-_{k+1},\ P^-_{k+1})$$

$$p(y_{k+1} \mid x_{k+1}) \sim N(y_{k+1} - Cx_{k+1};\ 0,\ Q_\xi)$$

$$p(y_{k+1} \mid y^k) \sim N(y_{k+1};\ C\hat{x}^-_{k+1},\ CP^-_{k+1}C^T + Q_\xi)$$

$$p(x_{k+1} \mid y^{k+1}) \sim \frac{p(y_{k+1} \mid x_{k+1})\, p(x_{k+1} \mid y^k)}{p(y_{k+1} \mid y^k)}$$

$$\sim N(x_{k+1};\ \hat{x}_{k+1},\ P_{k+1})$$

Algebra is pretty fierce ... (conditional Gaussians)

→ Kalman filter is just the same as Bayes filter assuming linear dynamics and Gaussian noise + init.

(linear combo of Gaussians is Gaussian ...)

To give an idea of how the calculations go, look at $x_{k+1} = x_k + \nu_k$ (ignore y's ...)

$$p(x_{k+1}) = \int dx_k \; p(x_{k+1}, x_k)$$

$$= \int dx_k \; p(x_{k+1} | x_k) \; p(x_k)$$

$$= \int dx_k \; N(x_{k+1} - x_k ; 0, \nu^2) \; N(x_k ; \hat{x}_k , P_k)$$

$$= N(x_k ; \hat{x}_k , P_k + \nu^2)$$ sum of Gaussians

Why choose the mean as "the" representative value?

Assume we are interested in $p(x|y) \sim p(y|x)\, p(x)$

We want to choose $\hat{x}$ to "best" represent $p(x|y)$.

Define cost (loss) function $J(\hat{x}) = \langle (x - \hat{x})^2 \rangle$

$$J = \int dx \, (x-\hat{x})^2 \, p(x|y)$$

$$\frac{\partial J}{\partial \hat{x}} = 2 \int dx \, (x-\hat{x}) \, p(x|y) = 0$$

$\int dx \cdot x \cdot p(x|y) \equiv \langle x \rangle_y$ conditional mean $\qquad \int dx \, \hat{x} \, p(x|y) = \hat{x} \overbrace{\int dx \, p(x|y)}^{1}$

$$\Rightarrow \quad \hat{x} = \langle x \rangle_y$$

- For linear dynamics, Gaussian noise (and init. cond.)
  $\rightarrow$ Kalman filter updates for $\hat{x}_{k+1}$, P_{k+1}

- For weak nonlinear dynamics, non-Gauss noise, etc.
  "perturbative approaches"

- For stronger nonlinearities, etc.
  direct numerical methods, Monte Carlo