Nonlinear filtering

1) What is nonlinear filtering?

$$\begin{tikzcd}
{} \arrow[r] & x_{t-1} \arrow[r] \arrow[d] & x_t \arrow[r] \arrow[d] & x_{t+1} \arrow[r] \arrow[d] & {} \\
& y_{t-1} & y_t & y_{t+1} &
\end{tikzcd}$$

transition probability $p(x_t \mid x_{t-1})$

emission probability $p(y_t \mid x_t)$

$p(x_t \mid y_{0:t}) \qquad p(x_{t+1} \mid y_{0:t+1})$ posterior

$= p(x_t \mid y_t, y_{0:t-1})$

Goal: We want to get an estimate about the real state of x_t according to the measurements taken at $y_{0:t} = y_0, \dots, y_t$

How can we solve this problem? Easy: Bayes' rule

$$p(A \mid B) = \frac{p(B \mid A)\, p(A)}{p(B)} \quad \& \quad \text{marginalization trick} \quad p(A) = \int p(A, c)\, dc$$

↳ do the same here with $A = x_t$ and $B = y_t$. Leave rest untouched.

$$p(x_t \mid y_{0:t}) = \frac{p(y_t \mid x_t, y_{0:t-1})\; p(x_t \mid y_{0:t-1})}{p(y_t \mid y_{0:t-1})} \quad \Big| \text{ marginalization trick}$$

$$= \frac{p(y_t \mid x_t, y_{0:t-1}) \int dx_{t-1}\, p(x_t \mid x_{t-1}, y_{0:t-1})\, p(x_{t-1} \mid y_{0:t-1})}{\int dx_t\, p(y_t \mid x_t, y_{0:t-1})\, p(x_t \mid y_{0:t-1})}$$

simplify with gen. model

$$= \frac{\overbrace{p(y_t \mid x_t)}^{\text{update}}\; \overbrace{\int dx_{t-1}\, p(x_t \mid x_{t-1})\, p(x_{t-1} \mid y_{0:t-1})}^{\text{prediction}}}{Z_t(y_{0:t})}$$

posterior of previous time step

normalization independent of latent variable

- problem: integrals often cannot be solved in closed form
  → either choose models that do close (sec 2) or introduce approximation

2) Linear filtering → Kalman filter

- linear-Gaussian state-space model: latent variables x_t as well as observations y_t are multivariate Gaussian distributions whose means are linear functions of the states of their parents in the graph

- LDS:

$$p(x_t \mid x_{t-1}) = \mathcal{N}(F_t x_{t-1}, Q_t) \quad \text{transition}$$

$$p(y_t \mid x_t) = \mathcal{N}(H_t x_t, R_t) \quad \text{emission}$$

- trick: we know everything is Gaussian, so the posterior is also a Gaussian (→ integrals are "closed")

$$p(x_{t-1} \mid y_{0:t-1}) = \mathcal{N}(\mu_{t-1}, \Sigma_{t-1})$$

$\downarrow$ prediction

$$p(x_t \mid y_{0:t-1}) = \int dx_{t-1}\, \mathcal{N}(x_t ; F x_{t-1}, Q_t)\, \mathcal{N}(x_{t-1} ; \mu_{t-1}, \Sigma_{t-1})$$

$$p(x_t \mid y_{0:t-1}) = \mathcal{N}(F_t \mu_{t-1}, \overbrace{F_t \Sigma_{t-1} F_t^T + Q_t}^{:= P_t})$$

$\downarrow$ update

$$p(x_t \mid y_{0:t}) \propto_{x_t} \mathcal{N}(y_t ; H_t x_t, R_t)\, \mathcal{N}(F_t \mu_t ; P_t)$$

$$p(x_t \mid y_{0:t}) = \mathcal{N}(x_t ; \mu_t, \Sigma_t)$$

$$\mu_t = F_t \mu_{t-1} + K_t (y_t - H_t F_t \mu_{t-1}) \qquad \text{with } K_t = P_t H_t^T (H_t P_t H_t^T + R_t)^{-1}$$

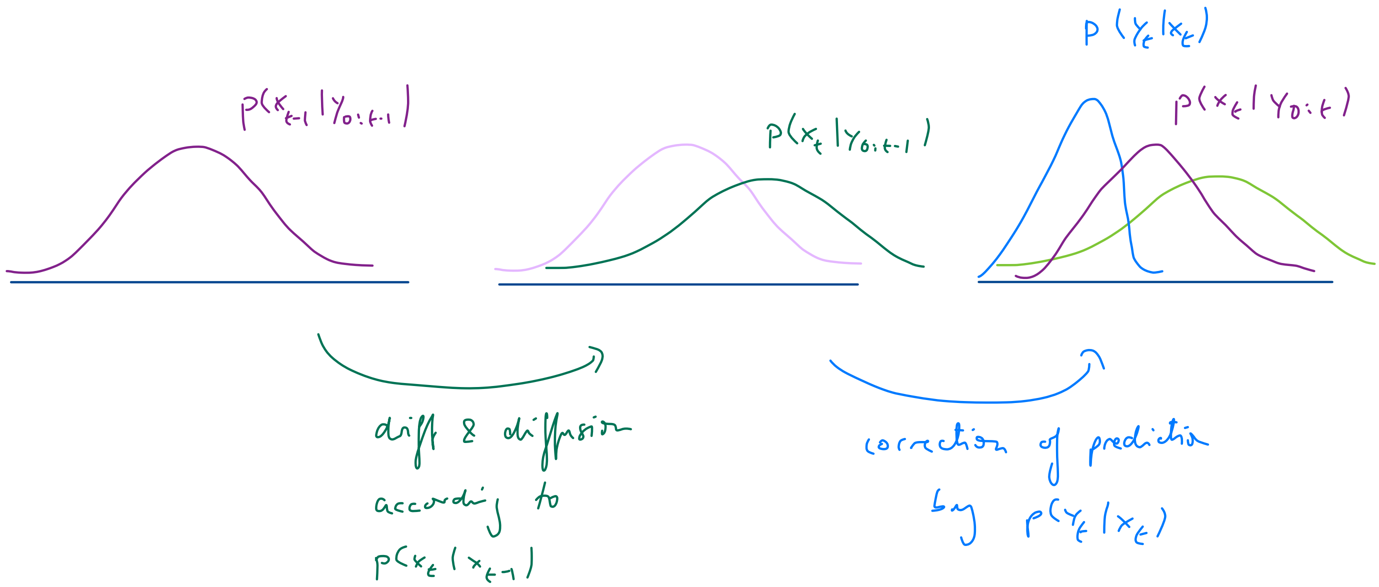
$$\Sigma_t = (\mathbb{1} - K_t H_t) P_t$$

Simple example: $x_t, y_t \in \mathbb{R}$, $Q_t = q^2$, $R_t = r^2$, $F = H = \mathbb{1}$ (Brownian motion)

$$P_t = \sigma_{t-1}^2 + q^2 = \tilde{\sigma}_t^2$$

$$K_t = \frac{\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2 + r^2}$$

$$\Sigma_t = \left(1 - \frac{\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2 + r^2}\right)\tilde{\sigma}_t^2 = \frac{r^2\,\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2 + r^2}$$



3) Nonlinear filtering: Particle filters

3.1. Importance sampling

Assume $p(x)$ can be evaluated, but is hard to sample from.

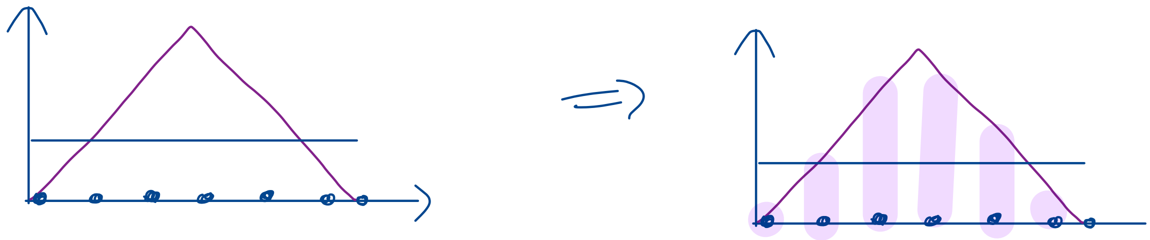
→ sample instead from proposal distribution $q(x)$

$$\langle f(x)\rangle_p = \int dx\, f(x)\, p(x) = \int dx\, f(x) \frac{p(x)}{q(x)} q(x) = \left\langle f(x) \frac{p(x)}{q(x)} \right\rangle_q$$

Now sample $x^i \sim q(x)$. Then

$$\left\langle f(x) \frac{p(x)}{q(x)} \right\rangle_q \approx \frac{1}{N}\sum_{i=1}^{N} f(x^i) \frac{p(x^i)}{q(x^i)} = \sum_i w^i f(x^i)$$

$$\text{s.th. } \sum_i w^i = 1 \quad (\text{normalization})$$



3.2. Sequential importance sampling

Consider importance sampling of the full posterior.

Ingredient 1: The full posterior can also be written recursively

$$p(x_{0:t} | y_{0:t}) = \frac{p(y_t | x_{0:t}, y_{0:t-1})\, p(x_{0:t} | y_{0:t-1})}{p(y_t | y_{0:t-1})}$$

$$= \frac{p(y_t | x_t)\; p(x_t | x_{t-1}, y_{0:t-1})\, p(x_{t-1} | y_{0:t-1})}{p(y_t | y_{0:t-1})}$$

$$= \frac{p(y_t | x_t)\, p(x_t | x_{t-1})}{p(y_t | y_{0:t-1})}\; p(x_{t-1} | y_{0:t-1})$$

Ingredient 2: The proposal $q(x_{0:t} | y_{0:t})$ factorizes

$$q(x_{0:t} | y_{0:t}) = \pi(x_t | x_{0:t-1}, y_{0:t})\; q(x_{0:t-1} | y_{0:t-1})$$

$$\Rightarrow \quad w_t^i \propto \frac{p(x_{0:t}^i | y_{0:t})}{q(x_{0:t}^i | y_{0:t})} \propto \frac{p(y_t | x_t^i)\, p(x_t^i | x_{t-1}^i)}{\pi(x_t^i | x_{0:t-1}^i, y_{0:t})} \underbrace{\frac{p(x_{t-1}^i | y_{0:t-1})}{q(x_{t-1}^i | y_{0:t-1})}}_{w_{t-1}^i}$$

$$p(x_{0:t} | y_{0:t}) = \sum_i w_t^i \prod_{s=0}^{t} \delta(x_s - x_s^i)$$

$$p(x_t | y_{0:t}) = \int dx_0 \dots \int dx_{t-1} \sum_i w_t^i \prod_{s=0}^{t} \delta(x_s - x_s^i) = \sum_i w_t^i\, \delta(x_t - x_t^i)$$

So we find

$$x_t^i \sim \pi(x_t \mid x_{0:t}^i, y_{0:t})$$

$$\tilde{w}_t^i \propto w_{t-1}^i \frac{p(y_t \mid x_t^i)\, p(x_t^i \mid x_{t-1}^i)}{\pi(x_t^i \mid x_{t-1}^i, y_{0:t})}$$

BPF:

$$x_t^i \sim p(x_t \mid x_t^i)$$

$$\tilde{w}_t^i \propto w_{t-1}^i \, p(y_t \mid x_t^i)$$

- normalization : $w_t^i = \frac{\tilde{w}_t^i}{\sum_j \tilde{w}_t^j}$
- resampling from weight distribution if needed!

If not enough time, just show this picture