The simplest filtering

D.A. Forsyth

University of Illinois at Urbana Champaign

Issue to solve in SLAM

- We have:
  - reconstructions of some points
  - configurations of some cameras
- A new frame arrives:
  - more information about old point positions
    - and so, more information about previous cameras
  - some new point positions
- Q:
  - how should we maintain point positions, cameras, in the light of new information?
    - in a reasonably lightweight way?

Filtering - Setup

- Very general model:
  - We assume there is an underlying state X
  - There are observations Y
    - some of which are functions of this state
  - There is a clock
    - at each tick, the state changes with known dynamics
    - at each tick, we get a new observation
    - (in principle, you can do this in continuous time, but…)
- Q:
  - construct "best" estimate of state
    - before/after observation

Filtering - Examples

- Tracking a ball:
  - object is ball
  - state is 3D position+velocity
  - observations are stereo pairs
- Tracking a person:
  - object is person
  - state is body configuration
  - observations are frames, clock is in camera (30 fps)
- SLAM:
  - state is structure of the world, location of agent
  - observations
    - images, lidar, angular measurements, etc.
  - clock is in camera

In probability language…

- Given
  - “Prior” $$p(X_{i-1}|Y_0, \ldots, Y_{i-1})$$
- We should like to know
  - “Predictive distribution” $$p(X_i|Y_0, \ldots, Y_{i-1})$$ haven’t got the new obs yet (Y_{i-1})
  - “Posterior” $$p(X_i|Y_0, \ldots, Y_i)$$ have the new obs (Y_i)

Key assumptions:

- **Only the immediate past matters:** formally, we require

$$P(\boldsymbol{X}_i|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{i-1}) = P(\boldsymbol{X}_i|\boldsymbol{X}_{i-1})$$

  This assumption hugely simplifies the design of algorithms, as we shall see; furthermore, it isn't terribly restrictive if we're clever about interpreting $\boldsymbol{X}_i$ as we shall show in the next section.

- **Measurements depend only on the current state:** we assume that $\boldsymbol{Y}_i$ is conditionally independent of all other measurements given $\boldsymbol{X}_i$. This means that

$$P(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \ldots \boldsymbol{Y}_k|\boldsymbol{X}_i) = P(\boldsymbol{Y}_i|\boldsymbol{X}_i)P(\boldsymbol{Y}_j, \ldots, \boldsymbol{Y}_k|\boldsymbol{X}_i)$$

  Again, this isn't a particularly restrictive or controversial assumption, but it yields important simplifications.

Filtering as Induction - base case

Firstly, we assume that we have $P(\boldsymbol{X}_0)$

Then we have

$$
\begin{aligned}
P(\boldsymbol{X}_0|\boldsymbol{Y}_0 = \boldsymbol{y}_0) &= \frac{P(\boldsymbol{y}_0|\boldsymbol{X}_0)P(\boldsymbol{X}_0)}{P(\boldsymbol{y}_0)} \\
&= \frac{P(\boldsymbol{y}_0|\boldsymbol{X}_0)P(\boldsymbol{X}_0)}{\int P(\boldsymbol{y}_0|\boldsymbol{X}_0)P(\boldsymbol{X}_0)d\boldsymbol{X}_0} \\
&\propto P(\boldsymbol{y}_0|\boldsymbol{X}_0)P(\boldsymbol{X}_0)
\end{aligned}
$$

Filtering as induction - induction step

Given

$$P(\boldsymbol{X}_{i-1}|\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}).$$

Prediction

Prediction involves representing

$$P(\boldsymbol{X}_i|\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1})$$

Notice this is i-1 current state based on previous measurements

Our independence assumptions make it possible to write

$$\begin{aligned} P(\boldsymbol{X}_i|\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}) &= \int P(\boldsymbol{X}_i, \boldsymbol{X}_{i-1}|\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}) d\boldsymbol{X}_{i-1} \\ &= \int P(\boldsymbol{X}_i|\boldsymbol{X}_{i-1}, \boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}) P(\boldsymbol{X}_{i-1}|\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}) d\boldsymbol{X}_{i-1} \\ &= \int P(\boldsymbol{X}_i|\boldsymbol{X}_{i-1}) P(\boldsymbol{X}_{i-1}|\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}) d\boldsymbol{X}_{i-1} \end{aligned}$$

Filtering as induction - induction step

Correction

Correction involves obtaining a representation of

$$P(\boldsymbol{X}_i|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_i)$$

Notice this is i
Prediction based on
current measurement
as well.

Our independence assumptions make it possible to write

$$\begin{aligned}
P(\boldsymbol{X}_i|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_i) &= \frac{P(\boldsymbol{X}_i,\boldsymbol{y}_0,\ldots,\boldsymbol{y}_i)}{P(\boldsymbol{y}_0,\ldots,\boldsymbol{y}_i)} \\
&= \frac{P(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{i-1})P(\boldsymbol{X}_i|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{i-1})P(\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{i-1})}{P(\boldsymbol{y}_0,\ldots,\boldsymbol{y}_i)} \\
&= P(\boldsymbol{y}_i|\boldsymbol{X}_i)P(\boldsymbol{X}_i|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{i-1})\frac{P(\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{i-1})}{P(\boldsymbol{y}_0,\ldots,\boldsymbol{y}_i)} \\
&= \frac{P(\boldsymbol{y}_i|\boldsymbol{X}_i)P(\boldsymbol{X}_i|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{i-1})}{\int P(\boldsymbol{y}_i|\boldsymbol{X}_i)P(\boldsymbol{X}_i|\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{i-1})d\boldsymbol{X}_i}
\end{aligned}$$

We need

- to know:
  - dynamical model (how X changes with time)
  - measurement model (how Y depends on X)
- to represent
  - all the probability distributions we deal with

Illuminating 1D example

- Drop a measuring device on a cable down a hole
  - where is it?
- Setup:
  - measurement of depth x
  - actual distance down the hole θ
  - known $p(\theta)$ which will be normal, $N(\theta_c; \sigma_c^2)$
  - known $p(x|\theta)$ which will be normal, $N(c\theta; \sigma_m^2)$
- Q: what is $p(\theta|x)$

A 1D problem, II

(Bayes rule), so that:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

And:

$$\begin{aligned} \log p(\theta|x) &= \log p(x|\theta) + \log p(\theta) + K \\ &= -\frac{(c\theta - x)^2}{2\sigma_m^2} - \frac{(\theta - \theta_c)^2}{2\sigma_c^2} + K' \end{aligned}$$

A 1D problem, III

$$\log p(\theta|x) = -\frac{(c\theta - x)^2}{2\sigma_m^2} - \frac{(\theta - \theta_c)^2}{2\sigma_c^2} + K'$$

$$= -\frac{\theta^2}{2}\left[\frac{\sigma_m^2 + c^2\sigma_c^2}{\sigma_m^2\sigma_c^2}\right] + \theta\left[\frac{\theta_c\sigma_m^2 + cx\sigma_c^2}{\sigma_m^2\sigma_c^2}\right] + K''$$

A 1D problem, IV

- Now *IF* $p(\theta|x)$ is normal
  - (say $N(\mu_t; \sigma_t^2)$)

- Then

$$\log p(\theta|x) = -\frac{(\theta - \mu_t)^2}{2\sigma_t^2} + K'''$$

$$= -\frac{\theta^2}{2\sigma_t^2} + \theta\frac{\mu_t}{\sigma_t^2} + K'''$$

Pattern match

$$
\begin{aligned}
\log p(\theta|x) &= -\frac{(c\theta - x)^2}{2\sigma_m^2} - \frac{(\theta - \theta_c)^2}{2\sigma_c^2} + K' \\
&= -\frac{\theta^2}{2}\left[\frac{\sigma_m^2 + c^2\sigma_c^2}{\sigma_m^2\sigma_c^2}\right] + \theta\left[\frac{\theta_c\sigma_m^2 + cx\sigma_c^2}{\sigma_m^2\sigma_c^2}\right] + K''
\end{aligned}
$$

$$
\begin{aligned}
\log p(\theta|x) &= -\frac{(\theta - \mu_t)^2}{2\sigma_t^2} + K''' \\
&= -\frac{\theta^2}{2\sigma_t^2} + \theta\frac{\mu_t}{\sigma_t^2} + K'''
\end{aligned}
$$

A 1D problem, V

$$\mu_t = \frac{\theta_c \sigma_m^2 + cx\sigma_c^2}{\sigma_m^2 + c^2\sigma_c^2}$$

$$\sigma_t^2 = \frac{\sigma_c^2 \sigma_m^2}{\sigma_m^2 + c^2\sigma_c^2}$$

- Important checks:
  - what happens if measurement is unreliable?
  - what happens if prior is very diffuse?

Summary, with change of notation

Useful Fact: 9.2 *The parameters of a normal posterior with a single measurement*

Assume we wish to estimate a parameter θ. The prior distribution for θ is normal, with known mean μ_π and known standard deviation σ_π. We receive a single data item x_1 and a scale c_1. The likelihood of x_1 is normal with mean $c_1\theta$ and standard deviation $\sigma_{m,1}$, where $\sigma_{m,1}$ is known. Then the posterior, $p(\theta|x_1, c_1, \sigma_{m,1}, \mu_\pi, \sigma_\pi)$, is normal, with mean

$$\mu_1 = \frac{c_1 x_1 \sigma_\pi^2 + \mu_\pi \sigma_{m,1}^2}{\sigma_{m,1}^2 + c_1^2 \sigma_\pi^2}$$

and standard deviation

$$\sigma_1 = \sqrt{\frac{\sigma_{m,1}^2 \sigma_\pi^2}{\sigma_{m,1}^2 + c^2 \sigma_\pi^2}}.$$

A second measurement arrives…

- We know that $p(\theta|x)$ is normal
  - think of this as the prior
- We know that $p(x_1|\theta)$ is normal
  - think of this as the likelihood
- So:
  - the posterior $p(\theta|x_1, x)$ must be normal
  - and we can update as before!

Key, and crucial, property

- IF
  - prior is normal
  - dynamics are linear + Gaussian noise
  - measurement is linear+Gaussian noise
- THEN
  - predictive distribution is normal
  - posterior is normal
- Modelling distributions is just maintenance on
  - means, covariances
- Next measurement:
  - update predictive dist.
  - update posterior
  - posterior for this is now prior for next