

# Filtering - more general view

D.A. Forsyth, UIUC

# Filtering - more general view

- 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
    - at each tick, we get a new observation
- Examples
  - object is ball, state is 3D position+velocity, observations are stereo pairs
  - object is person, state is body configuration, observations are frames, clock is in camera (30 fps)

# Formal statement

- Given

- “Prior”

$$p(X_{i-1} | Y_0, \dots, Y_{i-1})$$

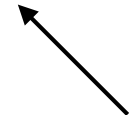
- We should like to know

- “Predictive distribution”

$$p(X_i | Y_0, \dots, Y_{i-1})$$

- “Posterior”

$$p(X_i | Y_0, \dots, Y_i)$$



## Key assumptions:

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

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

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

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

$$P(\mathbf{Y}_i, \mathbf{Y}_j, \dots, \mathbf{Y}_k | \mathbf{X}_i) = P(\mathbf{Y}_i | \mathbf{X}_i) P(\mathbf{Y}_j, \dots, \mathbf{Y}_k | \mathbf{X}_i)$$

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  $P(\mathbf{X}_0)$

Then we have

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

## Filtering as induction - induction step

Given

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

### Prediction

Prediction involves representing

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

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

Our independence assumptions make it possible to write

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

# Filtering as induction - induction step

## Correction

Correction involves obtaining a representation of

$$P(\mathbf{X}_i | \mathbf{y}_0, \dots, \mathbf{y}_i)$$

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

Our independence assumptions make it possible to write

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

# Required

- Clearly, we need to know
  - dynamical model (how  $X$  changes with time)
  - measurement model (how  $Y$  depends on  $X$ )
  - some way of representing all the probability distributions we deal with
    - 1D example suggests normals are particularly well behaved!