HMMs III: Linear Space Models

Many slides courtesy of
Dan Klein, Stuart Russell,
or Andrew Moore

CS 5300 / CS 6300
Artificial Intelligence
Spring 2009

Hal Daumé III
hal@cs.utah.edu

www.cs.utah.edu/~hal/courses/2009S_AI

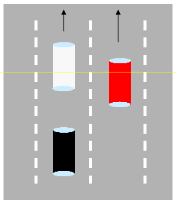
Announcements

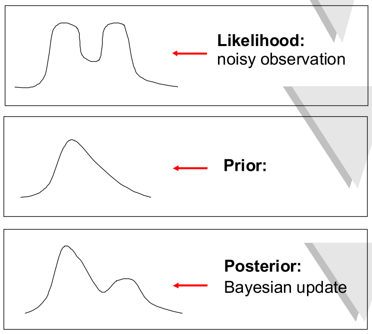
- HW7 solutions, HW8 solution up today
- Extra credit stuff posted to mailing list
  - Will send out assignments tonight
- Contest stuff up
  - You must register teams (email me per instructions)
  - Double elimination run every night
  - Top three U of U teams get some default extra credit

Tracking in Images

http://www.ai.sri.com/~beymer/mov/vsam-run.mov

Mathematical Formulation





Likelihood: noisy observation

Prior:

Posterior: Bayesian update

system states: car positions

observations: images

Notation

- $\mathbf{x}_k \in \mathbf{R}^d$: internal state at *k*th frame (hidden random variable, e.g. position of the object in the image).

  $\mathbf{X}_k = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]^T$: history up to time step *k*

- $\mathbf{z}_k \in \mathbf{R}^c$: measurement at *k*th frame (observable random variable, e.g. the given image).

  $\mathbf{Z}_k = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k]^T$: history up to time step *k*

Goal

Estimating the posterior probability $p(\mathbf{x}_k \mid \mathbf{Z}_k)$

How ???

One idea: recursion $p(\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1}) \Rightarrow p(\mathbf{x}_k \mid \mathbf{Z}_k)$

- How to realize the recursion ?
- What assumptions are necessary ?

Recursive Formula

$$\begin{aligned} & p(\mathbf{x}_k|\mathbf{Z}_k) \\ &= p(\mathbf{x}_k|\mathbf{Z}_{k-1},\mathbf{z}_k) \\ &\propto p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{Z}_{k-1})\,p(\mathbf{x}_k|\mathbf{Z}_{k-1}) \\ &\propto p(\mathbf{z}_k|\mathbf{x}_k)\,p(\mathbf{x}_k|\mathbf{Z}_{k-1}) \\ &\propto p(\mathbf{z}_k|\mathbf{x}_k)\int p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{Z}_{k-1})\,p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})\,d\mathbf{x}_{k-1} \\ &\propto p(\mathbf{z}_k|\mathbf{x}_k)\int p(\mathbf{x}_k|\mathbf{x}_{k-1})\,p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})\,d\mathbf{x}_{k-1} \end{aligned}$$

Bayes rule:
$p(a\,|\,b) = p(b\,|\,a)\,p(b)\,/\,p(a)$

Assumption:
$p(\mathbf{z}_k\,|\,\mathbf{x}_k,\mathbf{Z}_{k-1}) = p(\mathbf{z}_k\,|\,\mathbf{x}_k)$

Integration:
$p(a) = \int p(a\,|\,b)\,p(b)\,db$

Assumption:
$p(\mathbf{x}_k\,|\,\mathbf{x}_{k-1},\mathbf{Z}_{k-1}) = p(\mathbf{x}_k\,|\,\mathbf{x}_{k-1})$

Probabilistic Formulation

$$p(\mathbf{x}_k|\mathbf{Z}_k) = \kappa\; p(\mathbf{z}_k|\mathbf{x}_k)\int p(\mathbf{x}_k|\mathbf{x}_{k-1})\,p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})\,d\mathbf{x}_{k-1}$$

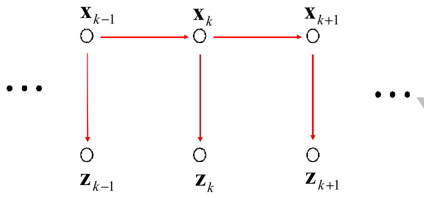
$p(\mathbf{z}_k\,|\,\mathbf{x}_k)$: *likelihood*

$p(\mathbf{x}_k\,|\,\mathbf{x}_{k-1})$: *temporal prior*

$p(\mathbf{x}_{k-1}\,|\,\mathbf{Z}_{k-1})$: posterior probability at previous time step

κ: normalizing term

Graphical Model



Markov assumptions:

$p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{Z}_{k-1}) = p(\mathbf{z}_k|\mathbf{x}_k)\,,\quad p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{Z}_{k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$

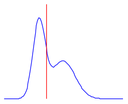
$p(\mathbf{x}_k\,|\,\mathbf{X}_{k-1}) = p(\mathbf{x}_k\,|\,\mathbf{x}_{k-1})$

Estimators

Assume the posterior probability $p(\mathbf{x}_k\,|\,\mathbf{Z}_k)$ is known:

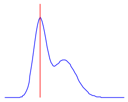
- posterior mean

$$\hat{\mathbf{x}}_k = E(\mathbf{x}_k\,|\,\mathbf{Z}_k)$$



- maximum *a posterior* (MAP)

$$\hat{\mathbf{x}}_k = \arg\max\, p(\mathbf{x}_k\,|\,\mathbf{Z}_k)$$

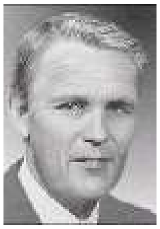


$p(\mathbf{x}_k\,|\,\mathbf{Z}_k)$

General Model

- $p(\mathbf{x}_k\,|\,\mathbf{Z}_k)$ can be an arbitrary, non-Gaussian, multi-modal distribution.
- The recursive equation has no explicit solution, but can be numerically approximated using Monte Carlo techniques.
- If both *likelihood* and *prior* are Gaussian, the solution has closed form and the two estimators (posterior mean & MAP) are the same. Such model is known as the Kalman filter. **(Kalman, 1960)**

Rudolph Emil Kalman



Born 1930 in Hungary

Professor Emeritus,
University of Florida

D.Sc., 1957,
Columbia University

His seminal paper:

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME, Journal of Basic Engineering*, *82*, 35–45

Broad Applications of Kalman Filter

- Engineering
  - Robotics, spacecraft, aircraft, automobiles
- Computer
  - Tracking, real-time graphics, computer vision
- Others
  - Forecasting economic indicators
  - Telephone and electricity loads
  - Encoding/decoding neural signals

Mathematical Properties of Kalman filter

- has a sound probabilistic framework
- makes explicit assumptions about the data and noise
- indicates the uncertainty of the estimate
- provides efficient estimation (closed-form and in real time)

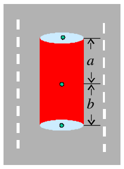
Likelihood Model

Generative model for the observation:

$$\boxed{\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{q}_k}$$

$$H_k \in R^{c\times d}, \quad \mathbf{q}_k \sim N(0, \mathbf{Q}_k), \quad \mathbf{Q}_k \in \mathbf{R}^{c\times c}, \quad k = 1,2,\cdots,M.$$

For example:



(x_k, y_k): centroid of car
(x_k^f, y_k^f): centroid of front bumper
(x_k^r, y_k^r): centroid of rear bumper

$$\begin{pmatrix} x_k^f \\ y_k^f \\ x_k^r \\ y_k^r \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & a \\ 1 & 0 & 0 \\ 0 & 1 & -b \end{pmatrix} \begin{pmatrix} x_k \\ y_k \\ 1 \end{pmatrix} + \begin{pmatrix} q_{k,1} \\ q_{k,2} \\ q_{k,3} \\ q_{k,4} \end{pmatrix}$$

Explicit Form

The likelihood model is equivalent to that

$$\mathbf{z}_k \sim N(\mathbf{H}_k \mathbf{x}_k, \mathbf{Q}_k)$$

The conditional probability has explicit form:

$$p(\mathbf{z}_k | \mathbf{x}_k) = \frac{1}{((2\pi)^c \det(\mathbf{Q}_k))^{1/2}} \exp(-\frac{1}{2}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)^T \mathbf{Q}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k))$$

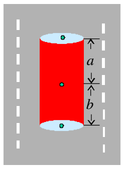
Temporal Prior

Temporal prior of the state:

$$\boxed{\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{w}_k}$$

$$\mathbf{A}_k \in \mathbf{R}^{d\times d}, \quad \mathbf{w}_k \sim N(0, \mathbf{W}_k), \quad \mathbf{W}_k \in \mathbf{R}^{d\times d}, \quad k = 2,3,\cdots M.$$

Same example:



(x_k, y_k): centroid of car

Assume the velocity is constant (v_x, v_y)

$$\begin{pmatrix} x_k \\ y_k \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & v_x\Delta t \\ 0 & 1 & v_y\Delta t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{k-1} \\ y_{k-1} \\ 1 \end{pmatrix} + \begin{pmatrix} w_{k,1} \\ w_{k,2} \\ 0 \end{pmatrix}$$

Explicit Form

The prior model is equivalent to that

$$\mathbf{x}_{k+1} \sim N(\mathbf{A}_k \mathbf{x}_k, \mathbf{W}_k)$$

The conditional probability has explicit form:

$$p(\mathbf{x}_{k+1} | \mathbf{x}_k) = \frac{1}{((2\pi)^d \det(\mathbf{W}_k))^{1/2}} \exp(-\frac{1}{2}(\mathbf{x}_{k+1} - \mathbf{A}_k \mathbf{x}_k)^T \mathbf{W}_k^{-1} (\mathbf{x}_{k+1} - \mathbf{A}_k \mathbf{x}_k))$$

Kalman Filter Model

Definition:

System Equation: (aka transition probabilities)

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{w}_k, \qquad \mathbf{w}_k \in N(0, \mathbf{W}_k)_{k=2,3,\cdots}$$

Measurement Equation: (aka observation probabilities)

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{q}_k, \qquad \mathbf{q}_k \in N(0, \mathbf{Q}_k)_{k=1,2,\cdots}$$

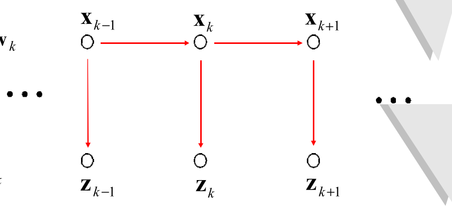
Assumption:

All random variables have Gaussian distributions and they are linearly related

---

Graphical Model

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{w}_k$$



$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{q}_k$$

$$p(\mathbf{X}_M, \mathbf{Z}_M) = p(\mathbf{X}_M)\, p(\mathbf{Z}_M \mid \mathbf{X}_M)$$
$$= [p(\mathbf{x}_1) \prod_{k=2}^{M} p(\mathbf{x}_k | \mathbf{x}_{k-1})][\prod_{k=1}^{M} p(\mathbf{z}_k | \mathbf{x}_k)]$$

Recursive Estimation

$$p(\mathbf{x}_k | \mathbf{Z}_k) = \kappa\, p(\mathbf{z}_k | \mathbf{x}_k) \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{Z}_{k-1}) d\mathbf{x}_{k-1}$$

Time update:

posterior at previous step:
$p(\mathbf{x}_{k-1} | \mathbf{Z}_{k-1})$

temporal prior:
$p(\mathbf{x}_k | \mathbf{x}_{k-1})$

prior distribution:
$p(\mathbf{x}_k | \mathbf{Z}_{k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{Z}_{k-1}) d\mathbf{x}_{k-1}$

Measurement update:

prior distribution:
$p(\mathbf{x}_k | \mathbf{Z}_{k-1})$

likelihood:
$p(\mathbf{z}_k | \mathbf{x}_k)$

posterior distribution:
$p(\mathbf{x}_k | \mathbf{Z}_k) = \kappa\, p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{Z}_{k-1})$

Basic Properties of Gaussian Distribution

$\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$ independent random vectors, $\mathbf{x} \sim N(0, \mathbf{A})$, $\mathbf{y} \sim N(0, \mathbf{B})$,

then, i) for any invertible matrix

$$\mathbf{C} \in \mathbf{R}^{d \times d}, \ \mathbf{Cx} \sim N(0, \mathbf{CAC}^T);$$

ii) $\mathbf{x} + \mathbf{y} \sim N(0, \mathbf{A} + \mathbf{B})$.

Kalman Filtering
Step I: Time Update

Assume: $\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1} \sim N(\hat{\mathbf{x}}_{k-1}, \mathbf{P}_{k-1})$

$$\Leftrightarrow \ \mathbf{x}_{k-1} = \hat{\mathbf{x}}_{k-1} + \mathbf{e}_{k-1}, \ \mathbf{e}_{k-1} \sim N(0, \mathbf{P}_{k-1})$$

System equation: $\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{w}_k, \ \mathbf{w}_k \sim N(0, \mathbf{W}_k)$

$$\Rightarrow \ \mathbf{x}_k = \mathbf{A}_k \hat{\mathbf{x}}_{k-1} + \mathbf{A}_k \mathbf{e}_{k-1} + \mathbf{w}_k$$

Use properties i) and ii):

$$\mathbf{A}_k \mathbf{e}_{k-1} + \mathbf{w}_k \sim N(0, \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{W}_k)$$

Let

$$\hat{\mathbf{x}}_k^- = \mathbf{A}_k \hat{\mathbf{x}}_{k-1}, \ \ \mathbf{P}_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{W}_k,$$

then,

$$\mathbf{x}_k \mid \mathbf{Z}_{k-1} \sim N(\hat{\mathbf{x}}_k^-, \mathbf{P}_k^-)$$

Kalman Filtering
Step II: Measurement Update

Time update:

$$p(\mathbf{x}_k \mid \mathbf{Z}_{k-1}) \propto \exp(-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)^T (\mathbf{P}_k^-)^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_k^-))$$

Measurement equation:

$$p(\mathbf{z}_k | \mathbf{x}_k) \propto \exp(-\frac{1}{2}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)^T \mathbf{Q}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k))$$

Recursive update:

(details omitted)

$$p(\mathbf{x}_k \mid \mathbf{Z}_k) \propto p(\mathbf{z}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{Z}_{k-1})$$
$$\propto \exp(-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T (\mathbf{P}_k)^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_k))$$

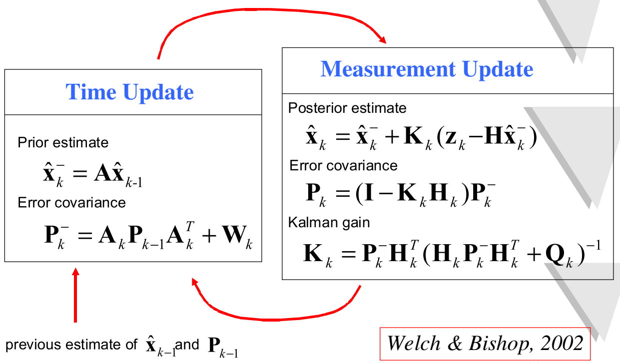
where, $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-), \ \ \mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-,$

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{Q}_k)^{-1}.$$

That is,

$$\mathbf{x}_k \mid \mathbf{Z}_k \sim N(\hat{\mathbf{x}}_k, \mathbf{P}_k)$$

Kalman Filter Algorithm



Time Update

Prior estimate

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1}$$

Error covariance

$$\mathbf{P}_k^- = \mathbf{A}_k\mathbf{P}_{k-1}\mathbf{A}_k^T + \mathbf{W}_k$$

Measurement Update

Posterior estimate

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-)$$

Error covariance

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-$$

Kalman gain

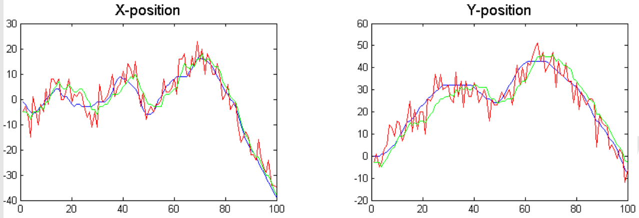
$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{Q}_k)^{-1}$$

previous estimate of $\hat{\mathbf{x}}_{k-1}$ and $\mathbf{P}_{k-1}$

Welch & Bishop, 2002

Estimation Examples

- Estimation by two approaches:
  - use only the likelihood;
  - use both likelihood and prior, the result looks smoother and more accurate.



blue: true, red: likelihood, green: Kalman filtering

Conclusion

- The Kalman filter is an extremely useful technique and is a fundamental tool in stochastic control theory.
- It is regularly rediscovered, and appears in different guises in different fields.
- It can be generalized by various non-linear and/or non-Gaussian models: extended Kalman filter, unscented Kalman filter, switching Kalman filter, mixture Kalman filter, particle filter, and so on.