Readings:
K&F: 6.1, 6.2, 6.3, 14.1, 14.2, 14.3, 14.4,

Kalman Filters
Gaussian MNs

Graphical Models – 10708
Carlos Guestrin
Carnegie Mellon University

December 1st, 2008

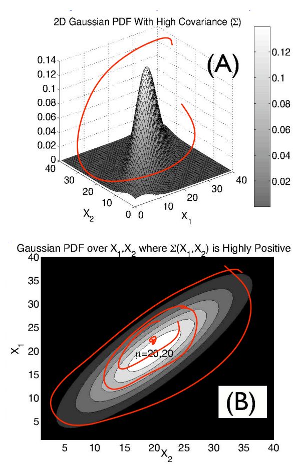
Multivariate Gaussian

$$p(X_1, \ldots, X_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

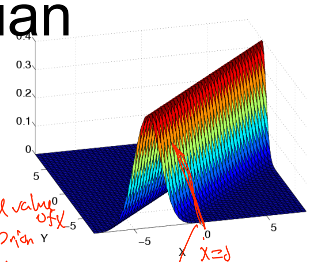
Mean vector: $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$

Covariance matrix: $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}$

$\sigma_{32} = \sigma_{23}$



Conditioning a Gaussian



- Joint Gaussian:
  - p(X,Y) ~ *N*(μ;Σ)
- Conditional linear Gaussian:
  - p(Y|X) ~ *N*($\mu_{Y|X}$; $\sigma^2_{Y|X}$) ← gaussian

$$\mu_{Y|X} = \mu_Y + \frac{\sigma_{YX}}{\sigma^2_X}(x - \mu_x)$$

$$\sigma^2_{Y|X} = \sigma^2_Y - \frac{\sigma^2_{YX}}{\sigma^2_X}$$

- observed value of X
- prior mean of X
- ↑ prior
- x=0
- P(Y|X=0)
- posterior variance doesn't depend on observed value!!
- posterior variance
- prior variance
- ≥ 0
- $\sigma^2_{Y|X} \le \sigma^2_Y$ ($\sigma^2_{Y|X} = \sigma^2_Y$ iff $Y \perp X$)
- observations always decrease variance

3

Gaussian is a "Linear Model"

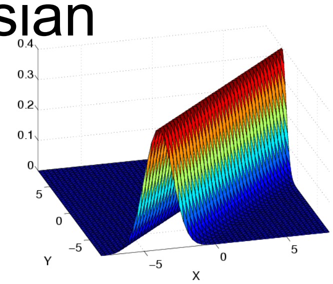
- Conditional linear Gaussian:
  - p(Y|X) ~ *N*(β_0+βX; σ^2)

$$\mu_{Y|X} = \mu_Y + \frac{\sigma_{YX}}{\sigma^2_X}(x - \mu_x)$$

$$\sigma^2_{Y|X} = \sigma^2_Y - \frac{\sigma^2_{YX}}{\sigma^2_X}$$

4

Conditioning a Gaussian



- Joint Gaussian:
  - p(X,Y) ~ *N*(μ;Σ)
- Conditional linear Gaussian:
  - p(Y|X) ~ *N*($\mu_{Y|X}$; $\Sigma_{YY|X}$)

$$\mu_{Y|X} = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(x - \mu_x)$$
$$\Sigma_{YY|X} = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$$

5

Conditional Linear Gaussian (CLG) – general case

- Conditional linear Gaussian:
  - p(Y|X) ~ *N*(β_0+BX; $\Sigma_{YY|X}$)

$$\mu_{Y|X} = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(x - \mu_x)$$
$$\Sigma_{YY|X} = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$$

6

Understanding a linear Gaussian – the 2d case

- Variance increases over time (motion noise adds up)
- Object doesn't necessarily move in a straight line

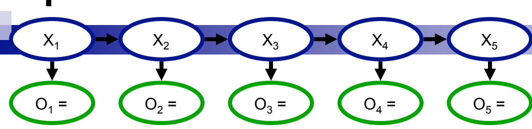
Tracking with a Gaussian 1

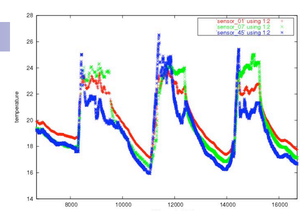
- $p(X_0) \sim N(\mu_0,\Sigma_0)$
- $p(X_{i+1}|X_i) \sim N(\mathrm{B}\, X_i + \beta;\, \Sigma_{Xi+1|Xi})$

Tracking with Gaussians 2 – Making observations

- We have $p(X_i)$
- Detector observes $O_i=o_i$
- Want to compute $p(X_i|O_i=o_i)$
- Use Bayes rule:
- Require a CLG observation model
  - $p(O_i|X_i) \sim N(W X_i + v; \Sigma_{Oi|Xi})$

Operations in Kalman filter





- Compute $p(X_t \mid O_{1:t} = o_{1:t})$
- Start with $p(X_0)$
- At each time step *t*:
  - **Condition** on observation
    $p(X_t \mid o_{1:t}) \propto p(X_t \mid o_{1:t-1}) p(o_t \mid X_t)$
  - **Prediction** (Multiply transition model)
    $p(X_{t+1}, X_t \mid o_{1:t}) = p(X_{t+1} \mid X_t) p(X_t \mid o_{1:t})$
  - **Roll-up** (marginalize previous time step)
    $p(X_{t+1} \mid o_{1:t}) = \int_{X_t} p(X_{t+1}, x_t \mid o_{1:t}) dx_t$
- I'll describe one implementation of KF, there are others
  - Information filter

Exponential family representation of Gaussian: Canonical Form

$$p(X_1, \ldots, X_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right\}$$

11

Canonical form

$$p(X_1, \ldots, X_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right\}$$
$$= K \exp\left\{\eta^T \mathbf{x} - \frac{1}{2}\mathbf{x}^T \Lambda \mathbf{x}\right\}$$

- Standard form and canonical forms are related:

$$\mu = \Lambda^{-1}\eta$$
$$\Sigma = \Lambda^{-1}$$

- Conditioning is easy in canonical form
- Marginalization easy in standard form

12

Conditioning in canonical form

$$p(X_t \mid o_{1:t}) \propto p(X_t \mid o_{1:t-1}) p(o_t \mid X_t)$$

- First multiply: $p(A,B) = p(A)p(B \mid A)$

$p(A): \quad \eta_1, \ \Lambda_1$

$p(B \mid A): \quad \eta_2, \ \Lambda_2$

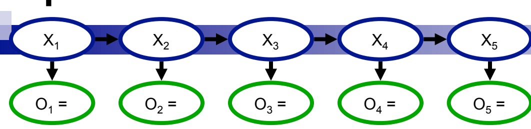
$p(A,B): \quad \eta_3 = \eta_1 + \eta_2, \ \Lambda_3 = \Lambda_1 + \Lambda_2$

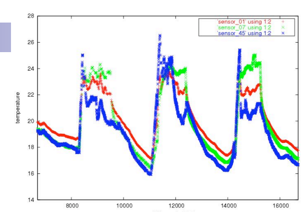
- Then, condition on value B = y $\quad p(A \mid B = y)$

$$\eta_{A|B=y} = \eta_A - \Lambda_{AB} \cdot y$$

$$\Lambda_{AA|B=y} = \Lambda_{AA}$$

Operations in Kalman filter





- Compute $\quad p(X_t \mid O_{1:t} = o_{1:t})$
- Start with $p(X_0)$
- At each time step *t*:
  - **Condition** on observation
    $p(X_t \mid o_{1:t}) \propto p(X_t \mid o_{1:t-1}) p(o_t \mid X_t)$
  - **Prediction** (Multiply transition model)
    $p(X_{t+1}, X_t \mid o_{1:t}) = p(X_{t+1} \mid X_t) p(X_t \mid o_{1:t})$
  - **Roll-up** (marginalize previous time step)
    $p(X_{t+1} \mid o_{1:t}) = \int_{X_t} p(X_{t+1}, x_t \mid o_{1:t}) dx_t$

Prediction & roll-up in canonical form

$$p(X_{t+1} \mid o_{1:t}) = \int_{X_t} p(X_{t+1} \mid x_t) p(x_t \mid o_{1:t}) dx_t$$

- First multiply: $p(A,B) = p(A)p(B \mid A)$
- Then, marginalize X_t: $p(A) = \int_B p(A,b)db$

$$\begin{aligned} \eta_A^m &= \eta_A - \Lambda_{AB}\Lambda_{BB}^{-1}\eta_B \\ \Lambda_{AA}^m &= \Lambda_{AA} - \Lambda_{AB}\Lambda_{BB}^{-1}\Lambda_{BA} \end{aligned}$$

What if observations are not CLG?

- Often observations are not CLG
  - CLG if $O_i = B X_i + \beta_o + \varepsilon$
- Consider a motion detector
  - O_i = 1 if person is likely to be in the region
  - Posterior is not Gaussian

Linearization: incorporating non-linear evidence

- $p(O_i|X_i)$ not CLG, but…
- Find a Gaussian approximation of $p(X_i,O_i)= p(X_i)\ p(O_i|X_i)$
- Instantiate evidence $O_i=o_i$ and obtain a Gaussian for $p(X_i|O_i=o_i)$
- Why do we hope this would be any good?
  - Locally, Gaussian may be OK

17

Linearization as integration

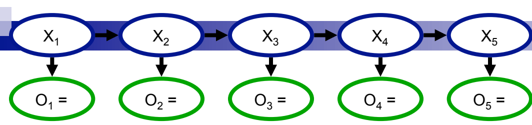
- Gaussian approximation of $p(X_i,O_i)= p(X_i)\ p(O_i|X_i)$
- Need to compute moments
  - $E[O_i]$
  - $E[O_i^2]$
  - $E[O_i\ X_i]$
- Note: Integral is product of a Gaussian with an arbitrary function

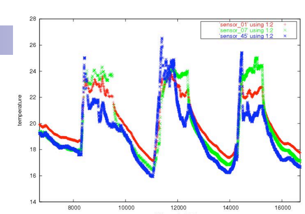
18

Linearization as numerical integration

- **Product of a Gaussian with arbitrary function**
- Effective numerical integration with **Gaussian quadrature** method
  - Approximate integral as **weighted sum over integration points**
  - Gaussian quadrature defines location of points and weights
- Exact if arbitrary function is **polynomial of bounded degree**
- **Number of integration points exponential** in number of dimensions *d*
- **Exact monomials** requires exponentially fewer points
  - For **2*d*+1 points**, this method is equivalent to effective **Unscented Kalman filter**
  - **Generalizes to many more points**

Operations in non-linear Kalman filter





- Compute $p(X_t \mid O_{1:t} = o_{1:t})$
- Start with $p(X_0)$
- At each time step *t*:
  - **Condition** on observation (use **numerical integration**)
    $$p(X_t \mid o_{1:t}) \propto p(X_t \mid o_{1:t-1}) p(o_t \mid X_t)$$
  - **Prediction** (Multiply transition model, use **numerical integration**)
    $$p(X_{t+1}, X_t \mid o_{1:t}) = p(X_{t+1} \mid X_t) p(X_t \mid o_{1:t})$$
  - **Roll-up** (marginalize previous time step)
    $$p(X_{t+1} \mid o_{1:t}) = \int_{X_t} p(X_{t+1}, x_t \mid o_{1:t}) dx_t$$

Canonical form & Markov Nets

$$p(X_1, \ldots, X_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$
$$= K \exp\left\{\eta^T\mathbf{x} - \frac{1}{2}\mathbf{x}^T\Lambda\mathbf{x}\right\}$$

21

What you need to know about Gaussians, Kalman Filters, Gaussian MNs

- **Kalman filter**
  - Probably most used BN
  - Assumes Gaussian distributions
  - Equivalent to linear system
  - Simple matrix operations for computations
- **Non-linear Kalman filter**
  - Usually, observation or motion model not CLG
  - Use numerical integration to find Gaussian approximation
- **Gaussian Markov Nets**
  - Sparsity in precision matrix equivalent to graph structure
- **Continuous and discrete (hybrid) model**
  - Much harder, but doable and interesting (see book)

22