Kalman Filters

- Prelim Review: Monday 4/3 6pm Upson 109
- Final Exam: 5/15/06, 7-9:30 pm, HO 401

OUTLINE:
- intro
- motivating example & derivation
- full discrete KF algorithm
- Matlab demo

INTRODUCTION
- popular model for Stochastic Estimation:
- estimate state of a system from noisy observations
- System: i) initial state distribution
  ii) transition model
  iii) sensor model
  all based on Normal distribution

Normal Distribution (Gaussian)
- continuous distribution over $(-\infty, +\infty$)
- parameters: Mean $(\mu) \in (-\infty, +\infty)$
  Variance $(\sigma^2) \in (0, +\infty)$
  $N(\mu, \sigma^2)$
- **Distribution function (pdf)**:

\[
f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

- **Additivity of independent variables**:

\[
N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)
\]

- **Central Limit Theorem**:

\[
\{X_i\} iid \text{ random variables, with } E(X_i) = \mu, \text{ Var}(X_i) = \sigma^2 \quad (\text{ANY distribution with } \mu < \infty, \sigma^2 < \infty)
\]

\[
S_n = \frac{\sum X_i}{n}, \text{ then } \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{n \to \infty} N(0,1)
\]

- **Multivariate Normal Distribution**: \(N(\mu, \Sigma)\)

- Higher-dimensional generalization of Normal

- Random vector \(\mathbf{X} = (X^{(1)}, \ldots, X^{(d)})\)

\[
\mu = (E(X^{(1)}), \ldots, E(X^{(d)}))
\]

\[
\Sigma = \text{cov} \{\mathbf{X}, \mathbf{X}\} = (E(X^{(i)} - \mu^{(i)})(X^{(j)} - \mu^{(j)}))_{i,j}
\]
Kalman Filter

i) initial distribution: \( P(w_0) \sim N(\mu_0, \Sigma_0) \)

ii) transition model: \( P(w_{t+1} | w_t) \sim N(\cdot, \Sigma) \)

iii) sensor model: \( P(z_t | w_t) \sim N(\cdot, \Sigma) \)

- posterior probability \( P(w_t) \) stays \( N(\mu_t, \Sigma_t) \) for all \( t \)
- continuous state & evidence, discrete time
  (discrete KF)

EXAMPLE & MODEL

DERIVATION

- lost at an unknown location \( x(t) \) on a boat

- 2 ways to estimate location: (assume normal error; \( Z = z + N(0, \sigma^2) \))
  - you (amateur) \( Z_t \sim N(\mu_1, \sigma_1^2) \)
  - friend (skilled) \( Z_t \sim N(\mu_2, \sigma_2^2) < \sigma_1^2 \)

1) you estimate at time \( t_1 \): \( Z_1 = Z_{1,t} \)

2) best estimate of the position:
   - mode \( \hat{x}(t) = Z_1 \)
   - median \( \hat{x}(t) = Z_1 \)
   - mean \( \hat{x}(t) = Z_1 \)
2) provide estimates at the same time: $z_s = z_t(t)$

$\Rightarrow$ how is the best estimate of $k^o(t)$?

how is the new information incorporated?
Model derivation (static)

- Linearly combine the observations:
  \[ \hat{x} = m \cdot z_1 + (1-m) \cdot z_2 \]
  
  \( m \) is unknown weight to be calculated.

\[ \Rightarrow \quad \sigma_x^2 = m^2 \sigma_1^2 + (1-m)^2 \sigma_2^2 \]

- Find \( m \) that minimizes the uncertainty:
  \[ \frac{\partial \sigma_x^2}{\partial m} = 0 \]

\[ 2m \sigma_1^2 - 2 \sigma_x^2 + 2m \sigma_2^2 = 0 \]

\[ \Rightarrow \quad m = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \]

So

\[ \sigma_x^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \]
OBSERVATIONS:

1) \( \hat{\alpha} \) vs. \((\alpha^2, \alpha^2)\) nicely follows intuition
\[ \alpha^2 = \alpha^2 \rightarrow \hat{\alpha} = \frac{1}{2} (\alpha_1 + \alpha_2) \]
\[ \alpha^2 \gg \alpha^2 \rightarrow \hat{\alpha} \sim \alpha_1 \]
\[ \alpha^2 \gg \alpha^2 \rightarrow \hat{\alpha} \sim \alpha_2 \]

2) \( \hat{\alpha} \) is smaller than both \( \alpha^2 \) and \( \alpha^2 \)
(follows from \( \frac{1}{\hat{\alpha}^2} = \frac{1}{\alpha^2} + \frac{1}{\alpha^2} \))

\( \Rightarrow \) any information is used (even if very noisy)

3) such \( \alpha \) also makes \( \hat{\alpha} \) minimize squared weighted distances from \( \alpha_1 \) and \( \alpha_2 \) to any \( \alpha \):
\[ \hat{\alpha} = \arg \min_{\alpha} \sum_{i=1}^{\alpha} \left( \frac{2\alpha_i - \alpha}{\alpha_i^2} \right)^2 \]
\[ \frac{\partial}{\partial \alpha} = \sum_{i=1}^{\alpha} -2 \frac{2\alpha_i - \alpha}{\alpha_i^2} (2\alpha_i - \alpha) = 0 \]
\[ \alpha_i^2 (2\alpha_i - \alpha) + \alpha_i^2 (2\alpha_i - \alpha) = 0 \]
\[ \alpha_i^2 \alpha_1 + \alpha_i^2 \alpha_2 = \alpha \left( \alpha_1^2 + \alpha_2^2 \right) \]
\[ m \alpha_1 + (1-m) \alpha_2 = \alpha \]

RECURSIVE FORMULATION

\[ \hat{\alpha} = m \hat{\alpha}_{\text{prev}} + (1-m) \hat{\alpha} = \hat{\alpha}_{\text{prev}} + (1-m) \frac{\left( \frac{\hat{\alpha}_2}{\hat{\alpha}_{\text{prev}}^2} \right)}{\hat{\alpha}_{\text{prev}}^2 + \hat{\alpha}^2} (\alpha_2 - \hat{\alpha}_{\text{prev}}) \]

Update gain \( k = \frac{\hat{\alpha}_{\text{prev}}^2}{\hat{\alpha}_{\text{prev}}^2 + \hat{\alpha}^2} \)

\( \Rightarrow \hat{\alpha} = \hat{\alpha}_{\text{prev}} + k (\alpha_2 - \hat{\alpha}_{\text{prev}}) \)
\[ \hat{\alpha}^2 = (1-k) \hat{\alpha}_{\text{prev}}^2 \]
Model derivation (dynamic)

- Similar situation as before, but the boat is moving with speed \( N \sim N(\bar{u}_N, \sigma_N^2) \) \( \bar{u} = \bar{u}_m + N \).

- Another measurement is done at time \( t_2 > t_1 \):
  \[ z_3 = z_3(t_3) \text{ with } \sigma_3^2 \]

- What is \( \hat{z}(t_2) \)?

Let \( t_2^- = t_2^- \) be time just before \( z_3 \) is taken.

Prediction:

\[
\begin{align*}
\hat{z}(t_2^-) &= \hat{z}(t_1) + \theta_\sigma(z_2^- - t_1) \\
\sigma_\theta^2(t_2^-) &= \sigma_\sigma^2(t_1) + \sigma^2_N \left( t_2^- - t_1 \right)^2
\end{align*}
\]

Observation \( z_3 \): again, we need to combine 2 Gaussians \( (z_3, \sigma_3) \) and \( (\hat{z}(t_2^-), \sigma_\theta^2(t_2^-)) \).

Correction:

\[
\begin{align*}
\hat{z}(t_2) &= \hat{z}(t_2^-) + K(z_2 - \hat{z}(t_2^-)) \\
\sigma^2(t_2) &= (1 - K) \sigma_\theta^2(t_2^-)
\end{align*}
\]

where

\[
K = \frac{\sigma_\theta^2(t_2^-)}{\sigma_\theta^2(t_2^-) + \sigma_3^2}
\]

Observations:

1) \( K \) and \( \sigma^2(t_2) \) does not depend on \( z_3 \), can be precomputed before observations are taken.

2) The correction step again makes an optimal decision between how much to trust the new observation vs. the prediction.
DISCRETE KALMAN FILTER

\[ \begin{align*}
x & \quad \text{system state} \\
\mathbf{u} & \quad \text{(optional) control input} \\
\mathbf{z} & \quad \text{observation (measurement)} \\
\mathbf{f} & \quad \text{state transition matrix} \\
\mathbf{b} & \quad \text{control input matrix} \\
\mathbf{w} & \quad \text{transition noise } \sim \mathcal{N}(0, \mathbf{Q}) \\
\mathbf{r} & \quad \text{observation relation} \\
\mathbf{n} & \quad \text{observation noise } \sim \mathcal{N}(0, \mathbf{R}) \\
\end{align*} \]

MODELS:
transition model:
\[ x(t) = F x(t-1) + B u(t) + w \]

sensor model:
\[ z(t) = H x(t) + n \]

ASSUMPTIONS:
(i) linear models (both transition & sensor)
(ii) uncertainty Gaussian (normally distributed)
(iii) white (uncorrelated in time)
Algorithm:

\( \hat{x}(t) \) .... estimate of \( x(t) \)
\( P(t) \) .... covariance matrix of \( \hat{x}(t) \) (uncertainty)

**INPUT:** \( \hat{x}(t-1), P(t-1), u(t-1) \)
**OUTPUT:** \( \hat{x}(t), P(t) \)

prediction: \[
\hat{x}(t^-) = F \hat{x}(t-1) + Bu(t) \\
P(t^-) = FP(t-1)F^T + Q
\]

correction: \[
\hat{x}(t) = \hat{x}(t^-) + K(Z(t) - H\hat{x}(t^-)) \\
P(t) = (I - KH)P(t^-)
\]

\( K = P(t^-)H^T S^{-1} \)
\( S = R + H P(t^-) H^T \)

**INITIAL ESTIMATE:** \( \hat{x}(0), P(0) \)
Extensions:
- Kalman Smoothing
- Extended KF (nonlinear transition & sensor models)
  - locally linearized using Hessian
- Switching KF

REFERENCES
An Introduction to the Kalman Filter, SIGGRAPH 2001 Course, Greg Welch and Gary Bishop
Kalman filtering chapter from Stochastic Models, Estimation, by Peter Maybeck
http://en.wikipedia.org/wiki/Kalman_filter