Kalman Filters

- Prelim Review: Monday 4/3 6pm Uson 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:
\[ N\left(\mu_1, \sigma_1^2\right) + N\left(\mu_2, \sigma_2^2\right) = N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right) \]

- Central Limit Theorem:
  \[ \{X_i\} \text{ iid random variables with } E(X_i) = \mu, \text{ Var}(X_i) = \sigma^2 \]
  \[ S_n = \frac{\sum_{i=1}^{n} X_i}{n}, \text{ then } \frac{S_n - n \mu}{\sqrt{n \sigma^2}} \xrightarrow{n \to \infty} N(0,1) \]
  ANR distribution with \( \mu < \infty, \sigma^2 < \infty \)

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

- Higher-dimensional generalization of Normal

- Random vector \( \mathbf{X} = (X^{(1)}, \ldots, X^{(k)}) \)
  \[ \mathbf{\mu} = (E(X^{(1)}), \ldots, E(X^{(k)})) \]
  \[ \Sigma = \text{cov}(\mathbf{X}, \mathbf{X}) = (E((X^{(i)} - \mu^{(i)})(X^{(j)} - \mu^{(j)})))_{i,j} \]
**Kalman Filter**

- Initial distribution: \( P(\mathbf{w}_0) \sim \mathcal{N}(\mu_0, \Sigma_0) \)

- Transition model: \( P(\mathbf{w}_{t+1} | \mathbf{w}_t) \sim \mathcal{N}(\cdot, \Sigma_w) \)

- Sensor model: \( P(\mathbf{z}_t | \mathbf{w}_t) \sim \mathcal{N}(\cdot, \Sigma_z) \)

- Posterior probability \( P(\mathbf{w}_t) \) stays \( \mathcal{N}(\mu_t, \Sigma_t) \) for all \( t \)

- Continuous state & evidence, discrete time

**Example & Model Derivation**

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

- 2 ways to estimate location: (Assume normal error: \( Z = z + \mathcal{N}(\mu, \sigma^2) \))

  - You (amateur) \( Z = z_1, \sigma_1^2 \)
  - Friend (skilled) \( Z = z_2, \sigma_2^2 < \sigma_1^2 \)

1) You estimate at time \( t_1 \): \( Z_1 = z_1(t_1) \)

\[
\begin{align*}
\int_{x(t_1)} f_{x(t_1) | z(t_1)}(x | z_1) dx
\end{align*}
\]

\( \Rightarrow \) Best estimate of the position:

- **Mode** \( \hat{x}(t) = Z_1 \)
- **Median** \( \hat{\sigma}^2(t) = \sigma_1^2 \)
- **Mean**
2) Train estimates at the same time: \( z_2 = z_2(t_1) \)

\[ f_{x(t_2)|z(t_2)}(x|z_2) \]

\[ \sigma_{z_2} \]

\[ x_1 \quad x_2 \]

\[ \Rightarrow \text{how what is the best estimate of } k^0(t_1) ? \]

\[ \text{how is the new information incorporated?} \]
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) \]

\[ \Rightarrow \quad \sigma_{\hat{x}}^2 = m^2 \sigma_1^2 + (1-m)^2 \sigma_2^2 \]

- find \( m \) that minimizes the uncertainty: \( \frac{\partial \sigma_{\hat{x}}^2}{\partial m} = 0 \)
  \[ 2m \sigma_1^2 - 2 \sigma_2^2 + 2m \sigma_2^2 = 0 \]
  \[ \Rightarrow \quad m = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \]

So \( \sigma_{\hat{x}}^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \)
OBSERVATIONS:

1) $\hat{\lambda}$ is nicely follows intuition

\[ \bar{d}^2 = \bar{d}_1^2 \rightarrow \hat{\lambda} = \frac{1}{2} (Z_1 + Z_2) \]

\[ \bar{d}_2^2 \gg \bar{d}_1^2 \rightarrow \hat{\lambda} \sim Z_1 \]

\[ \bar{d}_1^2 \gg \bar{d}_2^2 \rightarrow \hat{\lambda} \sim Z_2 \]

2) $\bar{d}^2$ is smaller than both $\bar{d}_1^2$ and $\bar{d}_2^2$

(follows from $\frac{1}{\bar{d}^2} = \frac{1}{\bar{d}_1^2} + \frac{1}{\bar{d}_2^2}$)

⇒ any information is used (even if very noisy)

3) such $\bar{d}$ also makes $\hat{\lambda}$ minimize squared weighted distances from $Z_1$ and $Z_2$ to any $\lambda$:

\[ \hat{\lambda} = \arg \min_{\lambda} \frac{1}{2} \sum_i \left( \frac{2\bar{d}_i - \lambda}{\bar{d}_i} \right)^2 \]

\[ \frac{\partial}{\partial \lambda} = \frac{3\bar{d}_i}{\bar{d}_i^2} \left( 2\bar{d}_i - \lambda \right) = 0 \]

\[ \bar{d}_2^2 (2\bar{d}_1 - \lambda) + \bar{d}_1^2 (2\bar{d}_2 - \lambda) = 0 \]

\[ \bar{d}_2^2 2\bar{d}_1 + \bar{d}_1^2 2\bar{d}_2 = \lambda (\bar{d}_1^2 + \bar{d}_2^2) \]

\[ m \bar{d}_1 + (1-m) \bar{d}_2 = \lambda \]

RECURSIVE FORMULATION

\[ \hat{\lambda} = m \hat{\lambda}_\text{prev} + (1-m) \bar{d}_2 = \hat{\lambda}_\text{prev} + (1-m) \frac{(Z_2 - \hat{\lambda}_\text{prev})}{\bar{d}_2} \]

\[ \text{Update gain } \hat{K} = \left( \frac{\bar{d}_2^2}{\bar{d}_2^2 + \bar{d}_1^2} \right) \]

⇒ $\hat{\lambda} = \hat{\lambda}_\text{prev} + \hat{K} (Z_2 - \hat{\lambda}_\text{prev})$

$\bar{d}_2^2 = (1-\hat{K}) \bar{d}_2^2$
Model derivation (dynamic)

- Similar situation as before, but the boat is moving with speed \( \mathbf{v} \sim \mathcal{N}(\mathbf{m}_o, \mathbf{V}_o) \), \( \mathbf{m} = \mathbf{m}_m + \mathbf{m}_o \)
- Another measurement is done at time \( t_2 > t_1 \)

\[ Z_3 = Z_3(t_3) \quad \text{with} \quad \sigma_3^2 \]

- What is \( \hat{\mathbf{x}}(t_2) \)?

Let \( t_2^-(=t_2^-) \) be time just before \( Z_3 \) is taken

**Prediction:** \( \hat{\mathbf{x}}(t_2^-) = \hat{\mathbf{x}}(t_1) + \mathbf{m}_o(t_2^- - t_1) \)

\[ \sigma^2(t_2^-) = \sigma^2(t_1) + \sigma_o^2(t_2^- - t_1)^2 \]

- Observation: \( Z_3 \) : again, we need to combine 2 Gaussians \((Z_3, \sigma_3^2)\) and \((\hat{\mathbf{x}}(t_2^-), \sigma^2(t_2^-))\)

**Correction:** \( \hat{\mathbf{x}}(t_2) = \hat{\mathbf{x}}(t_2^-) + K(Z_3 - \hat{\mathbf{x}}(t_2^-)) \)

\[ \sigma^2(t_2) = (1 - K) \sigma^2(t_2^-) \]

where \( K = \frac{\sigma^2(t_2^-)}{\sigma^2(t_2^-) + \sigma_3^2} \)

**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**

- \( x \): System state
- \( u \): (optional) control input
- \( z \): Observation (measurement)
- \( F \): State transition matrix
- \( B \): Control input matrix
- \( w \): Transition noise \( w \sim \mathcal{N}(0, Q) \)
- \( H \): Observation relation
- \( v \): Observation noise \( v \sim \mathcal{N}(0, R) \)

**Models:**

1. **Transition model:**
   \[ x(t) = F x(t-1) + B u(t) + w \]

2. **Sensor model:**
   \[ z(t) = H x(t) + v \]

**Assumptions:**

1. Linear models (both transition & sensor)
2. Uncertainty Gaussian (normally distributed)
3. White (uncorrelated in time)
Algorithm:
\[ \hat{x}(t) \ldots \text{estimate of } x(t) \]
\[ P(t) \ldots \text{covariance matrix of } \hat{x}(t) \ (\text{uncertainty}) \]

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

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

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

(K is very ugly...)

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