OPTIMAL FILTERING

LECTURE 6

- Smoothing
  - Using an extended state vector
  - General smoothing formulas
- Introduction to non-linear filtering

Reading instructions: Kailath, Sect. 10.1-10.4.

Arulampalam, S., Maskell, S., Gordon, N. and Clapp, T. (2002) "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking". IEEE Trans. SP, 50(2):174–188.

F. Gustafsson, "Particle filter theory and practice with positioning applications," IEEE A&E Syst. Mag. Part II: Tutorials, vol. 25, no. 7, pp. 53–82, 2010.

SMOOTHING

Often, an estimation delay is acceptable, or off-line processing can be done $\Longrightarrow$ More (future) data can be used

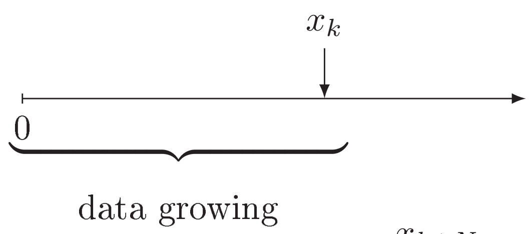
☺ More accurate estimates

☹ Increased computational complexity

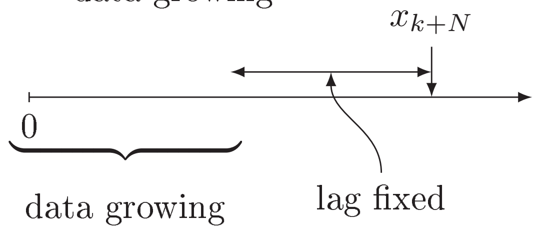
How to extend Kalman?

Different Forms of Smoothing

Filtering $\hat{x}_{k|k}$

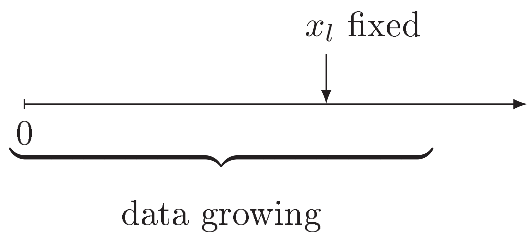


Prediction $\hat{x}_{k+N|k}$



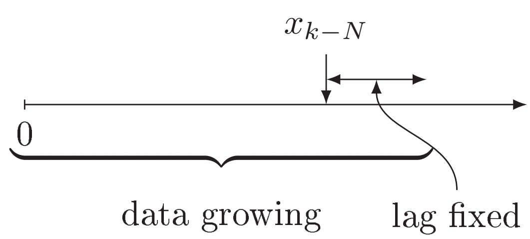
Fixed point smoothing

$\hat{x}_{l|k}$, l fixed, k growing



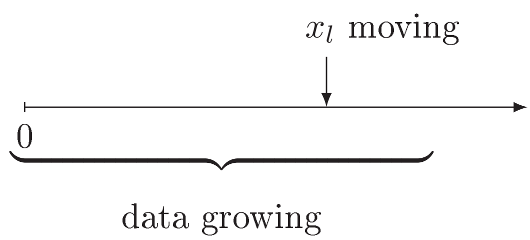
Different Forms of Smoothing, cont.

Fixed lag smoothing $\hat{x}_{k-N|k}$,



Fixed interval smoothing

$\hat{x}_{l|k}$, $k = M$, $0 \leq l \leq M$



Applications

Fixed point smoothing: Initial condition estimation

Fixed lag smoothing: on-line processing with delay

Fixed interval smoothing: off-line processing

. . .

Multiple-step Ahead Predictor

$$\hat{x}_{k+N|k} = F_{k+N-1}F_{k+N-2}\cdots F_k\hat{x}_{k|k}$$

Show!

Fixed Point Smoothing

Problem: For all $k > j$, j fixed, find $\hat{x}_{j|k}$ and
$P_k^{aa} = E[(x_j - \hat{x}_{j|k})(x_j - \hat{x}_{j|k})^*]$.

Solution: (assuming $S_k = 0$)

$$K_k^a = P_k^a H_k^* (H_k P_{k|k-1} H_k^* + R_k)^{-1}$$
$$\hat{x}_{j|k} = \hat{x}_{j|k-1} + K_k^a (y_k - H_k \hat{x}_{k|k-1})$$
$$P_{k+1}^a = P_k^a F_k^* - K_k^a H_k P_{k|k-1} F_k^*$$
$$P_{k+1}^{aa} = P_k^{aa} - K_k^a H_k P_k^a$$

Initial conditions: $\hat{x}_{j|j-1} = \hat{x}_{j|j-1}$, $P_j^{aa} = P_j^a = P_{j|j-1}$.

Improvement: $P_k^{aa} - P_{j|j-1} = \sum_{i=j}^{k-1} K_j^a H_j P_j^a$

Fixed Point Smoothing, Comments

- What happens when $R \to \infty$?
- Complexity?
- Most improvement obtained after 2-3 time constants of the system.
- Time varying even for time invariant systems!

Fixed Lag Smoothing

Problem: N fixed, k growing. Find $\hat{x}_{k-N|k}$ and
$P_{k-N|k} = E[(x_{k-N} - \hat{x}_{k-N|k})(x_{k-N} - \hat{x}_{k-N|k})^*]$.

Solution: For $i = 1, 2, \ldots, N$

$$K_k^{(i+1)} = P_{k|k-1}^{(i)} H_k^* (H_k P_{k|k-1} H_k^* + R_k)^{-1}$$

$$\hat{x}_{k-i|k} = \hat{x}_{k-i|k-1} + K_k^{(i+1)} e_k$$

$$P_{k+1|k}^{(i+1)} = P_{k|k-1}^{(i)} (F_k - K_k H_k)^*$$

$$P_{k-i|k} = P_{k-i|k-1} - P_{k|k-1}^{(i)} H_k^* \left(K_k^{(i+1)}\right)^*$$

Initialization: $\hat{x}_{0|-1} = 0$, $\hat{x}_{-i|-1} = 0$, $P_{k|k-1}^{(0)} = P_{k|k-1}$.

- Time invariant for time invariant systems!
- Complexity?

Fixed Interval Smoothing

Problem: M fixed, $0 \leq k \leq M$, find $\hat{x}_{k|M}$ and
$P_{k|M} = E[(x_k - \hat{x}_{k|M})(x_k - \hat{x}_{k|M})^*]$.

Solution 1: Fixed lag smoother with fixed lag $N = M$ and $k = M$.

Solution 2: Ordinary Kalman, $0 \leq k \leq M$, save $\hat{x}_{j|j-1}$, $\hat{x}_{j|j}$, $P_{j|j-1}$, $P_{j|j}$, followed by a backward recursion $j = M, M-1, \ldots, 1$:

$$\hat{x}_{j-1|M} = \hat{x}_{j-1|j-1} + P_{j-1|j-1} F_{j-1}^* P_{j|j-1}^{-1} (\hat{x}_{j|M} - \hat{x}_{j|j-1})$$

$$A_{j-1} = P_{j-1|j-1} F_{j-1}^* P_{j|j-1}^{-1}$$

$$P_{j-1|M} = P_{j-1|j-1} + A_{j-1}(P_{j|M} - P_{j|j-1}) A_{j-1}^*$$

- 2^{nd} approach requires less work
- Normally, fixed lag smoother with sufficiently large lag is adequate.

General Smoothing Formulas

"Bryson-Frazier" formulas (1963) $\quad 0 \leq k \leq N$

$$F_{p,k} = F_k - K_k H_k$$

$$\lambda_{k|N} = F_{p,k}^* \lambda_{k+1|N} + H_k^* R_{e_k}^{-1} e_k, \qquad \lambda_{N+1|N} = 0$$

$$\hat{x}_{k|N} = \hat{x}_{k|k-1} + P_{k|k-1} \lambda_{k|N}$$

$$\Lambda_{k|N} = F_{p,k}^* \Lambda_{k+1|N} F_{p,k} + H_k^* R_{e_k}^{-1} H_k, \qquad \Lambda_{N+1|N} = 0$$

$$P_{k|N} = P_{k|k-1} - P_{k|k-1} \Lambda_{k|N} P_{k|k-1}$$

KTH Electrical Engineering

Two-pass algorithm

(1) Forward $\hat{x}_{k|k-1}$, $P_{k|k-1}$

(2) Backward $\lambda_{k|N}$, $\Lambda_{k|N}$

(3) Combine to obtain $\hat{x}_{k|N}$, $P_{k|N}$

Filtered version (using $\hat{x}_{k|k}$ and $P_{k|k}$), see Kailath page 374.

Alternative form

Rauch-Tung-Striebel (RTS) Formulas (1965)

$$\hat{x}_{k|N} = \hat{x}_{k|k} + P_{k|k-1} F_{p,k}^* P_{k+1|k}^{-1} (\hat{x}_{k+1|N} - \hat{x}_{k+1|k})$$

$$P_{k|N} = P_{k|k} + P_{k|k-1} F_{p,k}^* P_{k+1|k}^{-1} (P_{k+1|N} - P_{k+1|k}) P_{k+1|k}^{-1} F_{p,k} P_{k|k-1}$$

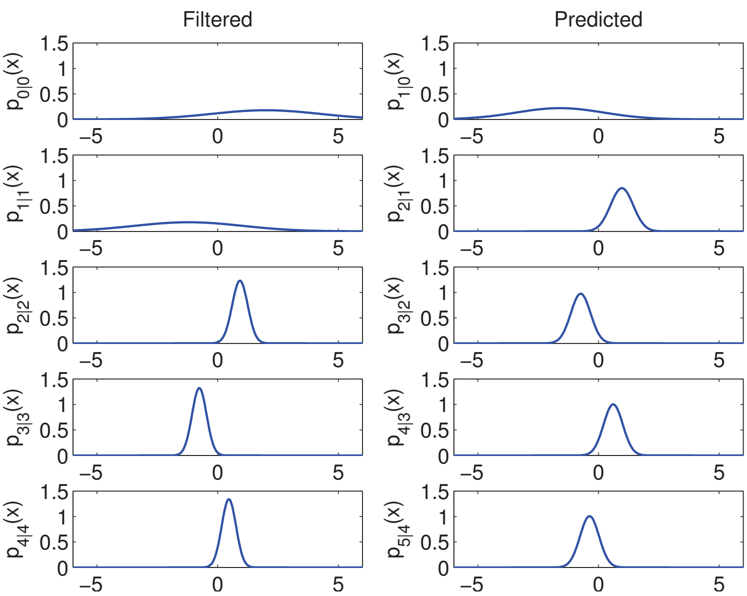
KTH Electrical Engineering

Also two-pass algorithm

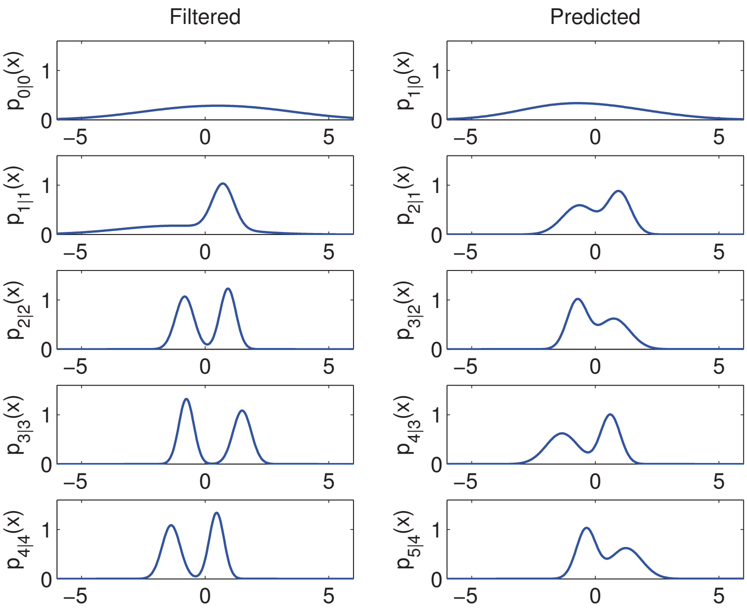
Non-Linear Filtering

- Assumption of linearity + Gaussianity often a rough approximation of reality!
- A single estimate of x_k is not always that useful.
- Full posterior distribution of x_k given observations contains more information, in general.
- Alternative view of Kalman:
  - Recursive calculation of posterior distributions $p(x_k|\{y_1, \ldots, y_{k-1}\})$ and $p(x_k|\{y_1, \ldots, y_{k-1}\})$
  - $p(x_k|\{y_1, \ldots, y_{k-1}\}) \in \mathcal{C}N(\hat{x}_{k|k-1}, P_{k|k-1})$
  - $p(x_k|\{y_1, \ldots, y_k\}) \in \mathcal{C}N(\hat{x}_{k|k}, P_{k|k})$
  - Parametric representation: $\hat{x}_{k|m}$ and $P_{k|m}$ specifies the Gaussian distribution fully

Kalman Filter – Alternative View



Non-Gaussian Example



General Model

$$x_{k+1} = f(x_k, w_k)$$
$$y_k = h(x_k) + v_k$$

More general:

$$f(x_{k+1}, x_k, w_k) = 0$$
$$h(y_k, x_k, v_k) = 0$$

Even more general:

$$x_{k+1} \sim p(x_{k+1}|x_k)$$
$$y_k \sim p(y_k|x_k)$$

General Bayesian Solution

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}$$

$$p(y_k|y_{1:k-1}) = \int p(y_k|x_k)p(x_k|y_{1:k-1})\, dx_k$$

$$p(x_{k+1}|y_{1:k}) = \int p(x_{k+1}|x_k)p(x_k|y_{1:k})\, dx_k$$

KTH Electrical Engineering

Problems:

- How represent the distributions?
- How calculate the integrals?

Alternative Representations of Distributions

KTH Electrical Engineering

- Parametric. Only possible for Gaussian + a few more $\Longrightarrow$ alternative derivation of Kalman
- Approximate by Gaussian, update mean and variance. Many possibilities.
- Discrete representation
  - Fixed grid
  - Stochastic grid (Monte Carlo, ...)
- Gaussian mixture model – linear combination of a number of Gaussian.

Importance Sampling

- Standard Monte Carlo:

$$\mathrm{E}[g(x)] = \int g(x)p(x)\, dx \approx \frac{1}{N}\sum_{n=1}^{N} g(x_n)$$

  where x_n i.i.d. drawn from distribution $p(x)$.

- Importance Sampling: Introduce suitable proposal density $q(x)$:

$$\mathrm{E}[g(x)] = \int g(x)p(x)\, dx = \int g(x) \underbrace{\frac{p(x)}{q(x)}}_{\text{weight}} \underbrace{q(x)}_{\text{proposal dens.}} \approx \sum_{n=1}^{N} g(x_n)w_n$$

  where x_n drawn from the proposal density $q(x)$ and $w_n = \frac{p(x_n)}{Nq(x_n)}$.

Importance Sampling Cont.

Equivalently $p(x)$ is approximated by

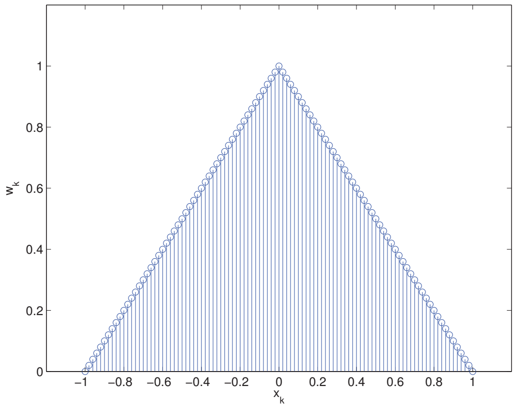
$$p(x) \approx \sum_{n=1}^{N} w_n \delta(x - x_n)$$

$\{x_n, w_n\}$ called **particle distribution**.

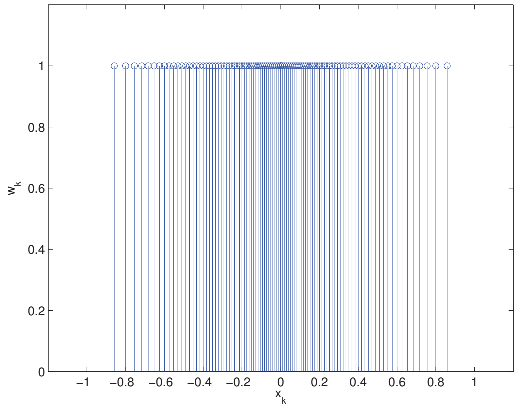
Sometimes $p(x)$ contain an unknown scaling $\Longrightarrow$ weights w_n are scaled. Easy solution is to normalize so $\sum w_n = 1$.

Illustration Importance Sampling

KTH Electrical Engineering



Case 1: Uniformly distributed particles



Case 2: Equally weighted particles

Note: Both represent the same distribution !