Statistical Filtering and Control for AI and Robotics

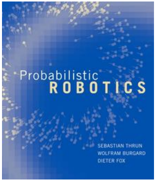
Part III. Extended Kalman filter, Particle filter, etc

Riccardo Muradore

UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

Reference

This lecture is based on the following book



Sebastian Thrun, Wolfram Burgard and Dieter Fox, “Probabilistic Robotics”, MIT Press, 2005

Several pictures from this book have been copied and pasted here

Contents

Introduction

Bayer recursive algorithm

Algorithm 1 $bel(x_t) = \text{BayerFilter}(bel(x_{t-1}), u_t, z_t)$

1: **forall** $x_t \in S_x$ **do**
2: $\quad \overline{bel}(x_t) = \int_{S_x} p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$
3: $\quad bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$
4: **end**
5: **return** $bel(x_t)$

- Line 2: prediction $\overline{bel}(x_t)$ computed using the old $bel(x_{t-1})$ and the current controls u_t
- Line 3: update $bel(x_t)$ computed using $\overline{bel}(x_t)$ and the new measurements z_t

Algorithm 2 $[\hat{x}_{k+1|k+1}, P_{k+1|k+1}]$ =KalmanFilter$(\hat{x}_{k|k}, P_{k|k}, u_k, y_{k+1})$

1: $\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k$
2: $P_{k+1|k} = AP_{k|k}A^T + Q$
3: $\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k})$
4: $P_{k+1|k+1} = P_{k+1|k} - P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}CP_{k+1|k}$
5: **return** $\hat{x}_{k+1|k+1}, P_{k+1|k+1}$

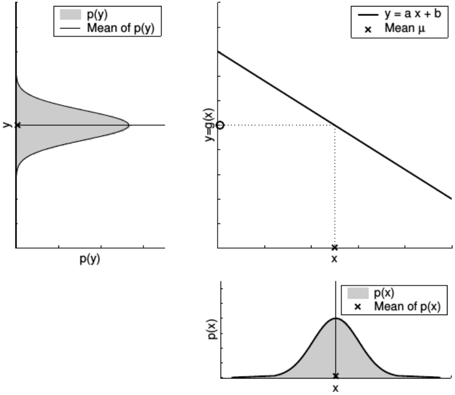
- Line 1-2: prediction $\overline{bel}(x_{k+1})$ computed using the old $bel(x_k)$ and the current controls u_k
- Line 3-4: update $bel(x_{k+1})$ computed using $\overline{bel}(x_{k+1})$ and the new measurements z_{k+1}

With linear Gaussian models

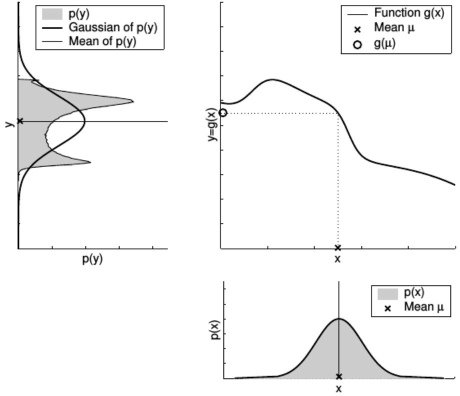
- $\overline{bel}(x_t)$ is fully described by the mean $\hat{x}_{k+1|k}$ and the variance $P_{k+1|k}$
- $bel(x_t)$ is fully described by the mean $\hat{x}_{k|k}$ and the variance $P_{k|k}$

The Kalman filter/predictor is extremely efficient because it is based on two strong assumptions:

1. the state equation and the measurement equation are linear
2. the random variables are Gaussian



What's happen when the assumption
" the state equation and the measurement equation are linear "
is not true?



Extended Kalman filter

Extended Kalman filter

With arbitrary nonlinear functions f and h,

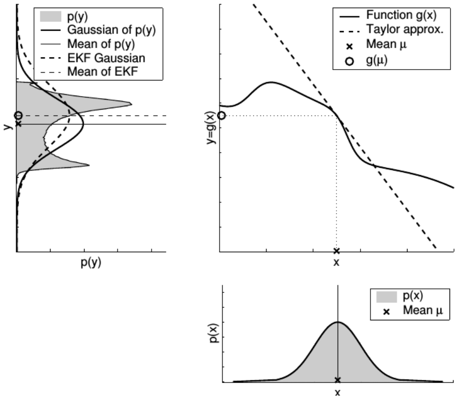
$$\begin{aligned} x_{t+1} &= f(x_t, u_t) + w_t \\ y_t &= h(x_t) + v_t \end{aligned}$$

the belief is no longer Gaussian and there is not a closed-form solution.

The extended Kalman filter (EKF) calculates an approximation to the true belief by a Gaussian. In particular, the belief $bel(x_t)$ at time t is represented by a mean and a covariance

Linearization of $y = g(x)$ via Taylor expansion around μ

$$g(x) = g(\mu) + \left.\frac{\partial g}{\partial x}\right|_{x=\mu} (x - \mu) + \text{higher orders}$$



Algorithm 3 $[\hat{x}_{k+1|k+1}, P_{k+1|k+1}] = \mathrm{EKF}(\hat{x}_{k|k}, P_{k|k}, u_k, y_{k+1})$

1: $\hat{x}_{k+1|k} = f(\hat{x}_{k|k}, u_k)$

2: $A_k = \left.\frac{\partial f(x,u)}{\partial x}\right|_{x=\hat{x}_{k|k}, u=u_k}$

3: $P_{k+1|k} = A_k P_{k|k} A_k^T + Q$

4: $C_k = \left.\frac{\partial h(x)}{\partial x}\right|_{x=\hat{x}_{k+1|k}}$

5: $K_{k+1} = P_{k+1|k} C_k^T \left(C_k P_{k+1|k} C_k^T + R\right)^{-1}$

6: $\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - h(\hat{x}_{k+1|k}))$

7: $P_{k+1|k+1} = P_{k+1|k} - P_{k+1|k} C_k^T (C_k P_{k+1|k} C_k^T + R)^{-1} C_k P_{k+1|k}$

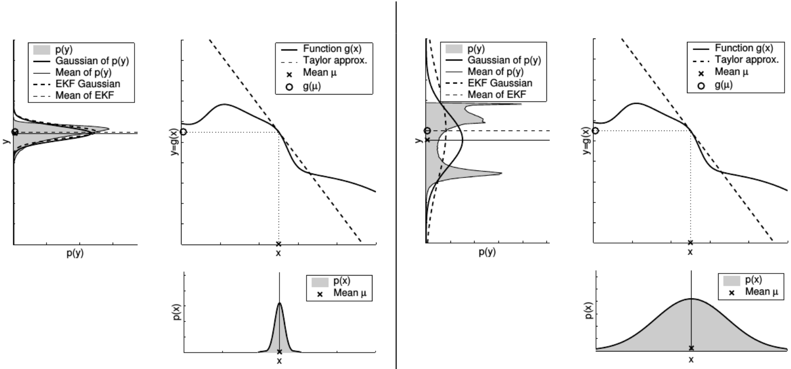
8: **return** $\hat{x}_{k+1|k+1}, P_{k+1|k+1}$

When the EKF works well:

1. f and h are approximately linear
2. the statistical description is not multi-modal

Observation: the less certain is the knowledge (i.e. high variance), the wider the Gaussian belief

Extended Kalman filter



good approximation

bad approximation

Non parametric filtering: the Histogram filter and the Particle filter

Non parametric filtering

What's happen when also the assumption
" the random variables are Gaussian "
is not true?

We have to resort to nonparametric filters: filters that do not rely on a fixed functional form of the posterior (e.g. Gaussian)

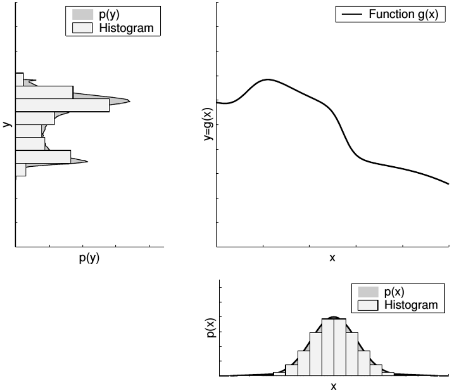
Q. How do they work? A. the probability are approximated by a finite number of values properly selected (each value corresponds to a region in state space)

The quality of the approximation depends on the number of parameters used to represent the probability!

ADVANTAGE nonparametric filters are well-suited to represent complex multimodal beliefs

DRAWBACK increased computational complexity

The Histogram Filter



Histogram filters decompose a continuous state space with support S_t into finitely many regions (K)

$$x_t \in \{\xi_{1,t}, \xi_{2,t}, \ldots, \xi_{K,t}\}$$

where

$$\bigcup_{k=1}^{K} \xi_{k,t} = S_t$$

$$\xi_{i,t} \cap \xi_{j,t} = \emptyset, \forall i \neq j$$

Continuous random variable $\longrightarrow$ Discrete random variable

Through the granularity of the decomposition, we can trade off accuracy and computational efficiency. Fine-grained decompositions infer smaller approximation errors than coarse ones, but at the expense of increased computational complexity.

Each region $\xi_{k,t}$ has a probability assigned $p_{k,t}$, then

$$x_t \in \xi_{k,t} \quad \Rightarrow \quad p(x_t) = \frac{p_{k,t}}{|\xi_{k,t}|}$$

What happens to the PDFs

$$\begin{aligned} p(x_t|u_t, x_{t-1}) \quad &\longrightarrow \quad p(\xi_{k,t}|u_t, \xi_{i,t-1}) = ??? \\ p(z_t|x_t) \quad &\longrightarrow \quad p(z_t|\xi_{k,t}) = ??? \end{aligned}$$

Educated Guess: the densities in each region $\xi_{k,t}$ is approximated by the density of a particular element within that region, e.g. the mean

$$\bar{x}_{k,t} = |\xi_{k,t}|^{-1} \int_{\xi_{k,t}} x_t dt$$

Then we have

$$\begin{aligned} p(x_t|u_t, x_{t-1}) \quad &\longrightarrow \quad p(\xi_{k,t}|u_t, \xi_{i,t-1}) \simeq \frac{\eta}{|\xi_{k,t}|} p(\bar{x}_{k,t}|u_t, \bar{x}_{i,t-1}) \\ p(z_t|x_t) \quad &\longrightarrow \quad p(z_t|\xi_{k,t}) \simeq p(z_t|\bar{x}_{k,t}) \end{aligned}$$

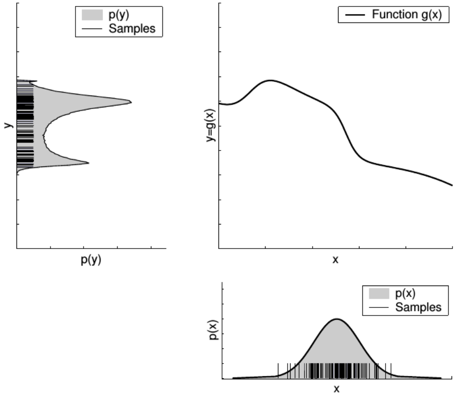
The Histogram Filter

Algorithm 4 $\{p_{k,t}\}_1^K = \text{HF}(\{p_{k,t-1}\}_1^K, u_t, z_t)$

for $k = 1$ **to** K **do**
    $\overline{p}_{k,t} = \sum_{i=1}^{K} p(X_t = \bar{x}_k | X_{t-1} = \bar{x}_i, u_t) p_{i,t-1}$
    $p_{k,t} = \eta p(z_t | X_t = \bar{x}_k) \overline{p}_{k,t}$
end

Also particle filters approximate the posterior by a finite number of parameters but the parameters representing the posterior $bel(x_t)$ are a set of random state samples drawn from the previous posterior

Like histogram filters, particle filters can represent a much broader space of distributions than Gaussian



The samples $x_t^{[i]}, i = 1, \ldots, M$ of a posterior distribution are called particles

$$\mathcal{X}_t = \{x_t^{[1]}, x_t^{[2]}, \ldots, x_t^{[M]}\}$$

The likelihood that a state hypothesis x_t belongs to $\mathcal{X}_t$ should be proportional to the Bayes filter posterior $bel(x_t)$

$$x_t^{[m]} \sim p(x_t | z_{1:t}, u_{1:t})$$

Algorithm 5 $\mathcal{X}_t$ =ParticleFilter($\mathcal{X}_{t-1}, u_t, z_t$)

$\bar{\mathcal{X}}_t = \mathcal{X}_t = \emptyset$
for $m = 1$ **to** M **do**
  sample $x_t^{[m]} \sim p(x_t | x_{t-1}^{[m]}, u_t)$
  $w_t^{[m]} = p(z_t | x_t^{[m]})$
  $\bar{\mathcal{X}}_t = \bar{\mathcal{X}}_t \cup \langle x_t^{[m]}, w_t^{[m]} \rangle$
end
for $m = 1$ **to** M **do**
  draw i with probability $\propto w_t^{[i]}$
  add $x_t^{[i]}$ to $\mathcal{X}_t$
end
return $\mathcal{X}_t$

Particle Filter

Algorithm 6 $\mathcal{X}_t = \text{ParticleFilter}(\mathcal{X}_{t-1}, u_t, z_t)$

$\bar{\mathcal{X}}_t = \mathcal{X}_t = \emptyset$
for $m = 1$ **to** M **do**
  sample $x_t^{[m]} \sim p(x_t | x_{t-1}^{[m]}, u_t)$
  $w_t^{[m]} = p(z_t | x_t^{[m]})$
  $\bar{\mathcal{X}}_t = \bar{\mathcal{X}}_t \cup \langle x_t^{[m]}, w_t^{[m]} \rangle$
end
for $m = 1$ **to** M **do**
  draw i with probability $\propto w_t^{[i]}$
  add $x_t^{[i]}$ to $\mathcal{X}_t$
end
return $\mathcal{X}_t$

Generation of new M samples $x_t^{[m]}$ using the state transition distribution $p(x_t | x_{t-1}, u_t)$ based on particles in $\mathcal{X}_{t-1}$ and the current control u_t

Particle Filter

Algorithm 7 $\mathcal{X}_t = \text{ParticleFilter}(\mathcal{X}_{t-1}, u_t, z_t)$

$\bar{\mathcal{X}}_t = \mathcal{X}_t = \emptyset$
for $m = 1$ **to** M **do**
  sample $x_t^{[m]} \sim p(x_t | x_{t-1}^{[m]}, u_t)$
  $w_t^{[m]} = p(z_t | x_t^{[m]})$
  $\bar{\mathcal{X}}_t = \bar{\mathcal{X}}_t \cup \langle x_t^{[m]}, w_t^{[m]} \rangle$
end
for $m = 1$ **to** M **do**
  draw i with probability $\propto w_t^{[i]}$
  add $x_t^{[i]}$ to $\mathcal{X}_t$
end
return $\mathcal{X}_t$

For each particle $x_t^{[m]}$, compute the importance factor $w_t^{[m]}$ based on the measurement probability $p(z_t|x_t)$ and the new measurement z_t

Particle Filter

Algorithm 8 $\mathcal{X}_t = \text{ParticleFilter}(\mathcal{X}_{t-1}, u_t, z_t)$

$\bar{\mathcal{X}}_t = \mathcal{X}_t = \emptyset$
for $m = 1$ **to** M **do**
  sample $x_t^{[m]} \sim p(x_t | x_{t-1}^{[m]}, u_t)$
  $w_t^{[m]} = p(z_t | x_t^{[m]})$
  $\bar{\mathcal{X}}_t = \bar{\mathcal{X}}_t \cup \langle x_t^{[m]}, w_t^{[m]} \rangle$
end
for $m = 1$ **to** M **do**
  draw i with probability $\propto w_t^{[i]}$
  add $x_t^{[i]}$ to $\mathcal{X}_t$
end
return $\mathcal{X}_t$

Resampling step: replace $\bar{\mathcal{X}}_t$ with another set of the same dimension M, $\mathcal{X}_t$

Particle Filter

Algorithm 9 $\mathcal{X}_t$ =ParticleFilter($\mathcal{X}_{t-1}, u_t, z_t$)

$\bar{\mathcal{X}}_t = \mathcal{X}_t = \emptyset$
for $m = 1$ **to** M **do**
  sample $x_t^{[m]} \sim p(x_t | x_{t-1}^{[m]}, u_t)$
  $w_t^{[m]} = p(z_t | x_t^{[m]})$
  $\bar{\mathcal{X}}_t = \bar{\mathcal{X}}_t \cup \langle x_t^{[m]}, w_t^{[m]} \rangle$
end
for $m = 1$ **to** M **do**
  draw i with probability $\propto w_t^{[i]}$
  add $x_t^{[i]}$ to $\mathcal{X}_t$
end
return $\mathcal{X}_t$

$\bar{\mathcal{X}}_t$ represents $\overline{bel}(x_t)$

Particle Filter

Algorithm 10 $\mathcal{X}_t = \text{ParticleFilter}(\mathcal{X}_{t-1}, u_t, z_t)$

$\bar{\mathcal{X}}_t = \mathcal{X}_t = \emptyset$
for $m = 1$ **to** M **do**
    sample $x_t^{[m]} \sim p(x_t | x_{t-1}^{[m]}, u_t)$
    $w_t^{[m]} = p(z_t | x_t^{[m]})$
    $\bar{\mathcal{X}}_t = \bar{\mathcal{X}}_t \cup \langle x_t^{[m]}, w_t^{[m]} \rangle$
end
for $m = 1$ **to** M **do**
    draw i with probability $\propto w_t^{[i]}$
    add $x_t^{[i]}$ to $\mathcal{X}_t$
end
return $\mathcal{X}_t$

$\mathcal{X}_t$ represents $bel(x_t) = \eta p(z_t | x_t^{[m]}) \overline{bel}(x_t)$