Recursive Bayesian State Estimation

Consider the problem of tracking the internal state vector of a dynamic system using (i) knowledge of the system equations, and (ii) measured outputs in time.

Let $\quad x_k = f_k(x_{k-1}, v_{k-1}) \quad$ State dynamics eqn.

$k \in \mathbb{N}$ {i.e. $k = 0, 1, 2, \ldots$} $\quad z_k = h_k(x_k, n_k) \quad$ Measurement equation

where $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \longrightarrow \mathbb{R}^{n_x}$ is possibly a nonlinear function of the state, $v_k \in \mathbb{R}^{n_v}$ is an iid noise process sequence, and $h_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_n} \longrightarrow \mathbb{R}^{n_z}$ is possibly a nonlinear function of the state and n_k is an iid noise process.

We seek filtered estimates of x_k based on $z_{1:k}$. From a Bayesian perspective, we need to obtain $p(x_k | z_{1:k})$. We assume that $p(x_0 | z_0) \triangleq p(x_0)$, the initial pdf of the state vector is also known as the prior.

Note that we have the Chapman-Kolmogorov equation:

$$p(x_k | z_{1:k-1}) = \int p(x_k | x_{k-1}) \, p(x_{k-1} | z_{1:k-1}) \, dx_{k-1}$$

and Bayes rule: $\quad p(x_k | z_{1:k}) = \dfrac{p(z_k | x_k) \, p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})}$

where the normalizing constant is

$$p(z_k | z_{1:k-1}) = \int p(z_k | x_k) p(x_k | z_{1:k-1}) dx_k$$

From the first two equations we obtain

$$p(x_k | z_{1:k}) \propto p(z_k | x_k) \int p(x_k | x_{k-1}) p(x_{k-1} | z_{1:k-1}) dx_{k-1}$$

where we replace the normalization constant with "proportionality".

<u>Kalman Filter</u>: The KF assumes that the posterior density at every time step is Gaussian, hence it updates the parameters mean and covariance.

If $p(x_{k-1} | z_{1:k-1})$ is Gaussian, one can prove that $p(x_k | z_{1:k})$ is also Gaussian provided that:

(i) v_{k-1} and n_k are drawn from Gaussians (with known params)

(ii) $f_k(x_{k-1}, v_{k-1})$ is a linear (known) function of x_{k-1} & v_{k-1}.

(iii) $h_k(x_k, n_k)$ is a linear (known) function of x_k & n_k.

$$\left.\begin{aligned} x_k &= F_k x_{k-1} + v_{k-1} \\ z_k &= H_k x_k + n_k \end{aligned}\right\}$$

F_k, H_k are known matrices
$Cov(v_{k-1}) = Q_{k-1}$ & $Cov(n_k) = R_k$ are known
Assume $E[v_{k-1}] = 0$, $E[n_k] = 0$ and they are iid.

Then, in KF we have: (for $N(x; m, P) \sim$ Gaussians)

$$p(x_{k-1} | z_{1:k-1}) = N(x_{k-1}; m_{k-1|k-1}, P_{k-1|k-1})$$

$$p(x_k | z_{1:k-1}) = N(x_k; m_{k|k-1}, P_{k|k-1}) \quad \text{(Prediction)}$$

$$p(x_k | z_{1:k}) = N(x_k; m_{k|k}, P_{k|k}) \quad \text{(Time-update)}$$

where $m_{k|k-1} = F_k m_{k-1|k-1}$

$$P_{k|k-1} = Q_{k-1} + F_k P_{k-1|k-1} F_k^T$$

$$m_{k|k} = m_{k|k-1} + K_k (z_k - H_k m_{k|k-1})$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1}$$

and $S_k = H_k P_{k|k-1} H_k^T + R_k \quad (\triangleq \operatorname{cov}(z_k - H_k m_{k|k-1}))$

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad \text{(the Kalman gain)}$$

<u>Grid-Based Methods</u>: These provide the optimal recursion of $p(x_k | z_{1:k})$ if the state space is <u>discrete</u> and consists of a finite number of states. Suppose that the state space at time $(k-1)$ consists of discrete states x^i_{k-1}, $i = 1, \dots, N_s$. Let $w^i_{k-1|k-1} \triangleq \Pr(x_{k-1} = x^i_{k-1} | z_{1:k-1})$. Then

$$\mathbf{p}(x_{k-1} | z_{1:k-1}) = \sum_{i=1}^{N_s} w^i_{k-1|k-1} \, \delta(x_{k-1} - x^i_{k-1})$$

From this, we obtain

$$p(x_k | z_{1:k-1}) = \sum_{i=1}^{N_s} w^i_{k|k-1} \delta(x_k - x^i_k) \quad \text{where } w^i_{k|k-1} \triangleq \sum_{j=1}^{N_s} w^j_{k-1|k-1} \, p(x^i_k | x^j_{k-1})$$

$$p(x_k | z_{1:k}) = \sum_{i=1}^{N_s} w^i_{k|k} \, \delta(x_k - x^i_k) \text{ where } w^i_{k|k} \triangleq \frac{w^i_{k|k-1} \, p(z_k | x^i_k)}{\sum_{j=1}^{N_s} w^j_{k|k-1} \, p(z_k | x^j_k)}$$

In this method, we assume that $p(x^i_k | x^j_{k-1})$ and $p(z_k | x^i_k)$ are known. Note that these are (Markov) state dynamics and (latent-variable) measurement equations.

Extended Kalman Filter: The EKF is based on the linearization of state dynamics and measurement equations, as well as the approximation of relevant state-conditional distributions with Gaussian densities:

$$p(x_{k-1} \mid z_{1:k-1}) \approx N(x_{k-1}; m_{k-1|k-1}, P_{k-1|k-1})$$

$$p(x_k \mid z_{1:k-1}) \approx N(x_k; m_{k|k-1}, P_{k|k-1})$$

$$p(x_k \mid z_{1:k}) \approx N(x_k; m_{k|k}, P_{k|k})$$

where

$$m_{k|k-1} = f_k(m_{k-1|k-1}) \quad \text{(propagate mean)}$$

$$P_{k|k-1} = Q_{k-1} + \hat{F}_k P_{k-1|k-1} \hat{F}_k^T \quad \text{(linearize around mean)}$$

$$m_{k|k} = m_{k|k-1} + K_k (z_k - h_k(m_{k|k-1})) \quad \text{(correct via mean)}$$

$$P_{k|k} = P_{k|k-1} - K_k \hat{H}_k P_{k|k-1} \quad \text{(linearize around mean)}$$

with the following approximations:

$$\hat{F}_k = \left.\frac{\partial f_k(x)}{\partial x}\right|_{x = m_{k-1|k-1}}, \quad \hat{H}_k = \left.\frac{\partial h_k(x)}{\partial x}\right|_{x = m_{k|k-1}}$$

$$S_k = \hat{H}_k P_{k|k-1} \hat{H}_k^T + R_k, \quad K_k = P_{k|k-1} \hat{H}_k^T S_k^{-1}$$

Thus EKF truncates the Taylor series at order-1; higher order EKF approximations are possible, but costly.

Note that mixtures of Gaussians combined with the EKF framework are possible, when state distribution is multimodal.

Approximate Grid-Based Methods: If the state-space is continuous but can be partitioned into N_s cells (i.e. histogram-bin like approach), then a grid-based method can be used to approximate the posterior density. Here, if $\bar{x}^i_k$ denotes the ~~k^{th} bin~~ i^{th} bin at time k, then

$$w^i_{k|k-1} \triangleq \sum_{j=1}^{N_s} w^j_{k-1|k-1} \int_{x \in x^i_k} p(x|\bar{x}^j_{k-1})\,dx$$

$$w^i_{k|k} \triangleq \frac{w^i_{k|k-1} \int_{x \in x^i_k} p(z_k|x)\,dx}{\sum_{j=1}^{N_s} w^j_{k|k-1} \int_{x \in x^j_k} p(z_k|x)\,dx}$$

If we let $\bar{x}^i_k$ be the center of a bin (cell), then the integrals can be approximated by conditionals evaluated at these center values; i.e. $p(\bar{x}^i_k|\bar{x}^j_{k-1})$ and $p(z_k|\bar{x}^i_k)$.

Particle Filtering Methods

Sequential Importance Sampling (SIS) Algorithm:

Consider the posterior $p(x_{0:k}|z_{1:k})$ and let $x^i_{0:k}$ $i=1,\dots,N_s$ be a set of support points with weights w^i_k, $i=1,\dots,N_s$ such that $\sum_i w^i_k = 1$. Then

$$p(x_{0:k}|z_{1:k}) \approx \sum_{i=1}^{N_s} w^i_k\,\delta(x_{0:k} - x^i_{0:k})$$

Let $x^i \sim q(x)$, $i = 1, \dots, N_s$ be samples generated from a proposal distribution $q(\cdot)$ (called importance density) and suppose $p(x) \propto \pi(x)$ is a desired prob. distribution for x^i. Then

$$p(x) \approx \sum_{i=1}^{N_s} w^i \delta(x - x^i) \quad \text{where } w^i \propto \frac{\pi(x^i)}{q(x^i)}$$

Therefore, if $x^i_{0:k}$ are drawn from an importance density $q(x_{0:k}|z_{1:k})$, then $w^i_k \propto \dfrac{p(x^i_{0:k}|z_{1:k})}{q(x^i_{0:k}|z_{1:k})}$

If the importance density is chosen such that

$$q(x_{0:k}|z_{1:k}) = q(x_k|x_{0:k-1}, z_{1:k})\, q(x_{0:k-1}|z_{1:k-1})$$

then one can obtain $x^i_{0:k} \sim q(x_{0:k}|z_{1:k})$ from $x^i_{0:k-1} \sim q(x_{0:k-1}|z_{1:k-1})$ with the addition of $x^i_k \sim q(x_k|x^i_{0:k-1}, z_{1:k})$.

Note that $p(x_{0:k}|z_{1:k}) = \dfrac{p(z_k|x_{0:k}, z_{1:k-1})\, p(x_{0:k}|z_{1:k-1})}{p(z_k|z_{1:k-1})}$

$\{(45)$ in paper$\}$ $\quad \propto p(z_k|x_k)\, p(x_k|x_{k-1})\, p(x_{0:k-1}|z_{1:k-1})$

Then the weight update is

$$w^i_k = w^i_{k-1} \frac{p(z_k|x^i_k)\, p(x^i_k|x^i_{k-1})}{q(x^i_k|x^i_{0:k-1}, z_{1:k})}$$

If $q(x_k | x_{0:k-1}, z_{1:k}) = q(x_k | x_{k-1}, z_k)$, then

$$w_k^i \propto w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k)}$$

and $p(x_k | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i)$

Algorithm 1: SIS Particle Filter

$$\{x_k^i, w_k^i\}_{i=1}^{N_s} \leftarrow SIS\left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k\right]$$

for $i = 1 : N_s$

Draw $x_k^i \sim q(x_k | x_{k-1}^i, z_k)$

Assign $w_k^i \propto w_{k-1}^i \dfrac{p(z | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k)}$

end

---

1) Degeneracy Problem: After a few iterations, almost all but one particle will have negligible weight.

One could monitor $N_{eff} \approx \left[\sum_{i=1}^{N_s} (w_k^i)^2\right]^{-1} \leq N_s$ and if N_{eff} becomes "too small", use <u>resampling</u> (or use a very large N_s).

2) Good Choice of Importance Density to reduce the effect of degeneracy. The optimal importance density (which maximizes N_{eff}) has been shown to be

$$q(x_k | x_{k-1}^i, z_k)_{opt} = p(x_k | x_{k-1}^i, z_k) = \frac{p(z_k | x_k, x_{k-1}^i) p(x_k | x_{k-1}^i)}{p(z_k | x_{k-1}^i)}$$

true density

Then $w_k^i = w_{k-1}^i \, p(z_k | x_{k-1}^i)$
$= w_{k-1}^i \int p(z_k | x_k') p(x_k' | x_{k-1}^i) dx_k'$

This choice is optimal since for x_{k-1}^i, $w_k^i = 1/N_s$.
(See paper for special cases where analytical integration is possible.)

3) Resampling: When N_{eff} < threshold, resampling could be employed. Resampling ~~by replacement~~ creates a new set $\{x_k^{i*}\}_{i=1}^{N_s}$ from $\{x_k^i\}_{i=1}^{N_s}$ so that $\cancel{p(x_k|z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i)}$ ~~and~~ $\Pr(x_k^{i*} = x_k^j) = w_k^j$. This is called resampling by replacement. The weights $\{w_k^{i*}\}_{i=1}^{N_s}$ are reset to $1/N_s$.

Algorithm 2: Resampling

$$\{x_k^{j*}, w_k^j, i^j\}_{j=1}^{N_s} = RESAMPLE\left[\{x_k^i, w_k^i\}_{i=1}^{N_s}\right]$$

```
Initialize the CDF:  c_1 = 0
for i = 2:Ns
    c_i = c_{i-1} + w_k^i
end
i=1; Draw u_1 ~ Uniform[0, 1/Ns];
for j = 1:Ns
    u_j = u_1 + (1/Ns)(j-1)
    while u_j > c_i , i=i+1, end
    x_k^{j*} <- x_k^i ;  w_k^j <- 1/Ns ;  i^j <- i      "parent of new sample j is i"
end
```

Algorithm 3: Generic Particle filter

$$\{x_k^i, w_k^i\}_{i=1}^{N_s} = PF\left[\{x_{k-1}^i, w_{k-1}^i\}, z_k\right]$$

for $i = 1 : N_s$

$$x_k^i \sim q(x_k | x_{k-1}^i, z_k)$$

$$w_k^i \propto w_{k-1}^i \frac{p(z_k|x_k^i)\, p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, z_k)}$$

end

for $i = 1:N_s$, $w_k^i \leftarrow w_k^i / \sum_j w_k^j$, end

Calculate $N_{eff} \approx \left(\sum_{i=1}^{N_s} {w_k^i}^2\right)^{-1}$

If $N_{eff} < thr$, $\{x_k^i, w_k^i\}_{i=1}^{N_s} \leftarrow RESAMPLE\left[\{x_k^i, w_k^i\}_{i=1}^{N_s}\right]$, end.

Sampling Importance Resampling Filter (SIR): (possibly inefficient and ~~~~ sensitive to outliers)

for $i = 1 : N_s$, $x_k^i \sim p(x_k|x_{k-1}^i)$; $w_k^i = p(z_k|x_k^i)$, end

for $i = 1:N_s$, $w_k^i \leftarrow w_k^i / \sum_j w_k^j$, end

RESAMPLE $\{x_k^i, w_k^i\}_{i=1}^{N_s}$

Auxiliary Particle filter

for $i = 1 : N_s$

$$\mu_k^i = E[x_k|x_{k-1}^i] \quad \text{OR} \quad \mu_k^i \sim p(x_k|x_{k-1}^i)$$

$$w_k^i = w_{k-1}^i\, p(z_k|\mu_k^i)$$

end

for $i = 1:N_s$, $w_k^i \leftarrow w_k^i / \sum_j w_k^j$, end

RESAMPLE $\{x_k^i, w_k^i\}_{i=1}^{N_s}$

parent particle

for $j=1:N_s$, $x_k^j \sim q(x_k | i^j, z_{1:k})$; $w_k^j \propto \dfrac{p(z_k | x_k^j)}{p(z_k | \mu_k^{i^j})}$, end

for $i=1:N_s$, $w_k^i = w_k^i / \sum_j w_k^j$, end

KDE: $K_h(x) = \frac{1}{h^{n_x}} K\left(\frac{x}{h}\right)$

<u>Regularized Particle Filter</u> (uses $p(x_k | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i K_h(x_k - x_k^i)$)

for $i=1:N_s$, $x_k^i \sim q(x_k | x_{k-1}^i, z_k)$; $w_k^i \propto w_{k-1}^i \dfrac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k)}$, end

for $i=1:N_s$, $w_k^i \leftarrow w_k^i / \sum_j w_k^j$; end

Calculate $N_{eff} = \left(\sum_i w_k^{i\,2} \right)^{-1}$

If $N_{eff} <$ thr

Compute the empirical covariance S_k from $\{x_k^i, w_k^i\}_{i=1}^{N_s}$.

Find D_k s.t. $D_k D_k^T = S_k$

Resample $\{x_k^i, w_k^i\}_{i=1}^{N_s}$

for $i=1:N_s$, $\varepsilon^i \sim K$ from the Epanechnikov kernel

$x_k^i = x_k^i + h_{opt} D_k \varepsilon^i$, ~~end~~

end

end

where the Epanechnikov kernel is

$$K_{opt} = \begin{cases} \frac{n_x+2}{2c_{n_x}} (1 - \|x\|^2), & \text{if } \|x\| < 1 \\ 0, & \text{o.w.} \end{cases}$$

and $h_{opt} = A N_s^{1/(n_x+4)}$ with $A = \left[8 c_{n_x}^{-1} (n_x+4)(2\sqrt{\pi})^{n_x} \right]^{1/(n_x+4)}$

according to the minimum integrated square error criterion.

Unscented Kalman Filters

Basic Idea: The Unscented Transformation

Let $\sum_{i=0}^{p} w^i = 1$, $w^i \geq 0$. Let x^i be vectors in $\mathbb{R}^n$.

Let $z = h(x)$ for some nonlinear mapping $h: \mathbb{R}^n \to \mathbb{R}^m$

Then if $\{x^i, w^i\}$ are chosen carefully, then

$$E[z] \approx \sum_i w^i z^i \triangleq \bar{z} \quad \text{where} \quad z^i = h(x^i)$$

and $$\mathrm{Cov}(z) \approx \sum_i w^i (z^i - \bar{z})(z^i - \bar{z})^T$$

One set of points: $$x^i = \bar{x} + (N_x \Sigma_x)_i^{1/2}, \quad w^i = 1/(2N_x)$$
$$x^{i+N_x} = \bar{x} - (N_x \Sigma_x)_i^{1/2}, \quad w^{i+N_x} = 1/(2N_x)$$

Another set of points: $$x^{(0)} = \bar{x}, \quad w^0 = w^0$$
$$x^i = \bar{x} + \left(\frac{N_x}{1-w^0}\Sigma_x\right)_i^{1/2}, \quad w^i = \frac{1-w^0}{2N_x}$$
$$x^{i+N_x} = \bar{x} - \left(\frac{N_x}{1-w^0}\Sigma_x\right)_i^{1/2}, \quad w^i = \frac{1-w^0}{2N_x}$$

where $(m)_i$ is the i^{th} row or column of the matrix

These points are usually called sigma-points.

The UKF algorithm is as follows:

1) Create sigma-points appropriately.

2) $\hat{x}^{(i)}_{a,n} = f(x^{(i)}_{a,n}, u_n)$ (system dynamics) propagation of σ-points.

3) $\hat{\mu}_{a,n} = \sum_{i=0}^{p} w^{(i)} \hat{x}^{(i)}_{a,n}$ (predicted mean)

4) $\hat{K}_{a,n} = \sum_i w^{(i)} (\hat{x}^{(i)}_{a,n} - \hat{\mu}_{a,n})(\hat{x}^{(i)}_{a,n} - \hat{\mu}_{a,n})^T$ (pred. cov.)

5) $\hat{y}^{(i)}_n = g(\hat{x}^{(i)}_{a,n}, u_n)$ (meas. eqn., pred. outputs)

6) $\hat{y}_n = \sum_i w^{(i)} \hat{y}^{(i)}_n$ (pred. expected output)

7) $\hat{S}_n = \sum_i w^{(i)} (\hat{y}^{(i)}_n - \hat{y}_n)(\hat{y}^{(i)}_n - \hat{y}_n)^T$ (innovation cov.)

8) $\hat{K}^{xy}_n = \sum_i w^{(i)} (\hat{x}^{(i)}_n - \hat{\mu}_n)(\hat{x}^{(i)}_n - \hat{\mu}_n)^T$ (cross-cov).

9) UKF updates
$$\mu_n = \hat{\mu}_n + W_n v_n$$
$$K_n = \hat{K}_n - W_n \hat{S}_n W_n^T$$
$$v_n = y_n - \hat{y}_n \quad \text{(meas. error)}$$
$$W_n = \hat{K}^{xy}_n \hat{S}_n^{-1}$$

~~Next,~~