Overview of Optimal Filtering Algorithms

Simo Särkkä <simo.sarkka@hut.fi>

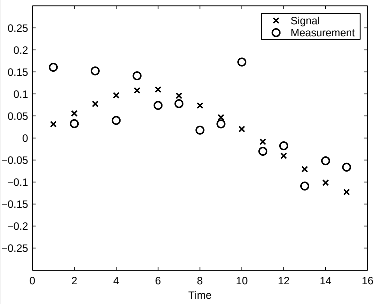
Laboratory of Computational Engineering
Helsinki University of Technology

October 5, 2006

Contents

1. Mathematical Description of the Problem
   - Discrete-Time Optimal Filtering
   - Continuous-Discrete-Time Optimal Filtering
   - Continuous-Time Optimal Filtering
2. Optimal Filtering in Practice
   - Linear Gaussian Optimal Filtering
   - Non-Linear Optimal Filtering
3. Approximation Methods
   - Gaussian Approximations
   - Particle Filtering
4. Summary

Discrete-Time Filtering Model



- Difference equation model for the state dynamics:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_{k-1}),$$

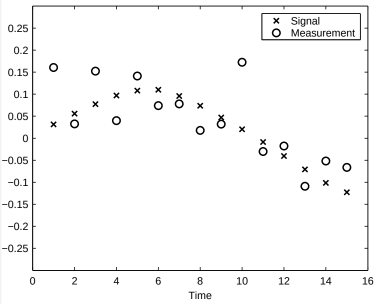
  where $\mathbf{q}_{k-1}$ is the process noise.

- Measurements $\mathbf{y}_k$ are modeled as

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k)$$

  where $\mathbf{r}_k$ is the measurement noise.

Discrete-Time Filtering Model [2/2]



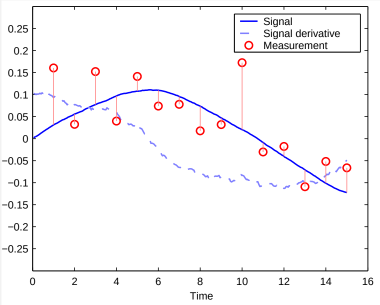
- More generally, discrete-time Markov model for the state:

$$\mathbf{x}_k \sim p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}).$$

- Likelihood distribution of the measurement:

$$\mathbf{y}_k \sim p(\mathbf{y}_k \,|\, \mathbf{x}_k).$$

Continuous-Discrete-Time Filtering Model



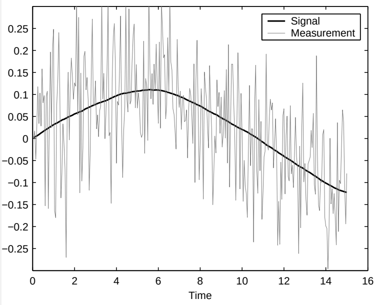
- The dynamics of the state $\mathbf{x}(t)$ modeled as stochastic differential equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(t)\,\mathbf{w}(t).$$

- Measurements $\mathbf{y}_k$ obtained at discrete instances of time:

$$\mathbf{y}_k \sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)).$$

Continuous-time Filtering Model



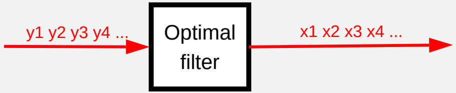
- The dynamics of the state $\mathbf{x}(t)$ modeled as stochastic differential equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(t)\,\mathbf{w}(t).$$

- Measurement signal $\mathbf{y}(t)$ is also modeled as stochastic differential equation

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t) + \mathbf{V}(t)\,\mathbf{n}(t).$$

Optimal Filter



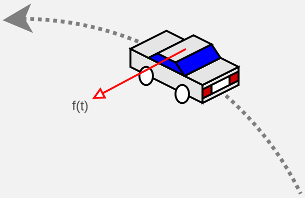
- The purpose of the optimal filter is to compute the posterior distribution of the state given the measurements:

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_1, \ldots, \mathbf{y}_k).$$

- The "filtered" state $\hat{\mathbf{x}}(t_k)$ typically is the posterior mean

$$\hat{\mathbf{x}}(t_k) = E(\mathbf{x}(t_k) \mid \mathbf{y}_1, \ldots, \mathbf{y}_k).$$

Dynamic Model for a Car [1/3]



- The dynamics of the car in 2d (x_1, x_2) are given by the Newton's law:

$$\mathbf{f}(t) = m\,\mathbf{a}(t),$$

where $\mathbf{a}(t)$ is the acceleration, m is the mass of the car, and $\mathbf{f}(t)$ is a vector of (unknown) forces acting the car.

- We shall now model $\mathbf{f}(t)/m$ as a 2-dimensional white noise process:

$$\begin{aligned} \mathrm{d}^2 x_1/\mathrm{d}t^2 &= w_1(t) \\ \mathrm{d}^2 x_2/\mathrm{d}t^2 &= w_2(t). \end{aligned}$$

Dynamic Model for a Car [2/3]

- If we define $x_3(t) = \mathrm{d}x_1/\mathrm{d}t$, $x_4(t) = \mathrm{d}x_2/\mathrm{d}t$, then the model can be written as a first order system of differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}_{\mathbf{L}} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

- In shorter matrix form:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}\,\mathbf{x} + \mathbf{L}\,\mathbf{w}.$$

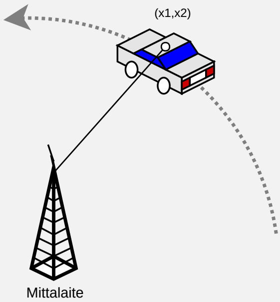
Dynamic Model for a Car [3/3]

- If the state of the car is measured (sampled) with sampling period Δt it suffices to consider the state of the car only at the time instances $t \in \{0, \Delta t, 2\Delta t, \ldots\}$.
- The dynamic model can be discretized, which leads to the linear difference equation model

$$\mathbf{x}_k = \mathbf{A}\,\mathbf{x}_{k-1} + \mathbf{q}_{k-1},$$

where $\mathbf{x}_k = \mathbf{x}(t_k)$, $\mathbf{A}$ is the transition matrix and $\mathbf{q}_k$ is a discrete-time Gaussian noise process.

Measurement Model for a Car



- Assume that the position of the car (x_1, x_2) is measured and the measurements are corrupted by Gaussian measurement noise $e_{1,k}, e_{2,k}$:

$$
\begin{aligned}
y_{1,k} &= x_{1,k} + e_{1,k} \\
y_{2,k} &= x_{2,k} + e_{2,k}.
\end{aligned}
$$

- The measurement model can be now written as

$$
\mathbf{y}_k = \mathbf{H}\,\mathbf{x}_k + \mathbf{e}_k, \qquad \mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}
$$

Kalman Filter

- The dynamic and measurement models of the car now form a linear Gaussian filtering model:

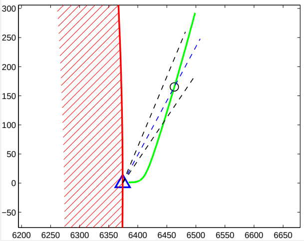
$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}\,\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}\,\mathbf{x}_k + \mathbf{r}_k, \end{aligned}$$

  where $\mathbf{q}_{k-1} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$.
- The posterior distribution is Gaussian

$$p(\mathbf{x}_k \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k) = \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k, \mathbf{P}_k).$$

- The recursion equations, for computing the mean $\mathbf{m}_k$ and covariance $\mathbf{P}_k$ of the posterior distribution are called the Kalman filter.

Re-Entry Vehicle Model [1/3]



- Gravitation law:

$$\mathbf{F} = m\,\mathbf{a}(t) = -\frac{G\,m\,M\,\mathbf{r}(t)}{|\mathbf{r}(t)|^3}.$$

- If we also model the friction and uncertainties:

$$\mathbf{a}(t) = -\frac{G\,M\,\mathbf{r}(t)}{|\mathbf{r}(t)|^3} - D(\mathbf{r}(t))\,|\mathbf{v}(t)|\,\mathbf{v}(t) + \mathbf{w}(t).$$

Re-Entry Vehicle Model [2/3]

- If we define $\mathbf{x} = (x_1 \; x_2 \; \frac{\mathrm{d}x_1}{\mathrm{d}t} \; \frac{\mathrm{d}x_2}{\mathrm{d}t} \;)^T$, the model is of the form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}) + \mathbf{L}\,\mathbf{w}(t).$$

  where $\mathbf{f}(\cdot)$ is non-linear.
- The radar measurement:

$$r = \sqrt{(x_1 - x_r)^2 + (x_2 - y_r)^2} + e_r$$
$$\theta = \tan^{-1}\left(\frac{x_2 - y_r}{x_1 - x_r}\right) + e_\theta,$$

  where $e_r \sim \mathrm{N}(0, \sigma_r^2)$ and $e_\theta \sim \mathrm{N}(0, \sigma_\theta^2)$.

Re-Entry Vehicle Model [3/3]

- By suitable discretization the model can be approximately written as discrete-time state space model:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_{k-1})$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k),$$

  where $\mathbf{y}_k$ is the vector of measurements, and $\mathbf{q}_{k-1} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$.
- The model is a special case of probabilistic state space models of the form

$$\mathbf{y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k)$$
$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}).$$

Discrete-Time Non-Linear Optimal Filtering [1/2]

Discrete-time Optimal Filter

- Prediction step uses *Chapman-Kolmogorov equation*

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1}$$

- Update step uses the *Bayes' rule*

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{Z_k}p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$$

Discrete-Time Non-Linear Optimal Filtering [2/2]

- Approximate Kalman filtering: Kalman filter, extended Kalman filter, unscented Kalman filter
- Particle Filtering: Sequential Monte Carlo, Sequential importance resampling, Bootstrap filter, Condensation
- Mixture Gaussian Approximations: Multiple Model Filtering, Rao-Blackwellized particle filtering
- Exponential family, assumed density filtering
- Finite grid approximations of posterior distributions

Continuous-Discrete-Time Filtering [1/3]

- If don't discretize the dynamics in car's or re-entry vehicle's models, we get a continuous-discrete filtering model.
- The dynamics of state $\mathbf{x}(t)$ modeled as a stochastic differential equation (diffusion process)

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(t)\,\mathbf{w}(t).$$

- Measurements $\mathbf{y}_k$ are obtained at discrete times

$$\mathbf{y}_k \sim p(\mathbf{y}_k \,|\, \mathbf{x}(t_k)).$$

- Formal solution: Compute the posterior distribution(s)

$$p(\mathbf{x}(t) \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k), \qquad t \geq t_k.$$

Continuous-Discrete-Time Filtering [2/3]

Optimal Continuous-Discrete Filter

1. Prediction step: Solve the Kolmogorov-forward (Fokker-Planck) partial differential equation.

$$\frac{\partial p}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\left(f_i(\mathbf{x}, t)\, p\right) + \frac{1}{2}\sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j}\left([\mathbf{L}\,\mathbf{Q}\,\mathbf{L}^T]_{ij}\, p\right)$$

2. Update step: Apply the Bayes' rule.

$$p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k \mid \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k \mid \mathbf{x}(t_k))\, p(\mathbf{x}(t_k) \mid \mathbf{y}_{1:k-1})\, \mathrm{d}\mathbf{x}(t_k)}$$

Continuous-Discrete-Time Filtering [3/3]

- Gaussian models and approximations: extended Kalman (-Bucy) filters and unscented Kalman (-Bucy) filters.
- FEM and finite difference based approximations to the Kolmogorov forward PDE.
- Bootstrap filter
- Sequential importance resampling

Continuous-Time Filtering

- The dynamics of state $\mathbf{x}(t)$ modeled as a stochastic differential equation (diffusion process)

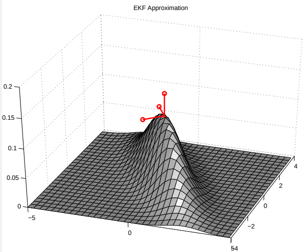
$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(t)\,\mathbf{w}(t).$$

- Measurement signal $\mathbf{y}(t)$ is a diffusion process, which is driven by $\mathbf{x}(t)$ and a measurement noise process:

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t) + \mathbf{V}(t)\,\mathbf{n}(t).$$

- Formal solutions given by Kushner-Stratonovich equation and Zakai equation, which are stochastic PDEs.
- Approximations: Extended/unscented Kalman-Bucy filters, particle filters.

Extended Kalman Filter (EKF) [1/2]



- Forms Gaussian approximation

$$p(\mathbf{x}_k \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_k) \approx \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k, \mathbf{P}_k).$$

- Linearize the dynamic and measurement models around the current mean estimate **m**:

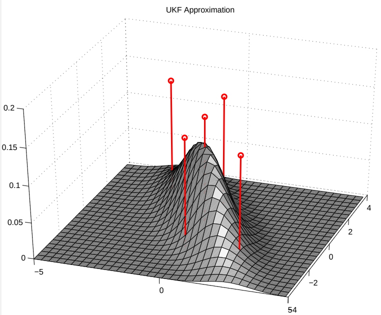
$$\mathbf{f}(\mathbf{x}, t) \approx \mathbf{f}(\mathbf{m}, t) + \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{m}, t)\right)(\mathbf{x} - \mathbf{m})$$

$$\mathbf{h}(\mathbf{x}, t) \approx \mathbf{h}(\mathbf{m}, t) + + \left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\mathbf{m}, t)\right)(\mathbf{x} - \mathbf{m}).$$

Extended Kalman Filter (EKF) [2/2]

- Recursion equations almost the same as in Kalman filter.
- Also higher order Taylor series approximations possible.
- Works well if the functions are approximately locally linear.
- Strong non-linearities are problematic.
- Derivatives have to exists and be known.

Unscented Kalman Filter (UKF) [1/2]



- Also forms a Gaussian approximation

$$p(\mathbf{x}_k \mid \mathbf{y}_1, \ldots, \mathbf{y}_k) \approx \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k).$$

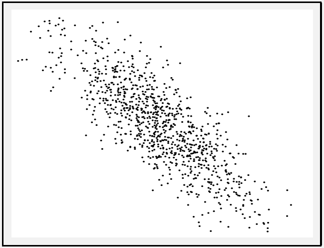
- The idea is to compute the Cholesky factorization $\mathbf{P} = \mathbf{L}\,\mathbf{L}^T$ and form sigma points:

$$\mathbf{X} = \begin{bmatrix} \mathbf{m} & (\mathbf{m} + \mathbf{L}_1) & \cdots & (\mathbf{m} + \mathbf{L}_n) & (\mathbf{m} - \mathbf{L}_1) & \cdots & (\mathbf{m} - \mathbf{L}_n) \end{bmatrix}.$$

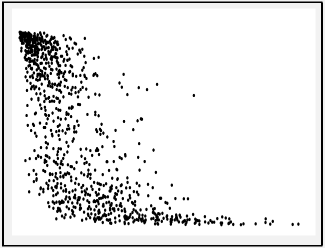
Unscented Kalman Filter (UKF) [2/2]

- Approximations are obtained by propagating the sigma points through the non-linearities.
- The equations of the filter are similar to the equations of the Kalman filter.
- Functions don't need to be differentiable or derivatives known.
- Works with stronger non-linearities.
- Still a local approximation, that is, Gaussian approximation around the mean.

EKF/UKF Example



$\Longrightarrow$

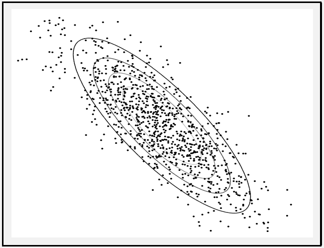


$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix}\right)$$

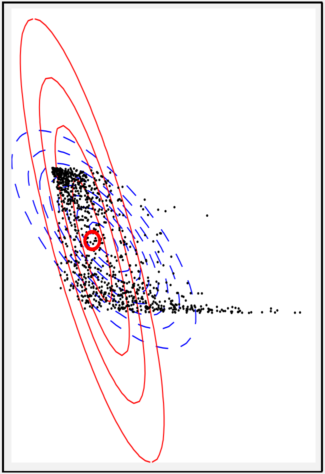
$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = \exp(-y_1), \quad y_1(0) = x_1$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = -\frac{1}{2}y_2^3, \qquad y_2(0) = x_2$$

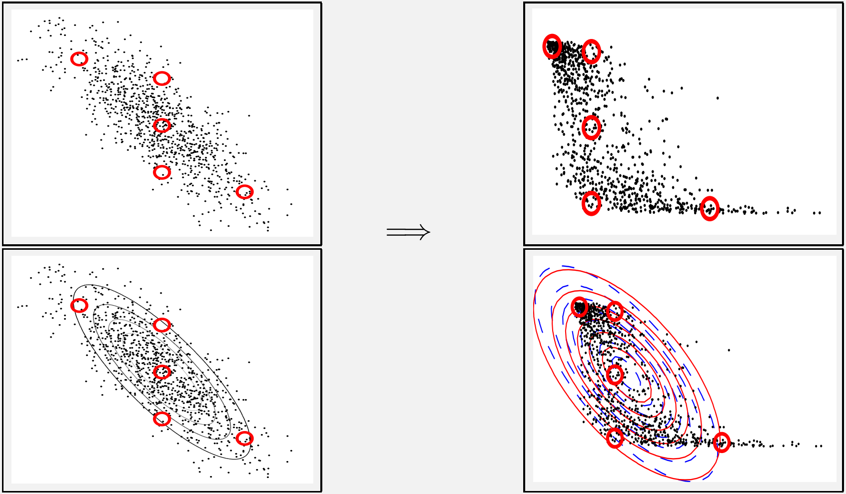
EKF Approximation



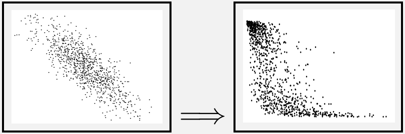
⟹



UKF Approximation



Particle Filtering 1/2



- The idea is to form a weighted particle presentation $(\mathbf{x}^{(i)}, w^{(i)})$ of the posterior distribution:

$$p(\mathbf{x}) \approx \sum_i w^{(i)} \, \delta(\mathbf{x} - \mathbf{x}^{(i)}).$$

- Sequential importance sampling, with additional resampling step.
- Bootstrap filter (also called Condensation) is the simplest particle filter.

Particle Filtering 2/2

- The efficiency of particle filter is determined by the selection of the importance distribution.
- The importance distribution can be formed by using e.g. EKF or UKF.
- Sometimes the optimal importance distribution can be used, and it minimizes the variance of the weights.
- Rao-Blackwellization: Some components of the model are marginalized in closed form $\Rightarrow$ hybrid particle/Kalman filter.

Summary

- Optimal filtering considers recursive Bayesian estimation of discrete-time, continuous-discrete-time or continuous-time state space models.
- The estimator for linear Gaussian models is called the Kalman filter.
- The formal solutions of non-linear filtering problems are known, but computationally intractable.
- The optimal estimators of non-linear and non-Gaussian models can be approximated with e.g. extended Kalman filters, unscented Kalman filters and particle filters.