Lecture 6: Multiple Model Filtering, Particle Filtering and Other Approximations

Simo Särkkä

Department of Biomedical Engineering and Computational Science
Aalto University

April 28, 2010

Contents

1. Multiple Model Filtering
2. Particle Filtering
3. Particle Filtering Properties
4. Further Filtering Algorithms
5. Continuous-Discrete-Time EKF
6. General Continuous-Discrete-Time Filtering
7. Continuous-Time Filtering
8. Linear Stochastic Differential Equations
9. What is Beyond This?
10. Summary

Multiple Model Kalman Filtering

- Algorithm for estimating true model or its parameter from a finite set of alternatives.
- Assume that we are given N possible dynamic/measurement models, and one of them is true.
- If s is the model index, the problem can be written in form:

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

  where $\mathbf{q}_{k-1} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}(s))$ and $\mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}(s))$.
- Can be solved in closed form with s parallel Kalman filters.

Interacting Multiple Models (IMM) Filter

- Assume that we have N possible models, but the true model is assumed to change in time.
- If the model index s_k is modeled as Markov chain, we have:

$$
\begin{aligned}
s_k &\sim P(s_k \mid s_{k-1}) \\
\mathbf{x}_k &= \mathbf{A}(s_{k-1})\,\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\
\mathbf{y}_k &= \mathbf{H}(s_k)\,\mathbf{x}_k + \mathbf{r}_k,
\end{aligned}
$$

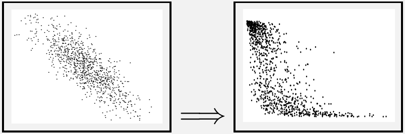
- Closed form solution would require running Kalman filters for each possible history $s_{1:k} \Rightarrow N^k$ filters, not feasible.
- Interacting Multiple Models (IMM) filter is an approximation algorithm, which uses only N parallel Kalman filters.

Particle Filtering: Overview [1/3]

Demo: Kalman vs. Particle Filtering:

- Kalman filter animation
- Particle filter animation

Particle Filtering: Overview [2/3]



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

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

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

Particle Filtering: Overview [3/3]

- 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.

Bootstrap Filter: Principle

- State density representation is set of samples $\{\mathbf{x}_k^{(i)} \; : \; i = 1, \ldots, N\}$.
- Bootstrap filter performs optimal filtering update and prediction steps using Monte Carlo.
- Prediction step performs prediction for each particle separately.
- Equivalent to integrating over the distribution of previous step (as in Kalman Filter).
- Update step is implemented with weighting and resampling.

Bootstrap Filter: Algorithm

Bootstrap Filter

1. Generate sample from predictive density of each old sample point $\mathbf{x}_{k-1}^{(i)}$:

$$\tilde{\mathbf{x}}_k^{(i)} \sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}).$$

2. Evaluate and normalize weights for each new sample point $\tilde{\mathbf{x}}_k^{(i)}$:

$$w_k^{(i)} = p(\mathbf{y}_k \mid \tilde{\mathbf{x}}_k^{(i)}).$$

3. Resample by selecting new samples $\mathbf{x}_k^{(i)}$ from set $\{\tilde{\mathbf{x}}_k^{(i)}\}$ with probabilities proportional to $w_k^{(i)}$.

Sequential Importance Resampling: Principle

- State density representation is set of weighted samples $\{(\mathbf{x}_k^{(i)}, w_k^{(i)}) \, : \, i = 1, \ldots, N\}$ such that for arbitrary function $\mathbf{g}(\mathbf{x}_k)$, we have

$$\mathrm{E}[\mathbf{g}(\mathbf{x}_k) \,|\, \mathbf{y}_{1:k}] \approx \sum_i w_k^{(i)} \, \mathbf{g}(\mathbf{x}_k^{(i)}).$$

- On each step, we first draw samples from an importance distribution $\pi(\cdot)$, which is chosen suitably.
- The prediction and update steps are combined and consist of computing new weights from the old ones $w_{k-1}^{(i)} \rightarrow w_k^{(i)}$.
- If the “sample diversity” i.e the effective number of different samples is too low, do resampling.

Sequential Importance Resampling: Algorithm

Sequential Importance Resampling

1. Draw new point $\mathbf{x}_k^{(i)}$ for each point in the sample set $\{\mathbf{x}_{k-1}^{(i)}, i = 1, \ldots, N\}$ from the importance distribution:

$$\mathbf{x}_k^{(i)} \sim \pi(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_{1:k}), \qquad i = 1, \ldots, N.$$

2. Calculate new weights

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(\mathbf{y}_k \mid \mathbf{x}_k^{(i)})\, p(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)})}{\pi(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_{1:k})}, \qquad i = 1, \ldots, N.$$

   and normalize them to sum to unity.

3. If the effective number of particles is too low, perform resampling.

Effective Number of Particles and Resampling

The estimate for the effective number of particles can be computed as:

$$n_{\text{eff}} \approx \frac{1}{\sum_{i=1}^{N} \left(w_k^{(i)} \right)^2},$$

Resampling

1. Interpret each weight $w_k^{(i)}$ as the probability of obtaining the sample index i in the set $\{\mathbf{x}_k^{(i)} \mid i = 1, \ldots, N\}$.
2. Draw N samples from that discrete distribution and replace the old sample set with this new one.
3. Set all weights to the constant value $w_k^{(i)} = 1/N$.

Constructing the Importance Distribution

- Use the dynamic model as the importance distribution ⇒ Bootstrap filter.
- Use the optimal importance distribution

$$\pi(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{y}_{1:k}).$$

- Approximate the optimal importance distribution by UKF ⇒ unscented particle filter.
- Instead of UKF also EKF or SLF can be, for example, used.
- Simulate availability of optimal importance distribution ⇒ auxiliary SIR (ASIR) filter.

Rao-Blackwellized Particle Filtering: Principle [1/2]

- Consider a conditionally Gaussian model of the form

$$\begin{aligned}
\mathbf{s}_k &\sim p(\mathbf{s}_k \,|\, \mathbf{s}_{k-1}) \\
\mathbf{x}_k &= \mathbf{A}(\mathbf{s}_{k-1})\,\mathbf{x}_{k-1} + \mathbf{q}_k, \qquad \mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}) \\
\mathbf{y}_k &= \mathbf{H}(\mathbf{s}_k)\,\mathbf{x}_k + \mathbf{r}_k, \qquad \mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R})
\end{aligned}$$

- The model is of the form

$$\begin{aligned}
p(\mathbf{x}_k, \mathbf{s}_k | \mathbf{x}_{k-1}, \mathbf{s}_{k-1}) &= \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{A}(\mathbf{s}_{k-1})\mathbf{x}_{k-1}, \mathbf{Q})\, p(\mathbf{s}_k \,|\, \mathbf{s}_{k-1}) \\
p(\mathbf{y}_k \,|\, \mathbf{x}_k, \mathbf{s}_k) &= \mathrm{N}(\mathbf{y}_k \,|\, \mathbf{H}(\mathbf{s}_k), \mathbf{R})
\end{aligned}$$

- The full model is non-linear and non-Gaussian in general.
- But given the values $\mathbf{s}_k$ the model is Gaussian and thus Kalman filter equations can be used.

Rao-Blackwellized Particle Filtering: Principle [1/2]

- The idea of the Rao-Blackwellized particle filter:
  - Use Monte Carlo sampling to the values $\mathbf{s}_k$
  - Given these values, compute distribution of $\mathbf{x}_k$ with Kalman filter equations.
  - Result is a Mixture Gaussian distribution, where each particle consist of value $\mathbf{s}_k^{(i)}$, associated weight $w_k^{(i)}$ and the mean and covariance conditional to the history $\mathbf{s}_{1:k}^{(i)}$
- The explicit RBPF equations can be found in the lecture notes.
- If the model is almost conditionally Gaussian, it is also possible to use EKF, SLF or UKF instead of the linear KF.

Particle Filter: Advantages

- No restrictions in model – can be applied to non-Gaussian models, hierarchical models etc.
- Global approximation.
- Approaches the exact solution, when the number of samples goes to infinity.
- In its basic form, very easy to implement.
- Superset of other filtering methods – Kalman filter is a Rao-Blackwellized particle filter with one particle.

Particle Filter: Disadvantages

- Computational requirements much higher than of the Kalman filters.
- Problems with nearly noise-free models, especially with accurate dynamic models.
- Good importance distributions and efficient Rao-Blackwellized filters quite tricky to implement.
- Very hard to find programming errors (i.e., to debug).

Variational Kalman Smoother

- Variation Bayesian analysis based framework for estimating the parameters of linear state space models.
- Idea: Fix $\mathbf{Q} = \mathbf{I}$ and assume that the joint distribution of states $\mathbf{x}_1, \ldots, \mathbf{x}_T$ and parameters $\mathbf{A}, \mathbf{H}, \mathbf{R}$ is approximately separable:

$$
\begin{aligned}
& p(\mathbf{x}_1, \ldots, \mathbf{x}_T, \mathbf{A}, \mathbf{H}, \mathbf{R} \mid \mathbf{y}_1, \ldots, \mathbf{y}_T) \\
& \qquad \approx p(\mathbf{x}_1, \ldots, \mathbf{x}_T \mid \mathbf{y}_1, \ldots, \mathbf{y}_T) \, p(\mathbf{A}, \mathbf{H}, \mathbf{R} \mid \mathbf{y}_1, \ldots, \mathbf{y}_T).
\end{aligned}
$$

- The resulting EM-algorithm consist of alternating steps of smoothing with fixed parameters and estimation of new parameter values.
- The general equations of the algorithm are quite complicated and assume that all the model parameters are to be estimated.

Recursive Variational Bayesian Estimation of Noise Variances

- Algorithm for estimating unknown time-varying measurement variances.
- Assume that the joint filtering distribution of state and measurement noise variance is approximately separable:

$$p(\mathbf{x}_k, \sigma_k^2 \mid y_1, \ldots, y_k) \approx p(\mathbf{x}_k \mid y_1, \ldots, y_k) \, p(\sigma_k^2 \mid y_1, \ldots, y_k)$$

- Variational Bayesian analysis leads to algorithm, where the natural representation is

$$\begin{aligned} p(\sigma_k^2 \mid y_1, \ldots, y_k) &= \mathrm{InvGamma}(\sigma_k^2 \mid \alpha_k, \beta_k) \\ p(\mathbf{x}_k \mid y_1, \ldots, y_k) &= \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k). \end{aligned}$$

- The update step consists of a fixed-point iteration for computing new $\alpha_k, \beta_k, \mathbf{m}_k, \mathbf{P}_k$ from the old ones.

Outlier Rejection and Multiple Target Tracking

- Outlier Rejection / Clutter Modeling:
  - Probabilistic Data Association (PDA)
  - Monte Carlo Data Association (MCDA)
  - Multiple hypothesis tracking (MHT)
- Multiple Target Tracking
  - Multiple hypothesis tracking (MHT)
  - Joint Probabilistic Data Association (JPDA)
  - Rao-Blackwellized Particle Filtering (RBMCDA) for Multiple Target Tracking

Continuous-Discrete Pendulum Model

- Consider the pendulum model, which was first stated as

$$d^2\theta/dt^2 = -g \sin(\theta) + w(t)$$
$$y_k = \sin(\theta(t_k)) + r_k,$$

where $w(t)$ is "Gaussian white noise" and $r_k \sim \mathrm{N}(0, \sigma^2)$.

- With state $\mathbf{x} = (\theta, d\theta/dt)$, the model is of the abstract form

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}) + \mathbf{w}(t)$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}(t_k)) + \mathbf{r}_k$$

where $\mathbf{w}(t)$ has the covariance (spectral density) $\mathbf{Q}_c$.

- Continuous-time dynamics + discrete-time measurement = Continuous-discrete (-time) filtering model.

Discretization of Continuous Dynamics [1/4]

- Previously we assumed that the measurements are obtained at times $t_k = 0, \Delta t, 2\Delta t, \ldots$
- The state space model was then Euler-discretized as

$$
\begin{aligned}
\mathbf{x}_k &= \mathbf{x}_{k-1} + \mathbf{f}(\mathbf{x}_{k-1})\,\Delta t + \mathbf{q}_{k-1} \\
\mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k
\end{aligned}
$$

- But what should be the variance of $\mathbf{q}_k$?
- Consistency: The same variance for single step of length Δt, and 2 steps of length $\Delta t/2$:

$$
\mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_c\,\Delta t)
$$

Discretization of Continuous Dynamics [2/4]

- Now the Extended Kalman fiter (EKF) for this model is
  - Prediction:

$$
\begin{aligned}
\mathbf{m}_k^- &= \mathbf{m}_{k-1} + \mathbf{f}(\mathbf{m}_{k-1})\,\Delta t \\
\mathbf{P}_k^- &= (\mathbf{I} + \mathbf{F}\,\Delta t)\,\mathbf{P}_{k-1}\,(\mathbf{I} + \mathbf{F}\,\Delta t)^T + \mathbf{Q}_c\,\Delta t \\
&= \mathbf{P}_{k-1} + \mathbf{F}\,\mathbf{P}_{k-1}\,\Delta t + \mathbf{P}_{k-1}\,\mathbf{F}^T\,\Delta t \\
&\quad + \mathbf{F}\,\mathbf{P}_{k-1}\,\mathbf{F}^T\,\Delta t^2 + \mathbf{Q}_c\,\Delta t
\end{aligned}
$$

  - Update:

$$
\begin{aligned}
\mathbf{S}_k &= \mathbf{H}(\mathbf{m}_k^-)\,\mathbf{P}_k^-\,\mathbf{H}^T(\mathbf{m}_k^-) + \mathbf{R} \\
\mathbf{K}_k &= \mathbf{P}_k^-\,\mathbf{H}^T(\mathbf{m}_k^-)\,\mathbf{S}_k^{-1} \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\,[\mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-)] \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T
\end{aligned}
$$

Discretization of Continuous Dynamics [3/4]

- But what happens if Δt is not "small", that is, if we get measurements quite rarely?
  - We can use more Euler steps between measurements.
  - We can perform the EKF prediction on each step.
  - We can even compute the limit of infinite number of steps.
- If we let $\delta t = \Delta t / n$, the prediction becomes:

$$
\begin{aligned}
&\hat{\mathbf{m}}_0 = \mathbf{m}_{k-1}; \quad \hat{\mathbf{P}}_0 = \mathbf{P}_{k-1} \\
&\text{for } i = 1 \ldots n \\
&\quad \hat{\mathbf{m}}_i = \hat{\mathbf{m}}_{i-1} + \mathbf{f}(\hat{\mathbf{m}}_{i-1})\, \delta t \\
&\quad \hat{\mathbf{P}}_i = \hat{\mathbf{P}}_{i-1} + \mathbf{F}\, \hat{\mathbf{P}}_{i-1}\, \delta t + \hat{\mathbf{P}}_{i-1}\, \mathbf{F}^T\, \delta t \\
&\qquad + \mathbf{F}\, \hat{\mathbf{P}}_{i-1}\, \mathbf{F}^T\, \delta t^2 + \mathbf{Q}_c\, \delta t \\
&\text{end} \\
&\mathbf{m}_k^- = \hat{\mathbf{m}}_n; \quad \mathbf{P}_k^- = \hat{\mathbf{P}}_n.
\end{aligned}
$$

Discretization of Continuous Dynamics [4/4]

- By re-arranging the equations in the for-loop, we get

$$(\hat{\mathbf{m}}_i - \hat{\mathbf{m}}_{i-1})/\delta t = \mathbf{f}(\hat{\mathbf{m}}_{i-1})$$

$$(\hat{\mathbf{P}}_i - \hat{\mathbf{P}}_{i-1})/\delta t = \mathbf{F}\,\hat{\mathbf{P}}_{i-1} + \hat{\mathbf{P}}_{i-1}\,\mathbf{F}^T + \mathbf{F}\,\hat{\mathbf{P}}_{i-1}\,\mathbf{F}^T\,\delta t + \mathbf{Q}_c$$

- In the limit $\delta t \to 0$, we get the differential equations

$$d\hat{\mathbf{m}}/dt = \mathbf{f}(\hat{\mathbf{m}}(t))$$

$$d\hat{\mathbf{P}}/dt = \mathbf{F}(\hat{\mathbf{m}}(t))\,\hat{\mathbf{P}}(t) + \hat{\mathbf{P}}(t)\,\mathbf{F}^T(\hat{\mathbf{m}}(t)) + \mathbf{Q}_c$$

- The initial conditions are

$$\hat{\mathbf{m}}(0) = \mathbf{m}_{k-1}$$

$$\hat{\mathbf{P}}(0) = \mathbf{P}_{k-1}$$

- The final prediction is

$$\mathbf{m}_k^- = \hat{\mathbf{m}}(\Delta t)$$

$$\mathbf{P}_k^- = \hat{\mathbf{P}}(\Delta t)$$

Continuous-Discrete EKF

Continuous-Discrete EKF

- **Prediction:** between the measurements integrate the following differential equations from t_{k-1} to t_k:

$$
\begin{aligned}
d\mathbf{m}/dt &= \mathbf{f}(\mathbf{m}(t)) \\
d\mathbf{P}/dt &= \mathbf{F}(\mathbf{m}(t))\,\mathbf{P}(t) + \mathbf{P}(t)\,\mathbf{F}^T(\mathbf{m}(t)) + \mathbf{Q}_c
\end{aligned}
$$

- **Update:** at measurements do the EKF update

$$
\begin{aligned}
\mathbf{S}_k &= \mathbf{H}(\mathbf{m}_k^-)\,\mathbf{P}_k^-\,\mathbf{H}^T(\mathbf{m}_k^-) + \mathbf{R} \\
\mathbf{K}_k &= \mathbf{P}_k^-\,\mathbf{H}^T(\mathbf{m}_k^-)\,\mathbf{S}_k^{-1} \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\,[\mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-)] \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\,\mathbf{S}_k\,\mathbf{K}_k^T,
\end{aligned}
$$

where $\mathbf{m}_k^-$ and $\mathbf{P}_k^-$ are the results of the prediction step.

Continuous-Discrete SLF, UKF, PF etc.

- The equations

$$
\begin{aligned}
d\mathbf{m}/dt &= \mathbf{f}(\mathbf{m}(t)) \\
d\mathbf{P}/dt &= \mathbf{F}(\mathbf{m}(t))\,\mathbf{P}(t) + \mathbf{P}(t)\,\mathbf{F}^T(\mathbf{m}(t)) + \mathbf{Q}_c
\end{aligned}
$$

  actually generate a Gaussian process approximation $\mathbf{x}(t) \sim \mathrm{N}(\mathbf{m}(t), \mathbf{P}(t))$ to the solution of non-linear stochastic differential equation (SDE)

$$
d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}) + \mathbf{w}(t)
$$

- We could also use statistical linearization or unscented transform and get a bit different limiting differential equations.
- Also possible to generate particle approximations by a continuous-time version of importance sampling (based on Girsanov theorem).

More general SDE Theory

- The most general SDE model usually considered is of the form

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}) + \mathbf{L}(\mathbf{x})\,\mathbf{w}(t)$$

- Formally, $\mathbf{w}(t)$ is a Gaussian white noise process with zero mean and covariance function

$$\mathrm{E}[\mathbf{w}(t)\,\mathbf{w}^T(t')] = \mathbf{Q}_c\,\delta(t'-t)$$

- The distribution $p(\mathbf{x}(t))$ is non-Gaussian and it is given by the following partial differential equation:

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

- Known as Fokker-Planck equation or Kolmogorov forward equation.

More general SDE Theory (cont.)

- In more rigorous theory, we actually must interpret the SDE as integral equation

$$\mathbf{x}(t) - \mathbf{x}(s) = \int_s^t \mathbf{f}(\mathbf{x})\, dt + \int_s^t \mathbf{L}(\mathbf{x})\, \mathbf{w}(t)\, dt$$

- In Ito's theory of SDE's the second integral is defined as stochastic integral w.r.t. Brownian motion $\beta(t)$:

$$\mathbf{x}(t) - \mathbf{x}(s) = \int_s^t \mathbf{f}(\mathbf{x})\, dt + \int_s^t \mathbf{L}(\mathbf{x})\, d\beta(t)$$

  i.e., formally $\mathbf{w}(t)\, dt = d\beta(t)$ or $\mathbf{w}(t) = d\beta(t)/dt$
- However, Brownian motion is nowhere differentiable!
- Brownian motion is also called as Wiener process.

More general SDE Theory (cont. II)

- In stochastics, the integral equation is often written as

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x})\, dt + \mathbf{L}(\mathbf{x})\, d\beta(t)$$

- In engineering (control theory, physics) it is customary to formally divide with *dt* to get

$$d\mathbf{x}(t)/dt = \mathbf{f}(\mathbf{x}) + \mathbf{L}(\mathbf{x})\, \mathbf{w}(t)$$

- So called Stratonovich's theory is more consistent with this white noise interpretation than Ito's theory.
- In mathematical sense Stratonovich's theory defines the stochastic integral $\int_s^t \mathbf{L}(\mathbf{x})\, d\beta(t)$ a bit differently – also the Fokker-Planck equation is different.

Cautions About White Noise

- White noise is actually only formally defined as derivative of Brownian motion.
- White noise can only be defined in distributional sense – for this reason non-linear functions of it $g(\mathbf{w}(t))$ are not well-defined.
- For this reason, the following more general type of SDE does not make sense:

$$d\mathbf{x}(t)/dt = \mathbf{f}(\mathbf{x}, \mathbf{w})$$

- We must also be careful in interpreting the multiplicative term in the equation

$$d\mathbf{x}(t)/dt = \mathbf{f}(\mathbf{x}) + \mathbf{L}(\mathbf{x})\,\mathbf{w}(t)$$

Formal Optimal Continuous-Discrete Filter

Optimal continuous-discrete filter

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

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

2. Update step: Apply the Bayes' rule.

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

General Continuous-Time Filtering

- We could also model the measurements as a continuous-time process:

$$dx/dt = \mathbf{f}(\mathbf{x}) + \mathbf{L}(\mathbf{x})\,\mathbf{w}(t)$$
$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{n}(t)$$

- Again, one must be very careful in interpreting the white noise processes $\mathbf{w}(t)$ and $\mathbf{n}(t)$.
- The filtering equations become a stochastic partial differential equation (SPDE) called Kushner-Stratonovich equation.
- The equation for the unnormalized filtering density is called the Zakai equation, which also is a SPDE.
- It is also possible to take the continuous-time limit of the Bayesian smoothing equations (result is a PDE).

Kalman-Bucy Filter

- If the system is linear

$$
\begin{aligned}
d\mathbf{x}/dt &= \mathbf{F}\,\mathbf{x} + \mathbf{w}(t) \\
\mathbf{y} &= \mathbf{H}\,\mathbf{x} + \mathbf{n}(t)
\end{aligned}
$$

  we get the continuous-time Kalman-Bucy filter:

$$
\begin{aligned}
d\mathbf{m}/dt &= \mathbf{F}\,\mathbf{m} + \mathbf{K}\,(\mathbf{y} - \mathbf{H}\,\mathbf{m}) \\
d\mathbf{P}/dt &= \mathbf{F}\,\mathbf{P} + \mathbf{P}\,\mathbf{F}^T + \mathbf{Q}_c - \mathbf{K}\,\mathbf{R}\,\mathbf{K}^T,
\end{aligned}
$$

  where $\mathbf{K} = \mathbf{P}\,\mathbf{H}^T\,\mathbf{R}^{-1}$.
- The stationary solution to these equations is equivalent to the continuous-time Wiener filter.
- Non-linear extensions (EKF, SLF, UKF, etc.) can be obtained similarly to the discrete-time case.

Solution of LTI SDE

- Let's return to linear stochastic differential equations:

$$d\mathbf{x}/dt = \mathbf{F}\,\mathbf{x} + \mathbf{w}$$

- Assume that $\mathbf{F}$ is time-independent. For example, in car-tracking model we had a model of this type.
- Given $\mathbf{x}(0)$ we can now actually solve the equation

$$\mathbf{x}(t) = \exp(t\,\mathbf{F})\,\mathbf{x}(0) + \int_0^t \exp((t-s)\,\mathbf{F})\,\mathbf{w}(s)\,ds,$$

where exp(.) is the matrix exponential function:

$$\exp(t\mathbf{F}) = \mathbf{I} + t\mathbf{F} + \frac{1}{2!}t^2\mathbf{F}^2 + \frac{1}{3!}t^3\mathbf{F}^3 + \ldots$$

- Note that we are treating $\mathbf{w}(s)$ as an ordinary function, which is not generally justified!

Solution of LTI SDE (cont.)

- We can also solve the equation on predefined time points $t_1, t_2, \ldots$ as follows:

$$\mathbf{x}(t_k) = \exp((t_k - t_{k-1})\,\mathbf{F})\,\mathbf{x}(t_{k-1}) + \int_{t_{k-1}}^{t_k} \exp((t_k - s)\,\mathbf{F})\,\mathbf{w}(s)\,ds$$

- The first term is of the form $\mathbf{A}\,\mathbf{x}(t_{k-1})$, where the matrix is a known constant $\mathbf{A} = \exp(\Delta t\,\mathbf{F})$.
- The second term is a zero mean Gaussian random variable and its covariance can be calculated as:

$$\begin{aligned}\mathbf{Q} &= \int_{t_{k-1}}^{t_k} \exp((t_k - s)\,\mathbf{F})\,\mathbf{Q}_c\,\exp((t_k - s)\,\mathbf{F})^T\,ds \\ &= \int_0^{\Delta t} \exp((\Delta t - s)\,\mathbf{F})\,\mathbf{Q}_c\,\exp((\Delta t - s)\,\mathbf{F})^T\,ds\end{aligned}$$

Solution of LTI SDE (cont. II)

- Thus the continuous-time system is in a sense equivalent to the discrete-time system

$$\mathbf{x}(t_k) = \mathbf{A}\,\mathbf{x}(t_{k-1}) + \mathbf{q}_k$$

  where $\mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q})$ and

$$\mathbf{A} = \exp(\Delta t\, A)$$

$$\mathbf{Q} = \int_0^{\Delta t} exp((\Delta t - s)\,\mathbf{F})\,\mathbf{Q}_c \exp((\Delta t - s)\,\mathbf{F})^T ds$$

- An analogous equivalent discretization is also possible with time-varying linear stochastic differential equation models.
- A continuous-discrete Kalman filter can be always implemented as a discrete-time Kalman filter by forming the equivalent discrete-time system.

Wiener Velocity Model

- For example, consider the Wiener velocity model (= white noise acceleration model):

$$d^2x/dt^2 = w(t),$$

which is equivalent to the state space model

$$d\mathbf{x}/dt = \mathbf{F}\,\mathbf{x} + \mathbf{w}$$

with $\mathbf{F} = (0\ 1;\ 0\ 0), \mathbf{x} = (x, dx/dt), \mathbf{Q}_c = (0\ 0;\ 0\ q)$.

- Then we have

$$\mathbf{A} = \exp(\Delta t\,\mathbf{F}) = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}$$

$$\begin{aligned} \mathbf{Q} &= \int_0^{\Delta t} \exp((\Delta t - s)\,\mathbf{F})\,\mathbf{Q}_c \exp((\Delta t - s)\,\mathbf{F})^T ds \\ &= \begin{pmatrix} \Delta t^3/3\,q & \Delta t^2/2\,q \\ \Delta t^2/2\,q & \Delta t\,q \end{pmatrix} \end{aligned}$$

which might look familiar.

Mean and Covariance Differential Equations

- Note that in the linear (time-invariant) case

$$d\mathbf{x}/dt = \mathbf{F}\,\mathbf{x} + \mathbf{w}$$

we could also write down the differential equations

$$\begin{aligned} d\mathbf{m}/dt &= \mathbf{F}\,\mathbf{m} \\ d\mathbf{P}/dt &= \mathbf{F}\,\mathbf{P} + \mathbf{P}\,\mathbf{F}^T + \mathbf{Q}_c \end{aligned}$$

which exactly give the evolution of mean and covariance.

- The solutions of these equations are

$$\begin{aligned} \mathbf{m}(t) &= \exp(t\,\mathbf{F})\,\mathbf{m}_0 \\ \mathbf{P}(t) &= \exp(t\,\mathbf{F})\,\mathbf{P}_0\,\exp(t\,\mathbf{F})^T \\ &\quad + \int_0^t \exp((t-s)\,\mathbf{F})\,\mathbf{Q}_c\,\exp((t-s)\,\mathbf{F})^T\,ds, \end{aligned}$$

which are consistent with the previous results.

Optimal Control Theory

- Assume that the physical system can be modeled with differential equation with input $\mathbf{u}$

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{u})$$

- Determine $\mathbf{u}(t)$ such that $\mathbf{x}(t)$ and $\mathbf{u}(t)$ satisfy certain constraints and minimize a cost functional.
- For example, steer a space craft to moon such that the consumed of fuel is minimized.
- If the system is linear and cost function quadratic, we get linear quadratic controller (or regulator).

Stochastic (Optimal) Control Theory

- Assume that the system model is stochastic:

$$
\begin{aligned}
d\mathbf{x}/dt &= \mathbf{f}(\mathbf{x}, \mathbf{u}) + \mathbf{w}(t) \\
\mathbf{y}_k &= \mathbf{h}(\mathbf{x}(t_k)) + \mathbf{r}_k
\end{aligned}
$$

- Given only the measurements $\mathbf{y}_k$, find $\mathbf{u}(t)$ such that $\mathbf{x}(t)$ and $\mathbf{u}(t)$ satisfy the constraints and minimize a cost function.
- If linear Gaussian, we have
  Linear Quadratic (LQ) controller + Kalman filter = Linear Quadratic Gaussian (LQG) controller
- In general, not simply a combination of optimal filter and deterministic optimal controller.
- Model Predictive Control (MPC) is a well-known approximation algorithm for constrained problems.

- Infinite dimensional generalization of state space model is the stochastic partial differential equation (SPDE)

$$\frac{\partial \mathbf{x}(t,\mathbf{r})}{\partial t} = \mathscr{F}_r \, \mathbf{x}(t,\mathbf{r}) + \mathscr{L}_r \, \mathbf{w}(t,\mathbf{r}),$$

  where $\mathscr{F}_r$ and $\mathscr{L}_r$ are linear operators (e.g. integro-differential operators) in $\mathbf{r}$-variable and $\mathbf{w}(\cdot)$ is a time-space white noise.
- Practically every SPDE can be converted into this form with respect to any variable (which is relabeled as t).
- For example, stochastic heat equation

$$\frac{\partial x(t,r)}{\partial t} = \frac{\partial^2 x(t,r)}{\partial r^2} + w(t,r).$$

Spatially Distributed Systems (cont.)

- The solution to the SPDE is analogous to finite-dimensional case:

$$\mathbf{x}(t, \mathbf{r}) = \mathscr{U}_r(t)\, \mathbf{x}(0, \mathbf{r}) + \int_0^t \mathscr{U}_r(t - s)\, \mathscr{L}_r\, \mathbf{w}(s, \mathbf{r})\, ds.$$

- $\mathscr{U}_r(t) = \exp(t\, \mathscr{F}_r)$ is the evolution operator – corresponds to propagator in quantum mechanics.
- Spatio-temporal Gaussian process models can be naturally formulated as linear SPDE's.
- Recursive Bayesian estimation with SPDE models lead to infinite-dimensional Kalman filters and RTS smoothers.
- SPDE's can be approximated with finite models by usage of finite-differences or finite-element methods.

Summary

- Other filtering algorithms than EKF, SLF, UKF and PF are, for example, multiple model Kalman filters and IMM algorithm.
- Particle filters use weighted set of samples (particles) for approximating the filtering distributions.
- Sequential importance resampling (SIR) is the general framework and bootstrap filter is a simple special case of it.
- In Rao-Blackwellized particle filters a part of the state is sampled and part is integrated in closed form with Kalman filter.
- Specialized filtering algorithms exist also, e.g., for parameter estimation, outlier rejection and multiple target tracking.

Summary (cont.)

- In continuous-discrete filtering, the dynamic model is a continuous-time process and measurement are obtained at discrete times.
- In continuous-discrete EKF, SLF and UKF the continuous-time non-linear dynamic model is approximated as a Gaussian process.
- In continuous-time filtering, the both the dynamic and measurements models are continuous-time processes.
- The theories of continuous and continuous-discrete filtering are tied to the theory of stochastic differential equations.