Lecture 2: Single object tracking in clutter
Version April 29, 2019

Multi-Object Tracking

Lennart Svensson

Section 1:
Introduction to SOT in clutter

Multi-Object Tracking

Lennart Svensson

An introduction to single object tracking in clutter

Multi-Object Tracking

Lennart Svensson

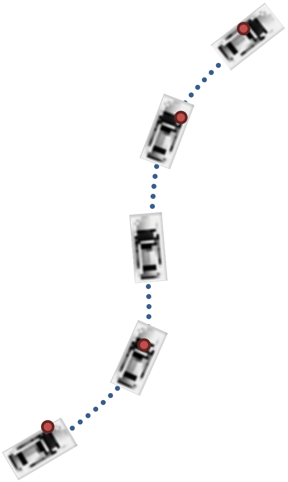
DEFINITION OF SINGLE OBJECT TRACKING (SOT)

- SOT in clutter is a special case of multi-object tracking (MOT).

Key property

- In SOT, there is always **precisely one object present** at all times.

- Easier than MOT? Yes!
  - No need to infer the number of objects, or when objects appear/disappear.
  - Many fewer data association hypotheses.



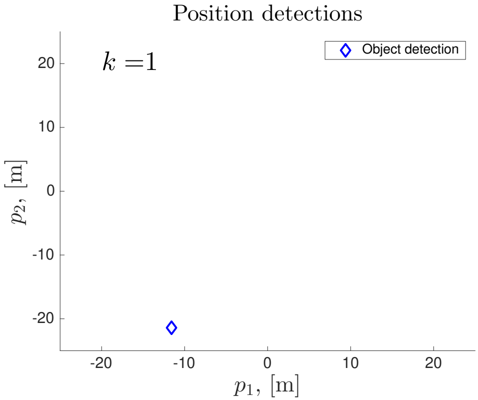
SOT IN CLUTTER: MAIN CHALLENGES

Previous challenges still relevant:

- Time varying state variables.
- Noisy measurements.

New challenges

- missed detections
- clutter detections
- unknown data associations.



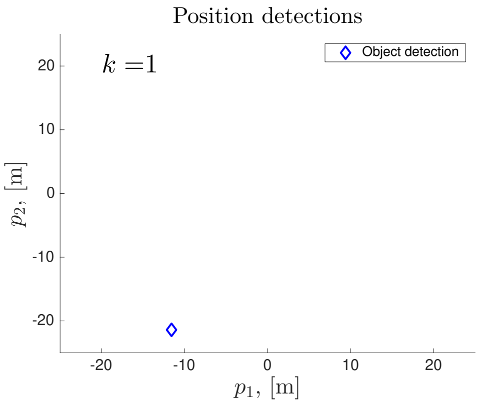
SOT IN CLUTTER: MAIN CHALLENGES

Previous challenges still relevant:

- Time varying state variables.
- Noisy measurements.

New challenges

- missed detections
- clutter detections
- unknown data associations.



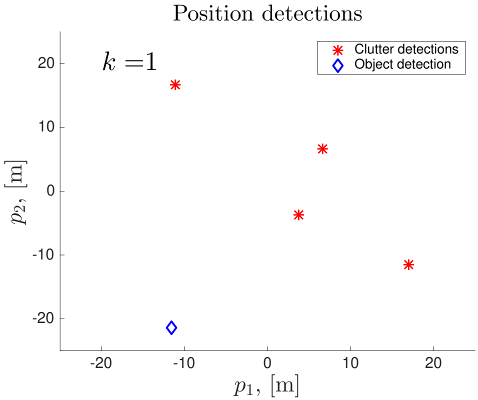
SOT IN CLUTTER: MAIN CHALLENGES

Previous challenges still relevant:

- Time varying state variables.
- Noisy measurements.

New challenges

- missed detections
- clutter detections
- unknown data associations.



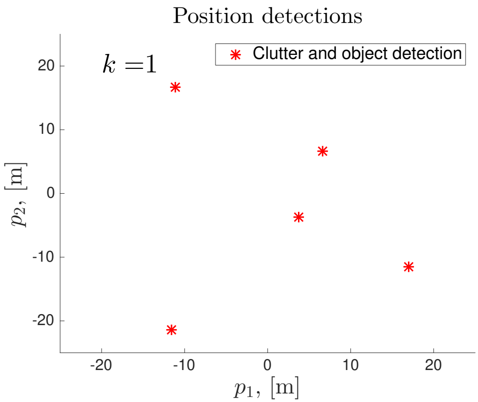
SOT IN CLUTTER: MAIN CHALLENGES

Previous challenges still relevant:

- Time varying state variables.
- Noisy measurements.

New challenges

- missed detections
- clutter detections
- unknown data associations.



WHY STUDY SOT IN CLUTTER?

Adding complexity bit by bit

- A simple(r) setting to learn about:
  - measurement models,
  - data association uncertainties,
  - tools to approximate posterior density.

- Important subproblem of MOT.
  - In some settings we only have one object to track (robot, athlete, vehicle).
  - Radar systems use SOT to control sensors and direct radar towards object.

Section 2: Motion and measurement models

Multi-Object Tracking

Lennart Svensson

Single object motion and measurement models

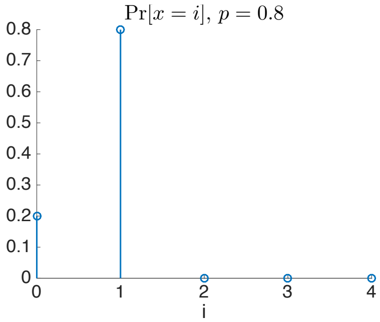
Multi-Object Tracking

Lennart Svensson

PRELIMINARIES – BERNOULLI DISTRIBUTION

- The **Bernoulli distribution** is central to both SOT and MOT.
- If x is Bernoulli distributed with probability, $p \in [0, 1]$, it takes the values

$$x = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1 - p. \end{cases}$$

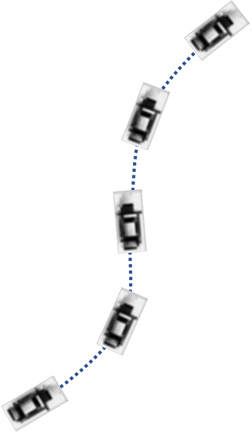


SINGLE OBJECT MOTION MODELS

Motion model

- The object state is a Markov chain that evolves as
$$p(x_k \middle| x_{k-1}) = \pi_k(x_k \middle| x_{k-1}).$$

- For instance, we often assume that
$$x_k = f_{k-1}(x_{k-1}) + q_{k-1}, \qquad q_{k-1} \sim \mathcal{N}(0, Q_{k-1}),$$
such that
$$\pi_k(x_k \middle| x_{k-1}) = N(x_k; f_{k-1}(x_{k-1}), Q_{k-1}).$$



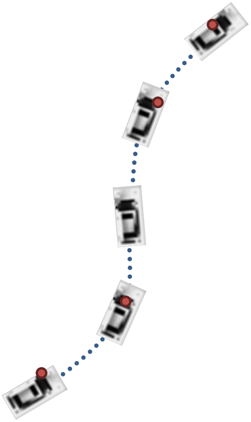
SINGLE OBJECT MEASUREMENT MODELS (1)

Measurement model

- The object is detected with probability $P^{\mathrm{D}}(x_k)$, and then generates a measurement from
$$p(o_k | x_k) = g_k(o_k | x_k).$$

- For instance, we often assume that
$$o_k = h_k(x_k) + v_k, \qquad v_k \sim \mathcal{N}(0, R_k),$$
such that
$$g_k(o_k | x_k) = N(o_k; h_k(x_k), R_k).$$

- **Note:** object is **not always detected**.



SINGLE OBJECT MEASUREMENT MODELS (2)

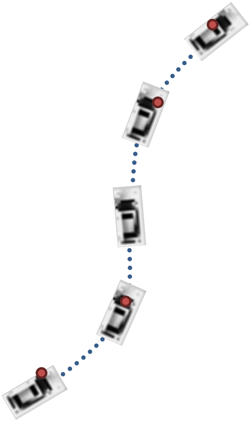
- We use a matrix (or a sequence) to represent the object detections,

$$O_k = \begin{cases} [\,] & \text{if object is undetected,} \\ o_k & \text{if object is detected.} \end{cases}$$

- We use $|O_k|$ to denote the number of column vectors in O_k.

- Given x_k, $|O_k|$ is Bernoulli distributed:

$$|O_k| = \begin{cases} 1 & \text{with probability } P^{\mathrm{D}}(x_k), \\ 0 & \text{with probability } 1 - P^{\mathrm{D}}(x_k). \end{cases}$$



SINGLE OBJECT MEASUREMENT MODEL (3)

- Using the matrix notation, we get

$$p(O_k | x_k) = \begin{cases} 1 - P^{\mathrm{D}}(x_k) & \text{if } O_k = [\,] \\ P^{\mathrm{D}}(x_k) g_k(o_k | x_k) & \text{if } O_k = o_k. \end{cases}$$

  Note: this captures both the probability of detection and, if detected, the distribution of the detection.

- Given x_k, the set of vectors in O_k is a **Bernoulli random finite set**.

SINGLE OBJECT MEASUREMENT MODEL (4)

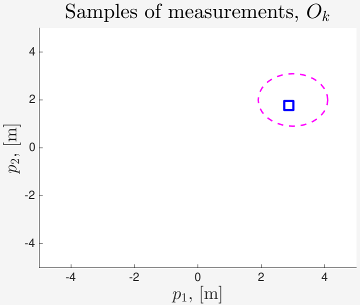
- Simple to generate object measurements O_k given x_k.

Algorithm Sampling O_k given x_k.

1: Initialize $O_k = [\,]$
2: **if** rand$< P^{\mathrm{D}}(x_k)$ **then**
3: $\quad o_k \sim g_k(\cdot \,|\, x_k)$
4: $\quad O_k = o_k$
5: **end if**

Example, samples of O_k

- Suppose $P^{\mathrm{D}}(x_k) = 0.85$ and $g_k(o_k|x_k) = \mathcal{N}(o_k; [3\ ,2]^T, 0.3\mathbf{I})$.



SOT with known associations

Multi-Object Tracking

Lennart Svensson

PROBLEM FORMULATION

- Suppose the data associations are known.
- Given a set of measurements, Z_k, we then know O_k.

Objective (this video)

- Recursively compute

$$p(x_k \,\middle|\, O_{1:k}).$$

PREDICTION STEP

- Given a motion model $\pi_k(x_k | x_{k-1})$, how can we perform prediction?

Chapman-Kolmogorov equation

- We use the Chapman-Kolmogorov equation

$$p(x_k | O_{1:k-1}) = \int \pi_k(x_k | x_{k-1}) p(x_{k-1} | O_{1:k-1}) \, \mathrm{d}x_{k-1}.$$

Linear and Gaussian models

- If $p(x_{k-1} | O_{1:k-1}) = \mathcal{N}(x_{k-1}; \bar{x}_{k-1|k-1}, P_{k-1|k-1})$ and $\pi_k(x_k | x_{k-1}) = \mathcal{N}(x_k; Fx_{k-1}, Q)$ then

$$p(x_k | O_{1:k-1}) = \mathcal{N}(x_k; F\bar{x}_{k-1|k-1}, FP_{k-1|k-1}F^T + Q).$$

UPDATE STEP

- Given a measurement model,

$$p(O_k|x_k) = \begin{cases} 1 - P^{\mathrm{D}}(x_k) & \text{if } O_k = [\,] \\ P^{\mathrm{D}}(x_k) g_k(o_k|x_k) & \text{if } O_k = o_k. \end{cases}$$

  how can we perform the update?

- **Recall:** posterior $\propto$ prior $\times$ likelihood.

Bayes' rule

- We get

$$\begin{aligned} p(x_k|O_{1:k}) &\propto p(x_k|O_{1:k-1}) p(O_k|x_k) \\ &= \begin{cases} p(x_k|O_{1:k-1})(1 - P^{\mathrm{D}}(x_k)) & \text{if } O_k = [\,] \\ p(x_k|O_{1:k-1}) P^{\mathrm{D}}(x_k) g_k(o_k|x_k) & \text{if } O_k = o_k. \end{cases} \end{aligned}$$

UPDATE STEP: INFORMATIVE $P^{\mathrm{D}}(x_k)$

Example 1: informative $P^{\mathrm{D}}(x_k)$

- Suppose we have a scalar state,

$$p(x_k \mid O_{1:k-1}) = \mathcal{N}(x_k; 0, 1).$$

- Consider a measurement model with $g_k(o_k \mid x_k) = p(o_k)$ and

$$P^D(x_k) = \begin{cases} 1 & \text{if } x_k \geq 0 \\ 0 & \text{if } x_k < 0. \end{cases}$$

- We get

$$p(x_k \mid O_{1:k}) \propto \begin{cases} p(x_k \mid O_{1:k-1})(1 - P^{\mathrm{D}}(x_k)) & \text{if } O_k = [\,] \\ p(x_k \mid O_{1:k-1}) P^{\mathrm{D}}(x_k) g_k(o_k \mid x_k) & \text{if } O_k = o_k. \end{cases}$$

UPDATE STEP: INFORMATIVE $P^{\mathrm{D}}(x_k)$

Example 1: informative $P^{\mathrm{D}}(x_k)$

- Suppose we have a scalar state,

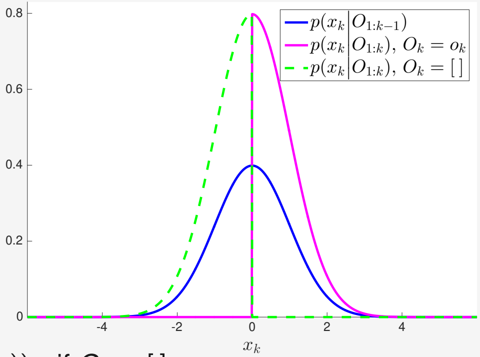
$$p(x_k | O_{1:k-1}) = \mathcal{N}(x_k; 0, 1).$$

- Consider a measurement model with $g_k(o_k | x_k) = p(o_k)$ and

$$P^D(x_k) = \begin{cases} 1 & \text{if } x_k \geq 0 \\ 0 & \text{if } x_k < 0. \end{cases}$$

- We get

$$p(x_k | O_{1:k}) \propto \begin{cases} p(x_k | O_{1:k-1})(1 - P^{\mathrm{D}}(x_k)) & \text{if } O_k = [\,] \\ p(x_k | O_{1:k-1}) P^{\mathrm{D}}(x_k) & \text{if } O_k = o_k. \end{cases}$$



UPDATE STEP: CONSTANT P^{D}

Example 2: constant $P^{\mathrm{D}}(x_k)$

- Suppose $P^{\mathrm{D}}(x_k)$ is constant.
- We have

$$p(x_k|O_{1:k}) \propto \begin{cases} p(x_k|O_{1:k-1})(1-P^{\mathrm{D}}(x_k)) & \text{if } O_k = [\,] \\ p(x_k|O_{1:k-1})P^{\mathrm{D}}(x_k)g_k(o_k|x_k) & \text{if } O_k = o_k, \end{cases}$$

  which simplifies to

$$p(x_k|O_{1:k}) \propto \begin{cases} p(x_k|O_{1:k-1}) & \text{if } O_k = [\,] \\ p(x_k|O_{1:k-1})g_k(o_k|x_k) & \text{if } O_k = o_k. \end{cases}$$

- In short, **standard update** using o_k but **only if object is detected**.

UPDATE STEP: CONSTANT P^{D}, LINEAR AND GAUSSIAN MODELS

Example 2: constant $P^{\mathrm{D}}(x_k)$, continued

- Specifically, consider

$$p(x_k \big| O_{1:k-1}) = \mathcal{N}(x_k; \bar{x}_{k|k-1}, P_{k|k-1}), \quad g_k(o_k \big| x_k) = \mathcal{N}(o_k; H_k x_k, R_k).$$

- Then, $p(x_k \big| O_{1:k}) = \mathcal{N}(x_k; \bar{x}_{k|k}, P_{k|k})$ where

$$\bar{x}_{k|k} = \bar{x}_{k|k-1}, \qquad P_{k|k} = P_{k|k-1}$$

when $O_k = [\,]$, and, when $O_k = o_k$,

$$\bar{x}_{k|k} = \bar{x}_{k|k-1} + K_k(o_k - H_k \bar{x}_{k|k-1})$$
$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1},$$

where $K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1}$ is the Kalman gain.

- **Standard Kalman filter update**, but only **if object is detected**.

PREDICTION AND UPDATE: ILLUSTRATION

Constant P^{D}, linear and Gaussian models

- **Motion model** (constant velocity):

$$x_k = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x_{k-1} + q_{k-1}$$

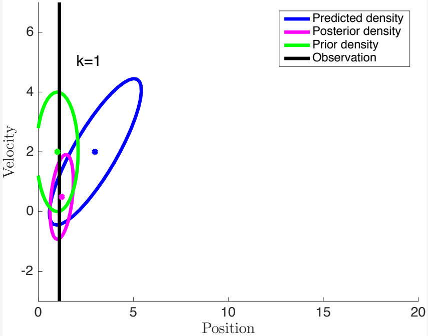
$$q_{k-1} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 0.5 \begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix} \right)$$

$$p(x_0) = \mathcal{N}\left(x_0; \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

- **Measurement model:**

$$P^{\mathrm{D}}(x_k) = 0.85$$

$$o_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k, \quad v_k \sim \mathcal{N}(0, 1).$$



Standard clutter model: motivation

Multi-Object Tracking

Lennart Svensson

PRELIMINARIES – BINOMIAL DISTRIBUTION

- This video uses two other scalar distributions, apart from the Bernoulli distribution.

Binomial distribution

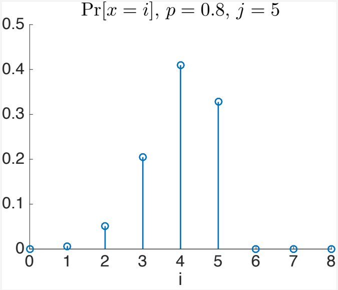
- If x is binomially distribution with parameters $p \in [0, 1]$ and $j \in \mathbb{N}$,

$$\Pr[x = i] = \binom{j}{i} p^i (1-p)^{j-i}$$

where

$$\binom{j}{i} = \frac{j!}{i!(j-i)!}.$$

Binomial example



- It holds that $\mathbb{E}[x] = pj$.

PRELIMINARIES – POISSON DISTRIBUTION

Poisson distribution

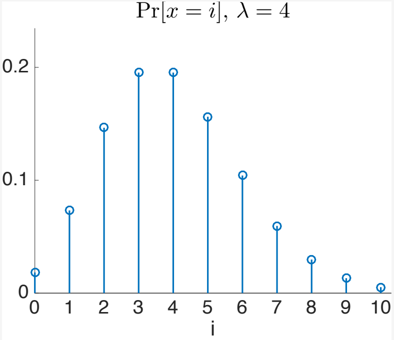
- If x is Poisson distributed with expected value $\lambda > 0$,

$$\Pr[x = i] = \mathrm{Po}[i; \lambda] = \frac{\lambda^i \exp(-\lambda)}{i!}.$$

- It is useful to know that

$$\mathbb{E}[x] = \mathrm{Var}(x) = \lambda.$$

Example of Poisson pmf



MODELLING CLUTTER

- Observed measurement matrix:

$$Z_k = \Pi(O_k, C_k),$$

where Π **randomly shuffles column vectors**, and C_k is **clutter**.

Parts of a clutter model

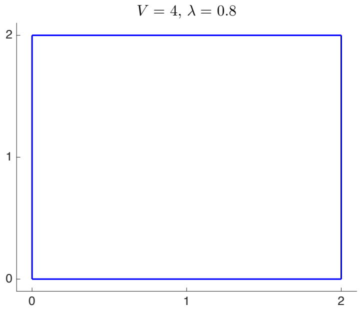
- We need a stochastic model for
  1. number of detections, $|C_k|$,
  2. vectors in C_k.

- Consider a field of view in $\mathbb{R}^{n_z}$ of volume V.
- Let λ denote "expected number of clutter detections per unit volume".

Example of Π

- If $Z = \Pi(o^1, c^1)$, then

$$\Pr\left[Z = [o^1, c^1]\right] = \Pr\left[Z = [c^1, o^1]\right] = 0.5.$$

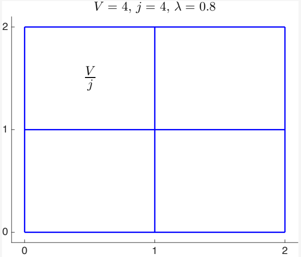


CLUTTER – LIMITED RESOLUTION

- Real sensors have limited resolution.
  $\Rightarrow$ Nearby objects generate at most one detection.

A possible clutter model

- Split volume into j cells, $C_k = \Pi(C_k^{(1)}, \cdots, C_k^{(j)})$, where $C_k^{(i)}$ denotes clutter in cell i.
- $C_k^{(1)}, \ldots, C_k^{(i)}$ are assumed independent.
- $|C_k^{(i)}|$ is Bernoulli with probability $\lambda V/j$.
- Detections are uniformly distribution within their cells.



- According to model, $|C_k|$ is binomially distributed with parameters j and $\lambda V/j$.

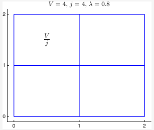
CLUTTER – UNLIMITED RESOLUTION

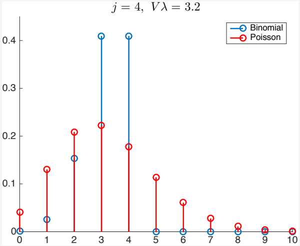
- In this course we assume unlimited sensor resolution.

A new clutter model

- Let us increase j and make cells smaller!
  Consequences:
  - Possible to obtain more detections.
  - Probability of detection in a single cell, $\lambda V/j$, decreases.
  - $\mathbb{E}\left[|C_k|\right] = V\lambda$ for all j.

- In the limit, as $j \to \infty$:
  - $|C_k|$ is Poisson distributed.
  - C_k is a **Poisson point process**.





Standard clutter model: the Poisson point process

Multi-Object Tracking

Lennart Svensson

POISSON POINT PROCESSES

- The Poisson point process (PPP) is the **default model for clutter** $C_k = [c_k^1, \ldots, c_k^{m_k^c}]$.

PPP introduced above

- The number of clutter is

$$m_k^c \sim \mathrm{Po}(\lambda V).$$

- Given m_k^c, the vectors $c_k^1, \ldots, c_k^{m_k^c}$ are i.i.d.

$$c_k^i \sim \mathrm{unif}(\mathbf{V}),$$

where $\mathbf{V}$ is our field of view.

Algorithm Sampling the PPP

```
Initialize C_k = [ ]
Generate m_k^c ~ Po(λV)
for i = 1 to m_k^c do
    Generate c_k^i ~ unif(V)
    Set C_k = [C_k, c_k^i]
end for
```

GENERAL PARAMETRIZATIONS OF PPPs

- More generally, we parametrize PPPs using either
  - an **intensity function**, $\lambda_c(c) \geq 0$, or
  - a combination of

$$\begin{cases} \bar{\lambda}_c = \int \lambda_c(c)\,\mathrm{d}c & \textbf{rate} \\ f_c(c) = \frac{\lambda_c(c)}{\bar{\lambda}_c} & \textbf{spatial pdf}. \end{cases}$$

- **Note:** the intensity can be computed from rate and spatial pdf:

$$\lambda_c(c) = \bar{\lambda}_c f_c(c).$$

In previous example

- Intensity function is

$$\lambda_c(c) = \begin{cases} \lambda & \text{if } c \in \mathbf{V} \\ 0 & \text{otherwise.} \end{cases}$$

- Rate and spatial pdf are

$$\bar{\lambda}_c = \lambda V$$

$$f_c(c) = \begin{cases} \frac{1}{V} & \text{if } c \in \mathbf{V} \\ 0 & \text{otherwise.} \end{cases}$$

GENERAL PPPs

Algorithm Sampling a general PPP

1: Initialize $C_k = [\,]$
2: Generate $m_k^c \sim \mathrm{Po}(\bar{\lambda}_c)$
3: **for** $i = 1$ to m_k^c **do**
4: Generate $c_k^i \sim f_c(\cdot)$
5: Set $C_k = \left[C_k, c_k^i\right]$
6: **end for**

PPP distributions

- For $C_k = \left[c_k^1, \ldots, c_k^{m_k^c}\right]$,

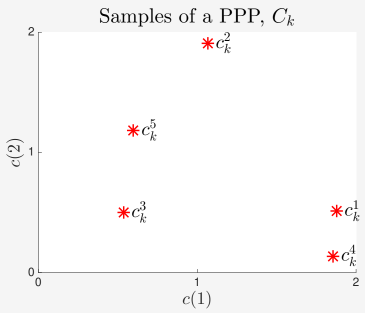
$$
\begin{aligned}
p(C_k) = p(C_k, m_k^c) &= p(m_k^c)p(C_k \middle| m_k^c) \\
&= \mathrm{Po}(m_k^c; \bar{\lambda}_c) \prod_{i=1}^{m_k^c} f_c(c_k^i) \\
&= \frac{\exp(-\bar{\lambda}_c)\bar{\lambda}_c^{m_k^c}}{m_k^c!} \prod_{i=1}^{m_k^c} \frac{\lambda_c(c_k^i)}{\bar{\lambda}_c} \\
&= \frac{\exp(-\bar{\lambda}_c)}{m_k^c!} \prod_{i=1}^{m_k^c} \lambda_c(c_k^i).
\end{aligned}
$$

PROPERTIES AND SAMPLES OF PPPs

Algorithm Sampling a general PPP

1: Initialize $C_k = [\,]$
2: Generate $m_k^c \sim \text{Po}(\bar{\lambda}_c)$
3: **for** $i = 1$ to m_k^c **do**
4: Generate $c_k^i \sim f_c(\cdot)$
5: Set $C_k = \left[C_k, c_k^i\right]$
6: **end for**

Samples of original PPP



Complete measurement model – part 1

Multi-Object Tracking

Lennart Svensson

MEASUREMENT MODEL

Objective

- We seek $p(Z_k \mid x_k)$, needed in the updated step.
- **Note:** $Z_k = \Pi(O_k, C_k)$.

- **Object detections:** the object is detected with probability $P^{\mathrm{D}}(x_k)$, and, if detected, generates $o_k \sim g_k(o_k \mid x_k)$.
- **Clutter detections:** the number of clutter measurements is $m_k^c \sim \mathrm{Po}(\bar{\lambda}_c)$, and the clutter measurements, $c_k^1, \ldots, c_k^{m_k^c}$, are i.i.d., $c_k^i \sim f_c(c_k^i)$.

Main challenges

1) The width of Z_k is random.
2) We do not know which, if any, detection in Z_k that is an object detection.

NOTATION FOR DATA ASSOCIATION

- For brevity, we omit time indexes.

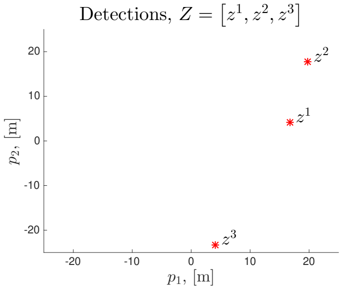
Data associations

- To describe the data association we use

$$\theta = \begin{cases} i > 0 & \text{if } z^i \text{ is an object detection,} \\ 0 & \text{if object is undetected.} \end{cases}$$

Data association example

- Suppose $Z = [z^1, z^2, z^3]$. If $\theta = 2$, then z^2 is an object detection whereas z^1 and z^3 are clutter. If $\theta = 0$, then the object is undetected and z^1, z^2 and z^3 are all clutter detections.



DERIVING THE MEASUREMENT MODEL

- To find $p(Z|x)$, we introduce the variables m and θ

$$
\begin{aligned}
p(Z|x) &= p(Z, m|x) \\
&= \sum_{\theta=0}^{m} p(Z, m, \theta|x) \\
&= \sum_{\theta=0}^{m} p(Z|m, \theta, x)p(\theta, m|x).
\end{aligned}
$$

- As we will see, $p(Z|m, \theta, x)$ and $p(\theta, m|x)$ have simple expressions. This enables us to **express the measurement model**!

- Let us find an expression for $p(Z, m, \theta|x) = p(Z|m, \theta, x)p(\theta, m|x)$!

Complete measurement model – part 2

Multi-Object Tracking

Lennart Svensson

PRIOR PROBABILITIES OF DATA ASSOCIATIONS

- **Objective:** find $p(Z, m, \theta | x) = p(Z | m, \theta, x) p(\theta, m | x)$.
- Expressions differ depending on:

$$\theta = 0, m : \begin{cases} \text{object is not detected,} \\ m \text{ clutter detections,} \end{cases}$$

$$\theta = i > 0, m > 0 : \begin{cases} \text{object is detected,} \\ m - 1 \text{ clutter detections,} \\ \text{object detection is given index } i. \end{cases}$$

- We now present $p(Z | m, \theta, x)$ and $p(\theta, m | x)$ for these two cases.

FINDING $p(Z, m, \theta | x)$, $\theta = 0$

- If $\theta = 0$,

$$p(Z | m, \theta, x) = \prod_{i=1}^{m} f_c(z^i)$$

$$p(\theta, m | x) = (1 - P^{\mathrm{D}}(x))\, \mathrm{Po}(m; \bar{\lambda}_c)$$

which implies that

$$p(Z, m, \theta | x) = (1 - P^{\mathrm{D}}(x)) \mathrm{Po}(m; \bar{\lambda}_c) \prod_{i=1}^{m} f_c(z^i).$$

- Using $\mathrm{Po}(m; \bar{\lambda}) = \frac{\exp(-\bar{\lambda}_c)\bar{\lambda}_c^m}{m!}$ and $f_c(z) = \lambda_c(z)/\bar{\lambda}_c$:

$$\begin{aligned} p(Z, m, \theta | x) &= (1 - P^{\mathrm{D}}(x)) \frac{\exp(-\bar{\lambda}_c)\bar{\lambda}_c^m}{m!} \prod_{i=1}^{m} \frac{\lambda_c(z^i)}{\bar{\lambda}_c} \\ &= (1 - P^{\mathrm{D}}(x)) \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i) \end{aligned}$$

FINDING $p(Z, m, \theta \big| x)$, $\theta = 1, \ldots, m$

- For $\theta = 1, \ldots, m$,

$$p(Z \big| m, \theta, x) = g_k(z^\theta|x) \prod_{\substack{i=1 \\ i \neq \theta}}^{N} f_c(z^i) = g_k(z^\theta|x) \frac{\prod_{i=1}^m f_c(z^i)}{f_c(z^\theta)}$$

$$p(\theta, m \big| x) = P^{\mathrm{D}}(x)\, \mathrm{Po}(m-1; \bar{\lambda}_c) \frac{1}{m}$$

which implies that

$$p(Z, m, \theta \big| x) = P^{\mathrm{D}}(x)\, \mathrm{Po}(m-1; \bar{\lambda}_c) \frac{1}{m} \frac{g_k(z^\theta|x)}{f_c(z^\theta)} \prod_{i=1}^{m} f_c(z^i).$$

- Using $\mathrm{Po}(m-1; \bar{\lambda}) = \frac{\exp(-\bar{\lambda}_c)\bar{\lambda}_c^{m-1}}{(m-1)!}$ and $f_c(z) = \lambda_c(z)/\bar{\lambda}_c$:

$$p(Z, m, \theta \big| x) = P^{\mathrm{D}}(x) \frac{\exp(-\bar{\lambda}_c)\bar{\lambda}_c^{m-1}}{(m-1)!} \frac{1}{m} \frac{\bar{\lambda}_c g_k(z^\theta|x)}{\lambda_c(z^\theta)} \prod_{i=1}^{m} \frac{\lambda_c(z^i)}{\bar{\lambda}_c}$$

$$= P^{\mathrm{D}}(x) \frac{g_k(z^\theta|x)}{\lambda_c(z^\theta)} \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i)$$

Example:

$$\begin{cases} Z = \{z^1,\, z^2\} \\ \theta = m = 2 \end{cases}$$

$$\Rightarrow p(Z \big| m, \theta, x)$$

$$= g_k(z^2 \big| x) f_c(z^1)$$

$$= g_k(z^2 \big| x) \frac{f_c(z^1) f_c(z^2)}{f_c(z^2)}$$

$$= g_k(z^2 \big| x) \frac{\prod_{i=1}^m f_c(z^i)}{f_c(z^\theta)}$$

THE COMPLETE MEASUREMENT MODEL

- Putting these equations together,

$$
\begin{aligned}
p(Z|x) &= \sum_{\theta=0}^{m} p(Z, m, \theta|x) \\
&= \overbrace{(1 - P^{\mathrm{D}}(x)) \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i)}^{\theta=0} + \sum_{\theta=1}^{m} P^{\mathrm{D}}(x) \frac{g_k(z^\theta|x)}{\lambda_c(z^\theta)} \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i) \\
&= \left[(1 - P^{\mathrm{D}}(x)) + P^{\mathrm{D}}(x) \sum_{\theta=1}^{m} \frac{g_k(z^\theta|x)}{\lambda_c(z^\theta)}\right] \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i).
\end{aligned}
$$

LIKELIHOOD VISUALIZATIONS

- We have found that

$$p(Z|x) = \sum_{\theta=0}^{m} p(Z, m, \theta|x) = \left[(1 - P^{\mathrm{D}}(x)) + P^{\mathrm{D}}(x) \sum_{\theta=1}^{m} \frac{g_k(z^\theta|x)}{\lambda_c(z^\theta)}\right] \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i).$$

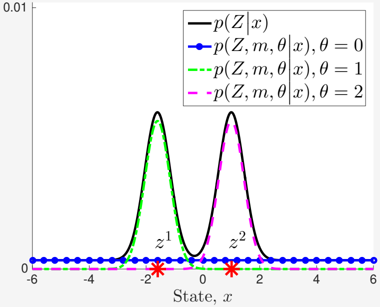
Examples with different P^{D}

- The original example

$$P^{\mathrm{D}}(x) = 0.85, \qquad g_k(o|x) = \mathcal{N}(o; x, 0.2)$$

$$\lambda(c) = \begin{cases} 0.3 & \text{if } |c| \leq 5, \\ 0 & \text{otherwise,} \end{cases} \qquad Z = [-1.6, 1].$$

- Likelihood is dominated by hypotheses $\theta > 0$, for x "near" z^1 or z^2.



LIKELIHOOD VISUALIZATIONS

- We have found that

$$p(Z|x) = \sum_{\theta=0}^{m} p(Z,m,\theta|x) = \left[(1-P^{\mathrm{D}}(x)) + P^{\mathrm{D}}(x)\sum_{\theta=1}^{m}\frac{g_k(z^\theta|x)}{\lambda_c(z^\theta)}\right]\frac{\exp(-\bar{\lambda}_c)}{m!}\prod_{i=1}^{m}\lambda_c(z^i).$$

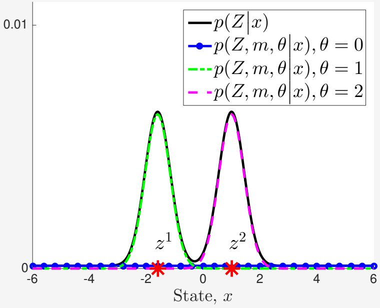
Examples with different P^{D}

- Now, with a larger P^{D}

$$P^{\mathrm{D}}(x) = 0.95, \qquad g_k(o|x) = \mathcal{N}(o; x, 0.2)$$

$$\lambda(c) = \begin{cases} 0.3 & \text{if } |c| \leq 5, \\ 0 & \text{otherwise,} \end{cases} \qquad Z = [-1.6, 1].$$

- The hypothesis $\theta = 0$ contributes even less to the likelihood.



LIKELIHOOD VISUALIZATIONS

- We have found that

$$p(Z|x) = \sum_{\theta=0}^{m} p(Z, m, \theta|x) = \left[(1 - P^{\mathrm{D}}(x)) + P^{\mathrm{D}}(x) \sum_{\theta=1}^{m} \frac{g_k(z^\theta|x)}{\lambda_c(z^\theta)}\right] \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i).$$

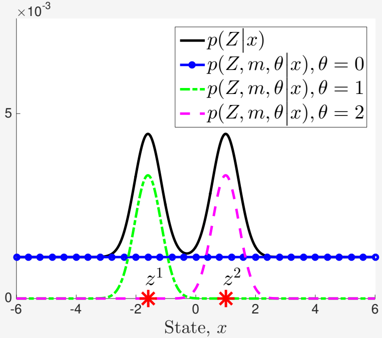
Examples with different P^{D}

- Finally, with a smaller P^{D}:

$$P^{\mathrm{D}}(x) = 0.5, \qquad g_k(o|x) = \mathcal{N}(o; x, 0.2)$$

$$\lambda(c) = \begin{cases} 0.3 & \text{if } |c| \leq 5, \\ 0 & \text{otherwise,} \end{cases} \qquad Z = [-1.6, 1].$$

- Now, $\theta = 0$ is more likely and the likelihood is less informative.



Section 3: SOT, conceptual solution

Multi-Object Tracking

Lennart Svensson

Visualizing the SOT filtering recursions

Multi-Object Tracking

Lennart Svensson

POSTERIOR DENSITIES: BASIC STRUCTURE

- Let the sequences of measurements and data association hypotheses up to time k be denoted

$$Z_{1:k} = (Z_1, Z_2, \ldots, Z_k), \qquad \theta_{1:k} = (\theta_1, \ldots, \theta_k).$$

Structure of posterior density

- In SOT, the filtering density can be written as

$$p(x_k | Z_{1:k}) = \sum_{\theta_{1:k}} p(x_k | Z_{1:k}, \theta_{1:k}) \Pr\left[\theta_{1:k} | Z_{1:k}\right],$$

- Proof using law of total probability

$$p(x_k | Z_{1:k}) = \sum_{\theta_{1:k}} p(x_k, \theta_{1:k} | Z_{1:k}).$$

POSTERIOR DENSITIES: BASIC STRUCTURE

- Let the sequences of measurements and data association hypotheses up to time k be denoted

$$Z_{1:k} = (Z_1, Z_2, \ldots, Z_k)\,, \qquad \theta_{1:k} = (\theta_1, \ldots, \theta_k)\,.$$

Structure of posterior density

- In SOT, the filtering density can be written as

$$p(x_k | Z_{1:k}) = \sum_{\theta_{1:k}} p(x_k | Z_{1:k}, \theta_{1:k}) \Pr\left[\theta_{1:k} \middle| Z_{1:k}\right],$$

whereas the predicted density is

$$p(x_{k+1} | Z_{1:k}) = \sum_{\theta_{1:k}} p(x_{k+1} | Z_{1:k}, \theta_{1:k}) \Pr\left[\theta_{1:k} \middle| Z_{1:k}\right].$$

MODEL ASSUMPTIONS AND OBSERVATIONS

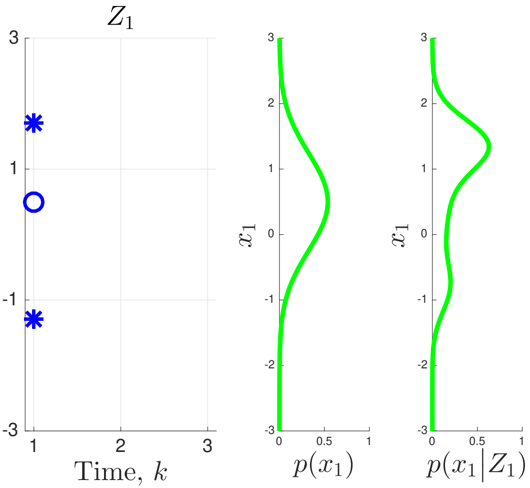
$$
\begin{aligned}
\text{Prior density :} \quad & p(x_1) = \mathcal{N}(x_1; 0.5, 0.2) \\
\text{Motion model :} \quad & \pi_k(x_k \mid x_{k-1}) = \mathcal{N}(x_k; x_{k-1}, 0.35) \\
\text{Probability of detection:} \quad & P^D(x) = 0.9 \\
\text{Object likelihood :} \quad & g_k(o_k \mid x_k) = \mathcal{N}(o_k; x_k, 0.2) \\
\text{Clutter intensity :} \quad & \lambda(c) = \begin{cases} 0.4 & \text{if } |c| \leq 4 \\ 0 & \text{otherwise} \end{cases} \\
\text{Observed detections :} \quad & Z_1 = [-1.3,\ 1.7], \quad Z_2 = [1.3], \\
& Z_3 = [-0.3,\ 2.3].
\end{aligned}
$$

- **Note:** in this example, $p(x_k \mid Z_{1:k}, \theta_{1:k})$ is computed using a Kalman filter (no update if undetected).

A VISUALIZATION OF THE UPDATE STEP, $k = 1$

- Circle corresponds to "object is undetected".
- **Green curves** illustrate $p(x_1)$ and $p(x_1 | Z_1)$, $Z_1 = [-1.3,\ 1.7]$.
- There are $m_1 + 1$ hypotheses.
- **Red dashed curve** is contribution from a single term to the posterior,

$$p(x_1 | Z_1) = \sum_{\theta_1} p(x_1 | \theta_1, Z_1) \Pr(\theta_1 | Z_1).$$



A VISUALIZATION OF THE PREDICTION STEP, $k = 2$

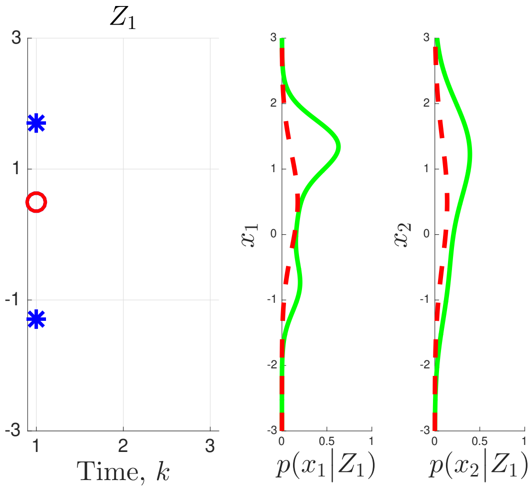
- Green curves illustrate

$$p(x_2|Z_1) = \sum_{\theta_1} p(x_2|Z_1, \theta_1) \Pr[\theta_1|Z_1]$$

$$p(x_1|Z_1) = \sum_{\theta_1} p(x_1|Z_1, \theta_1) \Pr[\theta_1|Z_1],$$

  whereas red dashed curves illustrate individual terms.

- There are $m_1 + 1$ hypotheses.



A VISUALIZATION OF THE UPDATE STEP, $k = 2$

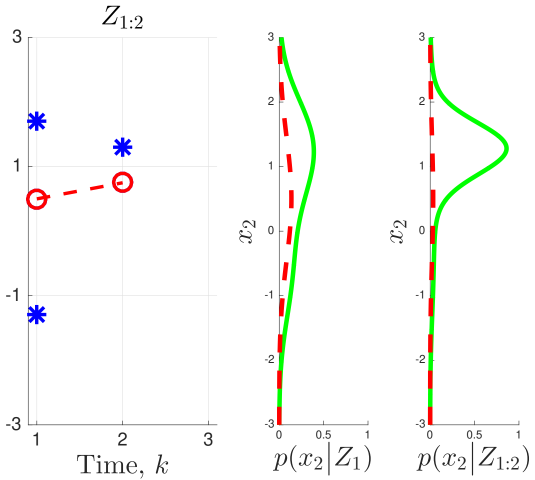
- Green curves illustrate

$$p(x_2|Z_{1:2}) = \sum_{\theta_{1:2}} p(x_2|Z_{1:2}, \theta_{1:2}) \Pr[\theta_{1:2}|Z_{1:2}]$$

$$p(x_2|Z_1) = \sum_{\theta_1} p(x_2|Z_1, \theta_1) \Pr[\theta_1|Z_1],$$

whereas red dashed curves illustrate individual terms.

- There are $(m_1 + 1) \times (m_2 + 1)$ hypotheses.



A VISUALIZATION OF THE PREDICTION STEP, $k = 3$

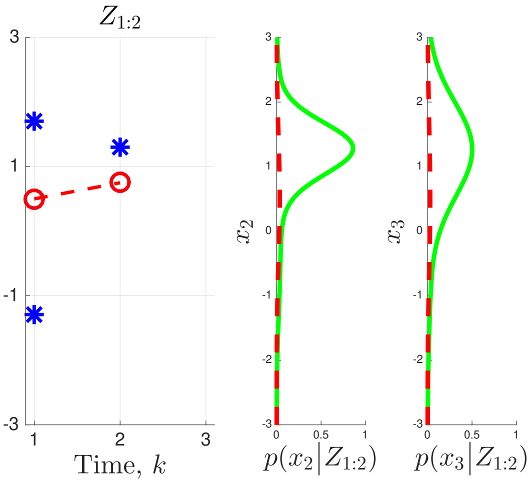
- Green curves illustrate

  $$p(x_2|Z_{1:2}) = \sum_{\theta_{1:2}} p(x_2|Z_{1:2}, \theta_{1:2}) \Pr[\theta_{1:2}|Z_{1:2}]$$

  $$p(x_3|Z_{1:2}) = \sum_{\theta_{1:2}} p(x_3|Z_{1:2}, \theta_{1:2}) \Pr[\theta_{1:2}|Z_{1:2}],$$

  whereas red dashed curves illustrate individual terms.

- There are $(m_1 + 1) \times (m_2 + 1)$ hypotheses.



A VISUALIZATION OF THE UPDATE STEP, $k = 3$

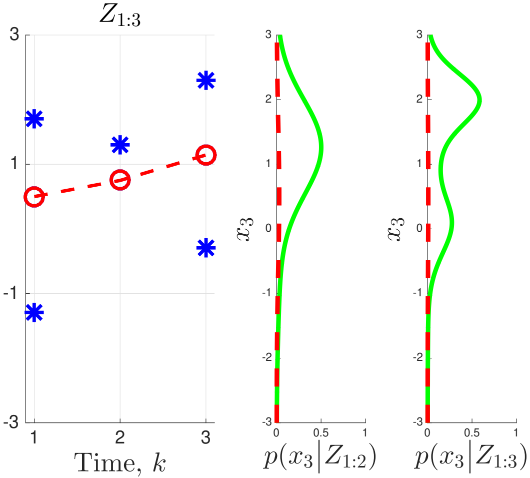
- Green curves illustrate

  $$p(x_3 | Z_{1:2}) = \sum_{\theta_{1:2}} p(x_3 | Z_{1:2}, \theta_{1:2}) \Pr[\theta_{1:2} | Z_{1:2}]$$

  $$p(x_3 | Z_{1:3}) = \sum_{\theta_{1:2}} p(x_3 | Z_{1:3}, \theta_{1:3}) \Pr[\theta_{1:3} | Z_{1:3}],$$

  whereas red dashed curves illustrate individual terms.

- There are $(m_1 + 1) \times (m_2 + 1) \times (m_3 + 1)$ hypotheses.



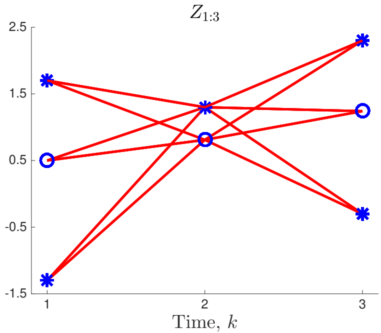
DATA ASSOCIATION HYPOTHESES

- We have $m_k + 1$ data association hypotheses at time k.

- The number of possible association sequences at time k is

$$\prod_{i=1}^{k} (m_i + 1) = (m_1 + 1) \times \cdots \times (m_k + 1),$$

which **grows quickly** with k.



Normalizing the posterior mixture of densities

Multi-Object Tracking

Lennart Svensson

MEASUREMENT UPDATE

- **Measurement likelihood:**

$$p(Z_k \middle| x_k) = \sum_{\theta_k=0}^{m_k} p(Z_k, m_k, \theta_k \middle| x_k).$$

- **Posterior density:**

$$\begin{aligned} p(x \middle| Z) &\propto p(x)p(Z \middle| x) \\ &= \sum_{\theta=0}^{m} p(x)p(Z, m, \theta \middle| x). \end{aligned}$$

- **Desired form:**

$$p(x \middle| Z) = \sum_{\theta=0}^{m} w_\theta p_\theta(x),$$

where

$$\begin{cases} w_\theta & \text{is a pmf } (\Pr[\theta | Z]) \\ p_\theta(x) & \text{is a pdf } (p(x | \theta, Z)). \end{cases}$$

PROBLEM FORMULATION

- Consider a probability density function

$$p(x) \propto g(x) = \sum_{\theta=0}^{m} g_\theta(x),$$

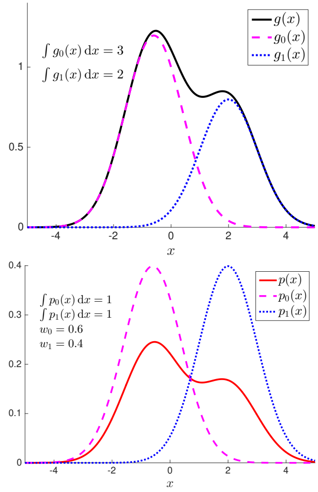
where $g_0(x), \dots, g_m(x)$ are non-negative functions with integrals

$$0 < \int g_i(x)\,\mathrm{d}x < \infty.$$

- How can we express this pdf as a mixture of pdfs

$$p(x) = \sum_{\theta=0}^{m} w_\theta p_\theta(x)?$$

Example:



NORMALIZING A FUNCTION

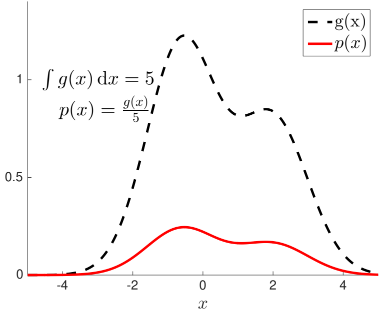
- For $p(x) \propto g(x)$ there is a c:

$$p(x) = c\, g(x).$$

- Given that $p(x)$ is a pdf:

$$1 = \int p(x)\,\mathrm{d}x = c \int g(x)\mathrm{d}x$$
$$\Rightarrow c = \frac{1}{\int g(x)\,\mathrm{d}x}.$$

Example:



Normalizing a density

- If $p(x) \propto g(x)$, then

$$p(x) = \frac{g(x)}{\int g(x')\,\mathrm{d}x'}.$$

FACTORIZING $g_\theta(x)$

- Introducing

$$\begin{cases} \tilde{w}_\theta = \int g_\theta(x)\,\mathrm{d}x \\ p_\theta(x) = \frac{g_\theta(x)}{\tilde{w}_\theta} \end{cases}$$

  we get the factorization

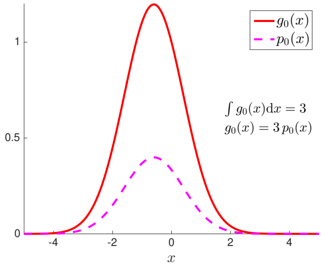
$$g_\theta(x) = \tilde{w}_\theta p_\theta(x)$$

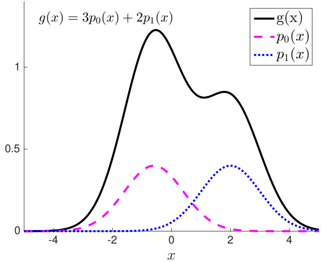
  where $p_\theta(x)$ is a pdf.

- It follows that

$$p(x) \propto g(x) = \sum_{\theta=0}^{m} \tilde{w}_\theta p_\theta(x).$$

Example:





NORMALIZING THE MIXTURE

- We know that

$$\begin{cases} p(x) \propto g(x) \Rightarrow p(x) = \frac{g(x)}{\int g(x')\,\mathrm{d}x'} \\ g(x) = \sum_{\theta=0}^{m} g_\theta(x) = \sum_{\theta=0}^{m} \tilde{w}_\theta p_\theta(x) \end{cases}$$

  where $\tilde{w}_\theta = \int g_\theta(x)\,\mathrm{d}x$ and $p_\theta(x) = \frac{g_\theta(x)}{\tilde{w}_\theta}$.

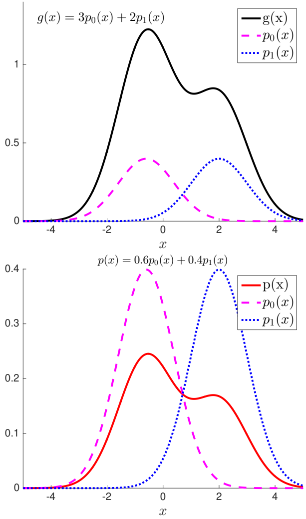
- **Note**: $\int g(x)\,\mathrm{d}x = \sum_{\theta=0}^{m} \int g_\theta(x)\,\mathrm{d}x = \sum_{\theta=0}^{m} \tilde{w}_\theta$.

- Introducing **normalized weights**

$$w_\theta = \frac{\tilde{w}_\theta}{\sum_{i=0}^{m} \tilde{w}_i},$$

$$\Rightarrow p(x) = \frac{\sum_{\theta=0}^{m} \tilde{w}_\theta p_\theta(x)}{\sum_{i=0}^{m} \tilde{w}_i} = \sum_{\theta=0}^{m} w_\theta p_\theta(x).$$

Example:



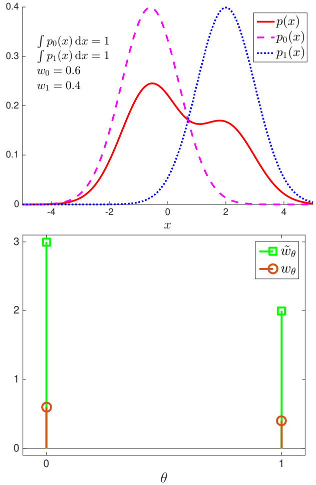
MIXTURES OF DENSITIES: SUMMARY

Normalizing a mixture

- If
$$p(x) \propto \sum_{\theta=0}^{m} g_\theta(x),$$
it follows that
$$p(x) = \sum_{\theta=0}^{m} w_\theta p_\theta(x),$$
where $p_\theta(x) \propto g_\theta(x)$ and $w_\theta \propto \int g_\theta(x)\mathrm{d}x$.
- **Note:** w_θ should be normalized to become a pmf.

- Specifically, we can set
$$\tilde{w}_\theta = \int g_\theta(x)\mathrm{d}x, \quad p_\theta(x) = \frac{g_\theta(x)}{\tilde{w}_\theta}, \quad w_\theta = \frac{\tilde{w}_\theta}{\sum_i \tilde{w}_i}.$$

Example:



Interpretation of weights and densities

Multi-Object Tracking

Lennart Svensson

MEASUREMENT UPDATE

- **Posterior density:**

$$\begin{aligned} p(x|Z) &\propto p(x)p(Z|x) \\ &= \sum_{\theta=0}^{m} \underbrace{p(x)p(Z, m, \theta|x)}_{g_\theta(x)} . \end{aligned}$$

- **Final expression:**

$$p(x|Z) = \sum_{\theta=0}^{m} w_\theta p_\theta(x),$$

where

$$\begin{cases} w_\theta & \text{is a pmf, } \Pr[\theta|Z], \\ p_\theta(x) & \text{is a pdf, } p(x|\theta, Z). \end{cases}$$

INTERPRETATIONS: MAIN RESULTS

- One can show that $$w_\theta p_\theta(x) = p(x, \theta \big| Z).$$

Three important consequences

- Weights are DA probabilities
$$w_\theta = \Pr(\theta \big| Z).$$

- Pdfs are conditional posterior pdfs
$$p_\theta(x) = \frac{p(x, \theta \big| Z)}{w_\theta} = \frac{p(x, \theta \big| Z)}{\Pr(\theta \big| Z)} = p(x \big| \theta, Z).$$

- An expression for posterior pdf
$$p(x \big| Z) = \sum_\theta w_\theta p_\theta(x).$$

A general update equation

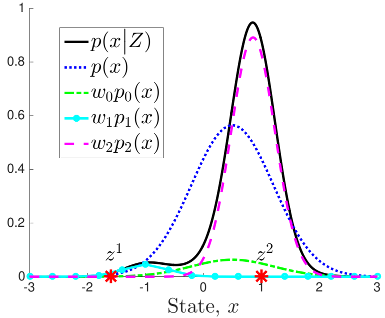
Multi-Object Tracking

Lennart Svensson

UPDATE STEP: AN ILLUSTRATION

- An illustrative example (revisited):

$$
\begin{aligned}
\text{Prior density :} \quad & p(x) = \mathcal{N}(x; 0.5, 0.5), \\
\text{Constant } P^{\mathrm{D}} \text{ :} \quad & P^{\mathrm{D}}(x) = 0.85, \\
\text{Object likelihood :} \quad & g(o|x) = \mathcal{N}(o; x, 0.2) \\
\text{Clutter intensity :} \quad & \lambda_c(c) = \begin{cases} 0.3 & \text{if } |c| < 5 \\ 0 & \text{otherwise,} \end{cases} \\
\text{Observed detections :} \quad & Z = [-1.6, 1].
\end{aligned}
$$



  - **We derive** an expression

$$p(x|Z) = \sum_{\theta=0}^{m} w_\theta p_\theta(x)$$

for **general models** $p(x)$, $P^{\mathrm{D}}(x)$, $\lambda_c(c)$ and $g(o|x)$.

UPDATE STEP (1)

- **Measurement model:**

$$p(Z|x) = \left[(1 - P^{\mathrm{D}}(x)) + P^{\mathrm{D}}(x) \sum_{\theta=1}^{m} \frac{g(z^\theta|x)}{\lambda_c(z^\theta)}\right] \frac{\exp(-\bar{\lambda}_c)}{m!} \prod_{i=1}^{m} \lambda_c(z^i).$$

- **Posterior density:**

$$\begin{aligned} p(x|Z) &\propto p(x)p(Z|x) \\ &\propto p(x) \left[(1 - P^{\mathrm{D}}(x)) + P^{\mathrm{D}}(x) \sum_{\theta=1}^{m} \frac{g(z^\theta|x)}{\lambda_c(z^\theta)}\right]. \end{aligned}$$

UPDATE STEP (2)

- The posterior density is,

$$p(x|Z) \propto p(x)\left[(1 - P^{\mathrm{D}}(x)) + P^{\mathrm{D}}(x)\sum_{\theta=1}^{m}\frac{g(z^{\theta}|x)}{\lambda_c(z^{\theta})}\right].$$

Posterior probabilities and densities

- We get $p(x|Z) = \sum_{\theta=0}^{m} w_\theta p_\theta(x)$, where

$\theta = 0$
Object is undetected
$$: \begin{cases} \tilde{w}_0 = \int p(x)(1 - P^{\mathrm{D}}(x))\,\mathrm{d}x \\ p_0(x) = \frac{p(x)(1-P^{\mathrm{D}}(x))}{\int p(x')(1-P^{\mathrm{D}}(x'))\,\mathrm{d}x'}, \end{cases}$$

$\theta \in \{1, 2, \ldots, m\}$
z^θ is object detection
$$: \begin{cases} \tilde{w}_\theta = \frac{1}{\lambda_c(z^\theta)} \int p(x)P^{\mathrm{D}}(x)g(z^\theta|x)\,\mathrm{d}x \\ p_\theta(x) = \frac{p(x)P^{\mathrm{D}}(x)g(z^\theta|x)}{\int p(x')P^{\mathrm{D}}(x')g(z^\theta|x')\,\mathrm{d}x'}, \end{cases}$$

and $w_\theta \propto \tilde{w}_\theta$.

POSTERIOR DENSITY GIVEN θ

- Recall the object measurement model

$$p(O|x) = \begin{cases} 1 - P^{\mathrm{D}}(x) & \text{if } O = [\,], \\ P^{\mathrm{D}}(x) g(o|x) & \text{if } O = o. \end{cases}$$

- Given O, we get

$$p(x|O) \propto \begin{cases} p(x)(1 - P^{\mathrm{D}}(x)) & \text{if } O = [\,], \\ p(x) P^{\mathrm{D}}(x) g(o|x) & \text{if } O = o. \end{cases}$$

- By comparison,

$$p_\theta(x) \propto \begin{cases} p(x)(1 - P^{\mathrm{D}}(x)) & \text{if } \theta = 0, \\ p(x) P^{\mathrm{D}}(x) g(z^\theta|x) & \text{if } \theta \in \{1, 2, \dots, m\}. \end{cases}$$

Conclusion: $p_\theta(x)$ is identical to $p(x|O)$, with O defined by θ and Z.

Update equations for linear and Gaussian models

Multi-Object Tracking

Lennart Svensson

MODEL ASSUMPTIONS

- Suppose:

$$\begin{aligned} \text{Prior density}: &\quad p(x) = \mathcal{N}(x; \mu, P), \\ \text{Constant } P^{\mathrm{D}}: &\quad P^{\mathrm{D}}(x) = P^{\mathrm{D}}, \\ \text{Object measurement likelihood}: &\quad g(o|x) = \mathcal{N}(o; Hx, R), \\ \text{Clutter intensity}: &\quad \lambda_c(c) \geq 0. \end{aligned}$$

- We express the posterior density on the form

$$p(x|Z) = \sum_{\theta=0}^{m} w_\theta p_\theta(x)$$

and study w_θ and $p_\theta(x)$.

POSTERIOR DENSITY GIVEN θ, (1)

- We found that

$$p_\theta(x) \propto \begin{cases} p(x)(1 - P^{\mathrm{D}}(x)) & \text{if } \theta = 0, \\ p(x)P^{\mathrm{D}}(x)g(z^\theta|x) & \text{if } \theta \in \{1, 2, \ldots, m\}. \end{cases}$$

- When $P^{\mathrm{D}}(x) = P^{\mathrm{D}}$, this simplifies to

$$p_\theta(x) \propto \begin{cases} p(x) & \text{if } \theta = 0, \\ p(x)g(z^\theta|x) & \text{if } \theta \in \{1, 2, \ldots, m\}. \end{cases}$$

- If $\theta > 0$, we update the prior using the likelihood $g(z^\theta|x)$.
  No update if $\theta = 0$.

POSTERIOR DENSITY GIVEN θ, (2)

- When $\theta \in \{1, 2, \ldots, m\}$, assuming $p(x) = \mathcal{N}(x; \mu, P)$ and $g(o|x) = \mathcal{N}(o; Hx, R)$,

$$p_\theta(x) \propto p(x) g(z^\theta | x) = \mathcal{N}(x; \mu, P)\mathcal{N}(z^\theta; Hx, R).$$

- With a Gaussian prior and a linear-Gaussian likelihood, we obtain a Gaussian posterior.
- We can use the **Kalman filter update** to compute the posterior density:

$$\begin{aligned}
\text{Predicted measurement covariance:} \quad & S = HPH^T + R \\
\text{Kalman gain:} \quad & K = PH^T S^{-1} \\
\text{Posterior mean:} \quad & \hat{x}_\theta = \mu + K(z^\theta - H\mu) \\
\text{Posterior covariance:} \quad & P_+ = P - KHP \\
\text{Posterior density:} \quad & p_\theta(x) = \mathcal{N}(x; \hat{x}_\theta, P_+).
\end{aligned}$$

VISUALIZING $p_\theta(x)$

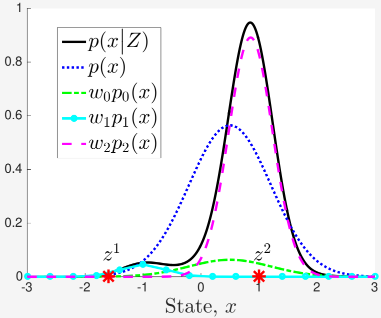
- When $\theta > 0$, $p_\theta(x)$ is obtained by a Kalman filter update of $p(x)$, assuming that z^θ is the object detection.

Example, revisited

- Suppose $p(x) = \mathcal{N}(x; 0.5, 0.5)$, $Z = [-1.6, 1]$, $P^{\mathrm{D}} = 0.85$, $g(o|x) = \mathcal{N}(o; x, 0.2)$ and

$$\lambda_c(c) = \begin{cases} 0.3 & \text{if } |c| < 5 \\ 0 & \text{otherwise.} \end{cases}$$

- $p_1(x)$ and $p_2(x)$ are obtained from a Kalman filter update using z^1 and z^2, respectively.



POSTERIOR PROBABILITIES OF θ, (1)

- We found that

$$\tilde{w}_\theta = \begin{cases} \int p(x)(1 - P^{\mathrm{D}}(x))\,\mathrm{d}x & \text{if } \theta = 0, \\ \frac{1}{\lambda_c(z^\theta)} \int p(x) P^{\mathrm{D}}(x) g(z^\theta|x)\,\mathrm{d}x & \text{if } \theta \in \{1, 2, \ldots, m\}. \end{cases}$$

- When $P^{\mathrm{D}}(x) = P^{\mathrm{D}}$, this simplifies to

$$\tilde{w}_\theta = \begin{cases} 1 - P^{\mathrm{D}} & \text{if } \theta = 0, \\ \frac{P^{\mathrm{D}}}{\lambda_c(z^\theta)} \int p(x) g(z^\theta|x)\,\mathrm{d}x & \text{if } \theta \in \{1, 2, \ldots, m\}. \end{cases}$$

- Here, $\int p(x) g(z^\theta|x)\,\mathrm{d}x$ is the predicted density for the object measurement, evaluated at z^θ.

POSTERIOR PROBABILITIES OF θ, (2)

- When $\theta \in \{1, 2, \ldots, m\}$, assuming $p(x) = \mathcal{N}(x; \mu, P)$ and $g(o \big| x) = \mathcal{N}(o; Hx, R)$,

$$\tilde{w}_\theta = \frac{P^{\mathrm{D}}}{\lambda_c(z^\theta)} \int \mathcal{N}(x; \mu, P)\mathcal{N}(z^\theta; Hx, R)\, \mathrm{d}x.$$

- Specifically, $\displaystyle\int \underbrace{\mathcal{N}(x; \mu, P)}_{p(x)} \underbrace{\mathcal{N}(z^\theta; Hx, R)}_{p(z^\theta | x, \theta)} \mathrm{d}x = \mathcal{N}(z^\theta; H\mu, HPH^T + R),$

  where we often use the Kalman filter notation: $\bar{z} = H\mu$ and $S = HPH^T + R$.

  Note: this is the density of $z^\theta = Hx + v$ where $x \sim \mathcal{N}(\mu, P)$, and $v \sim \mathcal{N}(0, R)$.

- We conclude that

$$\tilde{w}_\theta = \frac{P^{\mathrm{D}} \mathcal{N}(z^\theta; \bar{z}, S)}{\lambda_c(z^\theta)},$$

  where $\mathcal{N}(z^\theta; \bar{z}, S)$ is called the **predicted likelihood**.

VISUALIZING w_θ

- We conclude that

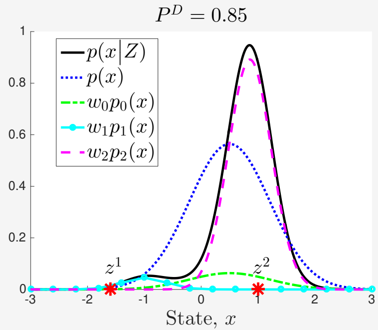
$$\tilde{w}_\theta = \begin{cases} 1 - P^{\mathrm{D}} & \text{if } \theta = 0, \\ \frac{P^{\mathrm{D}} \mathcal{N}(z^\theta; \bar{z}, S)}{\lambda_c(z^\theta)} & \text{if } \theta \in \{1, 2, \ldots, m\}. \end{cases}$$

Examples, revisited, with different P^{D}

- We assume $p(x) = \mathcal{N}(x; 0.5, 0.5)$ and $g(o|x) = \mathcal{N}(o; x, 0.2)$

$$\Rightarrow \bar{z} = 0.5, \qquad S = 0.5 + 0.2 = 0.7.$$

- $\tilde{w}_2 > \tilde{w}_1$ since z^2 is closer than z^1 to $\bar{z}$.



VISUALIZING w_θ

- We conclude that

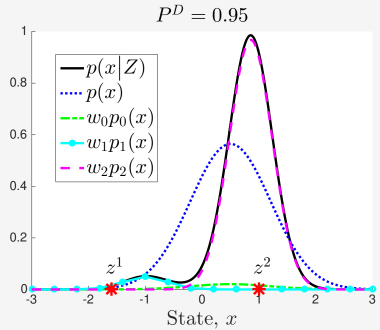
$$\tilde{w}_\theta = \begin{cases} 1 - P^{\mathrm{D}} & \text{if } \theta = 0, \\ \frac{P^{\mathrm{D}} \mathcal{N}(z^\theta; \bar{z}, S)}{\lambda_c(z^\theta)} & \text{if } \theta \in \{1, 2, \ldots, m\}. \end{cases}$$

Examples, revisited, with different P^{D}

- We assume $p(x) = \mathcal{N}(x; 0.5, 0.5)$ and $g(o|x) = \mathcal{N}(o; x, 0.2)$

$$\Rightarrow \bar{z} = 0.5, \qquad S = 0.5 + 0.2 = 0.7.$$

- $\tilde{w}_2 > \tilde{w}_1$ since z^2 is closer than z^1 to $\bar{z}$.
- w_0 decreases with P^{D}.



VISUALIZING w_θ

- We conclude that

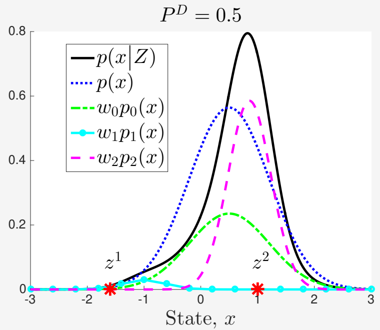
$$\tilde{w}_\theta = \begin{cases} 1 - P^{\mathrm{D}} & \text{if } \theta = 0, \\ \frac{P^{\mathrm{D}} \mathcal{N}(z^\theta; \bar{z}, S)}{\lambda_c(z^\theta)} & \text{if } \theta \in \{1, 2, \ldots, m\}. \end{cases}$$

Examples, revisited, with different P^{D}

- We assume $p(x) = \mathcal{N}(x; 0.5, 0.5)$ and $g(o|x) = \mathcal{N}(o; x, 0.2)$

$$\Rightarrow \bar{z} = 0.5, \qquad S = 0.5 + 0.2 = 0.7.$$

- $\tilde{w}_2 > \tilde{w}_1$ since z^2 is closer than z^1 to $\bar{z}$.
- w_0 decreases with P^{D}.



CONCLUSIONS

Closed form expressions

- If $p(x) = \mathcal{N}(x; \mu, P)$, $P^{\mathrm{D}}(x) = P^{\mathrm{D}}$ and $g(o|x) = \mathcal{N}(o; Hx, R)$:

$$p_\theta(x) = \begin{cases} p(x) & \text{if } \theta = 0, \\ \mathcal{N}(x; \hat{x}_\theta, P_+) & \text{if } \theta \in \{1, 2, \dots, m\}, \end{cases}$$

$$\tilde{w}_\theta = \begin{cases} 1 - P^{\mathrm{D}} & \text{if } \theta = 0, \\ \frac{P^{\mathrm{D}} \mathcal{N}(z^\theta; \bar{z}, S)}{\lambda_c(z^\theta)} & \text{if } \theta \in \{1, 2, \dots, m\}. \end{cases}$$

- **Remark:** Suppose

$$o = h(x) + v, \qquad v \sim \mathcal{N}(0, R),$$

  such that $g(o|x) = \mathcal{N}(o; h(x), R)$, where $h(x)$ is a nonlinear function.
- We can then **approximate** $\hat{x}_\theta$, P_+, $\bar{z}$ and S **using, e.g., an extended Kalman filter**.

Prediction and update steps: conceptual solution, part 1

Multi-Object Tracking

Lennart Svensson

OVERVIEW OF RESULTS

Main result

- Suppose
$$p(x_{k-1}|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p_{k-1|k-1}^{\theta_{1:k-1}}(x_{k-1})$$
where $w^{\theta_{1:k-1}} = \Pr[\theta_{1:k-1}|Z_{1:k-1}]$ and $p_{k-1|k-1}^{\theta_{1:k-1}}(x_{k-1}) = p(x_{k-1}|\theta_{1:k-1}, Z_{1:k-1})$.
- We can then express the predicted and updated densities as

Predicted density $$p(x_k|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p_{k|k-1}^{\theta_{1:k-1}}(x_k),$$

Updated density $$p(x_k|Z_{1:k}) = \sum_{\theta_{1:k}} w^{\theta_{1:k}} p_{k|k}^{\theta_{1:k}}(x_k).$$

1) Here $\theta_{1:k} = [\theta_1, \ldots, \theta_k]$ is a sequence of data association hypotheses.

OVERVIEW OF RESULTS

Main result

- Suppose
$$p(x_{k-1}|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p^{\theta_{1:k-1}}_{k-1|k-1}(x_{k-1})$$
where $w^{\theta_{1:k-1}} = \Pr[\theta_{1:k-1}|Z_{1:k-1}]$ and $p^{\theta_{1:k-1}}_{k-1|k-1}(x_{k-1}) = p(x_{k-1}|\theta_{1:k-1}, Z_{1:k-1})$.
- We can then express the predicted and updated densities as
$$p(x_k|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p^{\theta_{1:k-1}}_{k|k-1}(x_k), \qquad p(x_k|Z_{1:k}) = \sum_{\theta_{1:k}} w^{\theta_{1:k}} p^{\theta_{1:k}}_{k|k}(x_k).$$

2) The posterior at time k is written on the same form, but contains more terms
$$\sum_{\theta_{1:k}} = \sum_{\theta_1=0}^{m_1} \sum_{\theta_2=0}^{m_2} \cdots \sum_{\theta_k=0}^{m_k}.$$

OVERVIEW OF RESULTS

Main result

- Suppose
$$p(x_{k-1}|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p_{k-1|k-1}^{\theta_{1:k-1}}(x_{k-1})$$
where $w^{\theta_{1:k-1}} = \Pr[\theta_{1:k-1}|Z_{1:k-1}]$ and $p_{k-1|k-1}^{\theta_{1:k-1}}(x_{k-1}) = p(x_{k-1}|\theta_{1:k-1}, Z_{1:k-1})$.
- We can then express the predicted and updated densities as
$$p(x_k|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p_{k|k-1}^{\theta_{1:k-1}}(x_k), \qquad p(x_k|Z_{1:k}) = \sum_{\theta_{1:k}} w^{\theta_{1:k}} p_{k|k}^{\theta_{1:k}}(x_k).$$

3) We know how to compute $p(x_1|Z_1)$ on the above form.

We obtain a recursive algorithm to compute $p(x_k|Z_{1:k})$ for any k.

PREDICTION STEP

- If

$$p(x_{k-1}|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p_{k-1|k-1}^{\theta_{1:k-1}}(x_{k-1})$$

  then

$$\begin{aligned} p(x_k|Z_{1:k-1}) &= \int p(x_{k-1}|Z_{1:k-1}) p(x_k|x_{k-1}) \, \mathrm{d}x_{k-1} \\ &= \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} \underbrace{\int p_{k-1|k-1}^{\theta_{1:k-1}}(x_{k-1}) p(x_k|x_{k-1}) \, \mathrm{d}x_{k-1}}_{\triangleq p_{k|k-1}^{\theta_{1:k-1}}(x_k)} \\ &= \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p_{k|k-1}^{\theta_{1:k-1}}(x_k). \end{aligned}$$

Prediction step

- Weights are unchanged, standard prediction of densities for each hypothesis.

PREDICTION STEP: LINEAR AND GAUSSIAN MOTION

- Suppose
$$x_k = F_{k-1} x_{k-1} + q_{k-1}, \quad q_{k-1} \sim \mathcal{N}(0, Q_{k-1}),$$
such that $\pi_k(x_k \big| x_{k-1}) = \mathcal{N}(x_k; F_{k-1} x_{k-1}, Q_{k-1})$.
- If $p^{\theta_{1:k-1}}_{k-1|k-1}(x_{k-1}) = \mathcal{N}(x_{k-1}, \hat{x}^{\theta_{1:k-1}}_{k-1|k-1}, P^{\theta_{1:k-1}}_{k-1|k-1})$, then
$$p^{\theta_{1:k-1}}_{k|k-1}(x_k) = \mathcal{N}(x_k; F_{k-1} \hat{x}^{\theta_{1:k-1}}_{k-1|k-1}, F_{k-1} P^{\theta_{1:k-1}}_{k-1|k-1} F^T_{k-1} + Q_{k-1}).$$

- **Remark:** Suppose
$$x_k = f_{k-1}(x_{k-1}) + q_{k-1}, \qquad q_{k-1} \sim \mathcal{N}(0, Q_{k-1}),$$
such that $\pi_k(x_k \big| x_{k-1}) = \mathcal{N}(x_k; f_{k-1}(x_{k-1}), R)$, where $f_{k-1}(x_{k-1})$ is a nonlinear function.
- We can then **approximate $p^{\theta_{1:k-1}}_{k|k-1}(x_k)$ using, e.g., an extended Kalman filter**.

MODEL ASSUMPTIONS FOR VISUALIZATION

$$
\begin{aligned}
\text{Prior density :} &\quad p(x_1) = \mathcal{N}(x_1; 0.5, 0.2) \\
\text{Motion model :} &\quad \pi_k(x_k \,|\, x_{k-1}) = \mathcal{N}(x_k; x_{k-1}, 0.35) \\
\text{Probability of detection:} &\quad P^D(x) = 0.9 \\
\text{Object likelihood :} &\quad g_k(o_k \,|\, x_k) = \mathcal{N}(o_k; x_k, 0.2) \\
\text{Clutter intensity :} &\quad \lambda(c) = \begin{cases} 0.4 & \text{if } |c| \leq 4 \\ 0 & \text{otherwise} \end{cases} \\
\text{Observed detections :} &\quad Z_1 = [-1.3,\ 1.7], \quad Z_2 = [1.3].
\end{aligned}
$$

A VISUALIZATION OF THE PREDICTION STEP, $k = 2$

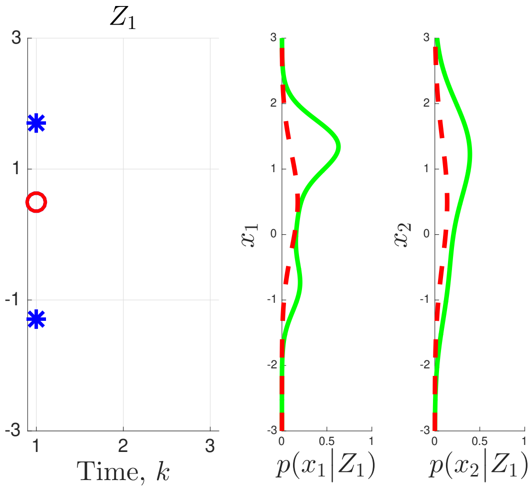
- Green curves illustrate

$$p(x_2|Z_1) = \sum_{\theta_1} w^{\theta_1} p^{\theta_1}_{2|1}(x_2)$$

$$p(x_1|Z_1) = \sum_{\theta_1} w^{\theta_1} p^{\theta_1}_{1|1}(x_1),$$

  whereas red dashed curves illustrate individual terms.

- There are $m_1 + 1$ hypotheses.



Prediction and update steps: conceptual solution, part 2

Multi-Object Tracking

Lennart Svensson

UPDATE STEP (1)

- **Measurement model**

$$p(Z_k|x_k) = \left[(1 - P^{\mathrm{D}}(x_k)) + P^{\mathrm{D}}(x_k) \sum_{\theta_k=1}^{m_k} \frac{g_k(z_k^{\theta_k}|x_k)}{\lambda_c(z_k^{\theta_k})}\right] \frac{\exp(-\bar{\lambda}_c)}{m_k!} \prod_{i=1}^{m} \lambda_c(z_k^i).$$

- For

$$p(x_k|Z_{1:k-1}) = \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} p_{k|k-1}^{\theta_{1:k-1}}(x_k),$$

this implies that

$$p(x_k|Z_{1:k}) \propto \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}}\, p_{k|k-1}^{\theta_{1:k-1}}(x_k)(1 - P^D(x_k)) + \sum_{\theta_{1:k-1}} \sum_{\theta_k=1}^{m_k} \frac{1}{\lambda_c(z_k^{\theta_k})} w^{\theta_{1:k-1}} p_{k|k-1}^{\theta_{1:k-1}}(x_k) P^D(x_k) g_k(z_k^{\theta_k}|x_k).$$

- **Note:** for every pair of hypotheses, $(\theta_{1:k-1}, \theta_k)$, we obtain a new hypothesis. We index this hypothesis using the vector $\theta_{1:k}$.

UPDATE STEP (2)

- The posterior density is,

$$p(x_k|Z_{1:k}) \propto \sum_{\theta_{1:k-1}} w^{\theta_{1:k-1}} \, p_{k|k-1}^{\theta_{1:k-1}}(x_k)(1 - P^D(x_k)) \\ + \sum_{\theta_{1:k-1}} \sum_{\theta_k=1}^{m_k} \frac{1}{\lambda_c(z_k^{\theta_k})} w^{\theta_{1:k-1}} p_{k|k-1}^{\theta_{1:k-1}}(x_k) P^D(x_k) g_k(z_k^{\theta_k}|x_k).$$

Posterior probabilities and densities

- We get $p(x_k|Z_{1:k}) = \sum_{\theta_{1:k}} w^{\theta_{1:k}} p_{k|k}^{\theta_{1:k}}(x_k)$, where $w^{\theta_{1:k}} \propto \tilde{w}^{\theta_{1:k}}$ and

$\theta_k = 0$
Object is undetected

$$: \begin{cases} \tilde{w}^{\theta_{1:k}} = w^{\theta_{1:k-1}} \int p_{k|k-1}^{\theta_{1:k-1}}(x_k)(1 - P^D(x_k)) \, \mathrm{d}x_k \\ p_{k|k}^{\theta_{1:k}}(x_k) \propto p_{k|k-1}^{\theta_{1:k-1}}(x_k)(1 - P^D(x_k)), \end{cases}$$

$\theta_k \in \{1, 2, \ldots, m_k\}$
z_k^{θ} is object detection

$$: \begin{cases} \tilde{w}^{\theta_{1:k}} = \frac{w^{\theta_{1:k-1}}}{\lambda_c(z_k^{\theta_k})} \int p_{k|k-1}^{\theta_{1:k-1}}(x_k) P^D(x_k) g_k(z_k^{\theta_k}|x_k) \, \mathrm{d}x_k \\ p_{k|k}^{\theta_{1:k}}(x_k) \propto p_{k|k-1}^{\theta_{1:k-1}}(x_k) P^D(x_k) g_k(z_k^{\theta_k}|x_k). \end{cases}$$

A LINEAR AND GAUSSIAN UPDATE STEP

Closed form expressions

- If $p_{k|k-1}^{\theta_{1:k-1}}(x_k) = \mathcal{N}(x_k; \mu_{k|k-1}^{\theta_{1:k-1}}, P_{k|k-1}^{\theta_{1:k-1}})$, $P^{\mathrm{D}}(x_k) = P^{\mathrm{D}}$ and $g_k(o_k \middle| x_k) = \mathcal{N}(o_k; H_k x_k, R_k)$:

$$p_{k|k}^{\theta_{1:k}}(x_k) = \begin{cases} p_{k|k-1}^{\theta_{1:k-1}}(x_k) & \text{if } \theta_k = 0, \\ \mathcal{N}(x_k; \mu_{k|k}^{\theta_{1:k}}, P_{k|k}^{\theta_{1:k}}) & \text{if } \theta_k \in \{1, 2, \dots, m\}, \end{cases}$$

$$\tilde{w}^{\theta_{1:k}} = \begin{cases} w^{\theta_{1:k-1}}(1 - P^{\mathrm{D}}) & \text{if } \theta_k = 0, \\ w^{\theta_{1:k-1}} \frac{P^{\mathrm{D}} \mathcal{N}(z_k^{\theta}; \bar{z}_{k|k-1}^{\theta_{1:k-1}}, S_{k|k-1}^{\theta_{1:k-1}})}{\lambda_c(z_k^{\theta})} & \text{if } \theta_k \in \{1, 2, \dots, m\}. \end{cases}$$

- Here $\mu_{k|k}^{\theta_{1:k}}$ and $P_{k|k}^{\theta_{1:k}}$ are the posterior mean and covariance given $Z_{1:k}$ and $\theta_{1:k}$.
- Similarly, $\bar{z}_{k|k-1}^{\theta_{1:k-1}}$ and $S_{k|k-1}^{\theta_{1:k-1}}$ are the predicted object measurement mean and covariance assuming the predicted density $p_{k|k-1}^{\theta_{1:k-1}}(x_k)$.

THE KALMAN FILTER UPDATE

$$
\begin{aligned}
\text{Object measurement prediction:} \quad & \bar{z}_{k|k-1}^{\theta_{1:k-1}} = H_k \mu_{k|k-1}^{\theta_{1:k-1}} \\
\text{Predicted measurement covariance:} \quad & S_{k|k-1}^{\theta_{1:k-1}} = H_k P_{k|k-1}^{\theta_{1:k-1}} H_k^T + R_k \\
\text{Kalman gain:} \quad & K_k^{\theta_{1:k}} = P_{k|k-1}^{\theta_{1:k-1}} H_k^T (S_{k|k-1}^{\theta_{1:k-1}})^{-1} \\
\text{Posterior mean:} \quad & \mu_{k|k}^{\theta_{1:k}} = \mu_{k|k-1}^{\theta_{1:k-1}} + K_k^{\theta_{1:k}} (z_k^{\theta_k} - \bar{z}_{k|k-1}^{\theta_{1:k-1}}) \\
\text{Posterior covariance:} \quad & P_{k|k}^{\theta_{1:k}} = P_{k|k-1}^{\theta_{1:k-1}} - K_k^{\theta_{1:k}} H_k P_{k|k-1}^{\theta_{1:k-1}}.
\end{aligned}
$$

A VISUALIZATION OF THE UPDATE STEP, $k = 2$

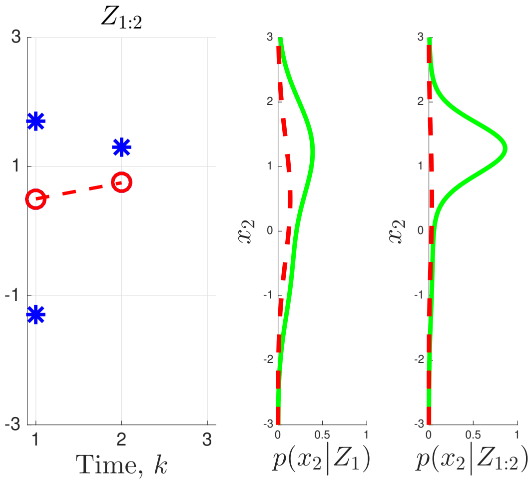
- Green curves illustrate

$$p(x_2 | Z_{1:2}) = \sum_{\theta_{1:2}} w^{\theta_{1:2}} p_{2|2}^{\theta_{1:2}}(x_2)$$

$$p(x_2 | Z_1) = \sum_{\theta_1} w^{\theta_1} p_{2|1}^{\theta_1}(x_2),$$

  whereas red dashed curves illustrate individual terms.

- There are $(m_1 + 1) \times (m_2 + 1)$ hypotheses.



CONCLUDING REMARKS

- We have presented a recursive algorithm to compute

$$p(x_k \big| Z_{1:k}) = \sum_{\theta_{1:k}} w^{\theta_{1:k}} p_{k|k}^{\theta_{1:k}}(x_k).$$

- The posterior contains one term for **every possible sequence** of data associations.
- In linear and Gaussian settings (with constant P^{D}) the densities are computed **using a Kalman filter**.
- The number of hypotheses grows quickly

$$\prod_{i=1}^{k}(m_i + 1)$$

  which means that we **need to introduce approximations**.

Section 4: SOT algorithms

Multi-Object Tracking

Lennart Svensson

An overview of different SOT algorithms

Multi-Object Tracking

Lennart Svensson

THE NEED FOR APPROXIMATIONS

- The number of hypotheses grows as

$$\prod_{i=1}^{k}(m_i + 1).$$

- It is therefore generally intractable to compute

$$p(x_k | Z_{1:k})$$

  exactly, except for a small number of time steps.
- To obtain a feasible algorithm, we **need to introduce approximations**.
- We focus specifically on the **Gaussian mixture setting**, though principles apply more generally.

GAUSSIAN MIXTURE REDUCTION

- **Problem:** $p(x_k|Z_{1:k})$ is a Gaussian mixture with too many components.
- **Standard solution:** Find
$$\hat{p}(x_k|Z_{1:k}) \approx p(x_k|Z_{1:k})$$
where $\hat{p}(x_k|Z_{1:k})$ is a Gaussian mixture with fewer components.
- Once we have selected $\hat{p}(x_k|Z_{1:k})$, we start the next recursion assuming
$$p(x_k|Z_{1:k}) = \hat{p}(x_k|Z_{1:k}).$$
- By limiting the number of components in $\hat{p}(x_k|Z_{1:k})$ we obtain a feasible algorithm.

PRUNING AND MERGING

- The main techniques for mixture reduction are **pruning and merging**.

Pruning

- Remove hypotheses with small weights (and renormalize).

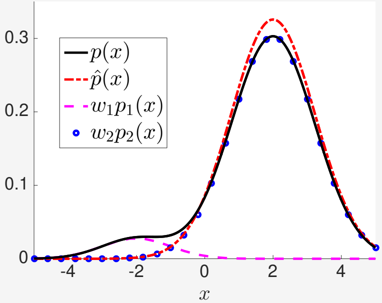
A pruning example

- Suppose $p(x)$ is given by

$$p(x) = w_1 p_1(x) + w_2 p_2(x) \text{ where}$$

$$\begin{cases} w_1 = 0.07, & p_1(x) = \mathcal{N}(x; -2, 1) \\ w_2 = 0.93, & p_2(x) = \mathcal{N}(x; 2, 1.5) \end{cases}$$

- Pruning first hypothesis gives $\hat{p}(x) = p_2(x)$.



PRUNING AND MERGING

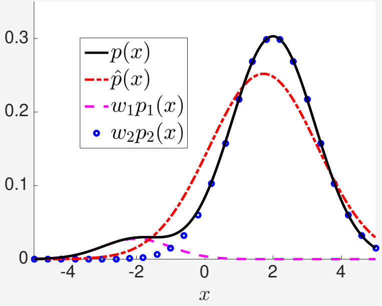
- The main techniques for mixture reduction are **pruning and merging**.

Merging

- Approximate a mixture of densities by a single density (often Gaussian).

A merging example

- Consider again $p(x) = w_1 p_1(x) + w_2 p_2(x)$ as above.
- We can select $\hat{p}(x)$ to match the first two moments of $p(x)$.
- Approximation also depends on w_1 and $p_1(x)$.



PRESENTED ALGORITHMS

- In the next videos, we present three algorithms for SOT in clutter:
  - Nearest neighbour (NN) filter, [**pruning**],
  - Probabilistic data association (PDA) filter [**merging**],
  - Gaussian sum filter (GSF). [**pruning**/**merging**]

- All of these are examples of **assumed density filters**
  - NN and PDA: Gaussian densities,
  - GSF: Gaussian mixture densites,

  that is, every recursion starts and ends with a density in that family.

- Apart from the above tracking algorithms, we also present **gating**, which is a technique to disregard unreasonable detections. [**pruning**]

Nearest neighbour filtering

Multi-Object Tracking

Lennart Svensson

LINEAR AND GAUSSIAN MODELS, PREDICTION STEP

- NN and PDA both assume **Gaussian posterior at time $k-1$**:
$$p(x_{k-1}|Z_{1:k-1}) = \mathcal{N}(x_{k-1}; \bar{x}_{k-1|k-1}, P_{k-1|k-1}).$$

- We also assume a **linear and Gaussian motion model**:
$$x_k = F_{k-1}x_{k-1} + q_{k-1}, \qquad q_{k-1} \sim \mathcal{N}(0, Q).$$

- **Predicted density** is therefore
$$p(x_k|Z_{1:k-1}) = \mathcal{N}(x_k; \bar{x}_{k|k-1}, P_{k|k-1})$$
where $\bar{x}_{k|k-1} = F_{k-1}\bar{x}_{k-1|k-1}$ and $P_{k|k-1} = F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q$.

- We sometimes use superscript NN, e.g., $p^{\mathrm{NN}}(x_k|Z_{1:k-1})$, to clarify that it is an approximation obtained using the NN algorithm.

LINEAR AND GAUSSIAN MODELS, UPDATE STEP

Measurement model

We assume $P^{\mathrm{D}}(x) = P^{\mathrm{D}}, \quad g_k(o|x) = \mathcal{N}(o; H_k x, R_k), \quad$ general $\lambda_c(c)$.

Posterior density, given $p^{\mathrm{NN}}(x_k | Z_{1:k-1}) = \mathcal{N}(x_k; \bar{x}^{\mathrm{NN}}_{k|k-1}, P^{\mathrm{NN}}_{k|k-1})$

Posterior is $\breve{p}^{\mathrm{NN}}(x_k | Z_{1:k}) = \sum_{\theta_k=0}^{m_k} \breve{w}_k^{\theta_k} p_{k|k}^{\theta_k}(x_k)$ where $p_{k|k}^{\theta_k}(x_k) = \mathcal{N}(x_k; \hat{x}_k^{\theta_k}, P_k^{\theta_k})$

and

$$\theta_k = 0 : \quad \begin{cases} \tilde{w}_k^{\theta_k} = 1 - P^{\mathrm{D}}, \quad \hat{x}_k^{\theta_k} = \bar{x}^{\mathrm{NN}}_{k|k-1}, \\ P_k^{\theta_k} = P^{\mathrm{NN}}_{k|k-1}, \end{cases}$$

$$\theta_k \in \{1, \ldots, m_k\} : \quad \begin{cases} \tilde{w}_k^{\theta_k} = \frac{P^{\mathrm{D}} \mathcal{N}(z_k^{\theta_k}; \bar{z}_{k|k-1}, S_k)}{\lambda_c(z_k^{\theta_k})}, \quad \hat{x}_k^{\theta_k} = \bar{x}^{\mathrm{NN}}_{k|k-1} + K_k(z_k^{\theta_k} - \bar{z}_{k|k-1}), \\ P_k^{\theta_k} = P^{\mathrm{NN}}_{k|k-1} - K_k H_k P^{\mathrm{NN}}_{k|k-1}. \end{cases}$$

- How can we **approximate** $\breve{p}^{\mathrm{NN}}(x_k | Z_{1:k})$ as Gaussian?

NEAREST NEIGHBOUR FILTERING

Basic idea

- Prune all hypotheses except the most probable one.

Algorithm The NN filtering update.

1: Compute $\tilde{w}_k^{\theta_k}$, $\theta_k = 0, 1 \ldots, m_k$.

2: Find

$$\theta_k^\star = \arg\max_\theta \tilde{w}_k^\theta.$$

3: Compute $\hat{x}_k^{\theta_k^\star}$ and $P_k^{\theta_k^\star}$.

4: Set $\bar{x}_{k|k}^{\text{NN}} = \hat{x}_k^{\theta_k^\star}$ and $P_{k|k}^{\text{NN}} = P_k^{\theta_k^\star}$.

- **Note:** we then assume that $p^{\text{NN}}(x_k \big| Z_{1:k}) = \mathcal{N}(x_k; \bar{x}_{k|k}^{\text{NN}}, P_{k|k}^{\text{NN}})$.

EXAMPLE FOR VISUALIZATION (SIMPLE)

Prior density : $p(x_1) = \mathcal{N}(x_1; 0.5, 0.2)$

Motion model : $\pi_k(x_k \middle| x_{k-1}) = \mathcal{N}(x_k; x_{k-1}, 0.35)$

Probability of detection: $P^D(x) = 0.9$

Object likelihood : $g_k(o_k \middle| x_k) = \mathcal{N}(o_k; x_k, 0.2)$

Clutter intensity : $\lambda(c) = \begin{cases} 0.4 & \text{if } |c| \leq 4 \\ 0 & \text{otherwise} \end{cases}$

Observed detections : $Z_1 = [-1.3,\ 1.7], \quad Z_2 = [1.3],$

$Z_3 = [-0.3,\ 2.3], \quad Z_4 = [-2,\ 3]$

$Z_5 = [2.6], \quad Z_6 = [-3.5,\ 2.8]$

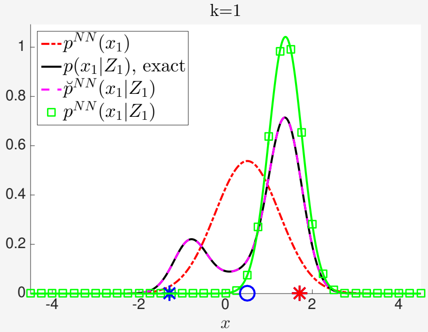
A NEAREST NEIGHBOUR FILTERING EXAMPLE

Example: NN vs exact posterior

- We compare four densities:

$$p^{\text{NN}}(x_k|Z_{1:k-1}), \quad p(x_k|Z_{1:k})$$
$$\breve{p}^{\text{NN}}(x_k|Z_{1:k}), \quad p^{\text{NN}}(x_k|Z_{1:k})$$

- We visualize the $m_k + 1$ hypotheses, and mark the most probable in red.
- The NN algorithm approximates the posterior fairly well.



Nearest neighbour filtering – additional remarks

Multi-Object Tracking

Lennart Svensson

WHY THE NAME "NEAREST NEIGHBOUR"?

- To find $\theta_k^\star$, we can do

$$\theta_k^+ = \operatorname*{arg\,max}_{\theta \in \{1,2,\ldots,m_k\}} \tilde{w}_k^{\theta_k},$$

$$\theta_k^\star = \begin{cases} \theta_k^+ & \text{if } \tilde{w}_k^{\theta_k^+} \geq \tilde{w}_k^0 \\ 0 & \text{if } \tilde{w}_k^{\theta_k^+} < \tilde{w}_k^0. \end{cases}$$

- Roughly speaking, $z_k^{\theta_k^+}$ is the **"nearest" measurement** to $\bar{z}_{k|k-1}$.

NEAREST NEIGHBOUR?

$$
\begin{aligned}
\theta_k^+ &= \underset{\theta \in \{1,2,\ldots,m_k\}}{\arg\max} \; \tilde{w}_k^{\theta_k} \\
&= \underset{\theta \in \{1,2,\ldots,m_k\}}{\arg\max} \; \frac{P^{\mathrm{D}} \mathcal{N}(z_k^\theta; \bar{z}_{k|k-1}, S_k)}{\lambda_c(z_k^\theta)} \\
&= \quad \left\{\text{If } \lambda_c(z_k^\theta) = \lambda_c, \ \forall \theta \in \{1, 2, \ldots, m_k\}\right\} \\
&= \underset{\theta \in \{1,2,\ldots,m_k\}}{\arg\max} \; \frac{\exp\left(-\frac{1}{2}\left(z_k^\theta - \bar{z}_{k|k-1}\right)^T S_k^{-1} \left(z_k^\theta - \bar{z}_{k|k-1}\right)\right)}{|2\pi S_k|^{1/2}} \\
&= \underset{\theta \in \{1,2,\ldots,m_k\}}{\arg\min} \; \left(z_k^\theta - \bar{z}_{k|k-1}\right)^T S_k^{-1} \left(z_k^\theta - \bar{z}_{k|k-1}\right)
\end{aligned}
$$

The nearest neighbour

- Under certain assumptions, $z_k^{\theta_k^+}$ is the **"nearest" neighbour** to $\bar{z}_{k|k-1}$, where S_k is used to define the distance.

EXAMPLE FOR VISUALIZATION (HARD)

$$\begin{aligned}
\text{Prior density :} \quad & p(x_1) = \mathcal{N}(x_1; 0.5, 0.2) \\
\text{Motion model :} \quad & \pi_k(x_k \middle| x_{k-1}) = \mathcal{N}(x_k; x_{k-1}, 0.35) \\
\text{Probability of detection:} \quad & P^D(x) = 0.9 \\
\text{Object likelihood :} \quad & g_k(o_k \middle| x_k) = \mathcal{N}(o_k; x_k, 0.2) \\
\text{Clutter intensity :} \quad & \lambda(c) = \begin{cases} 0.4 & \text{if } |c| \leq 4 \\ 0 & \text{otherwise} \end{cases} \\
\text{Observed detections :} \quad & Z_1 = [-1.3,\ 1.7], \quad Z_2 = [1.3], \\
& Z_3 = [-0.3,\ 2.3], \quad Z_4 = [-0.7,\ 3] \\
& Z_5 = [-1], \quad Z_6 = [-1.3]
\end{aligned}$$

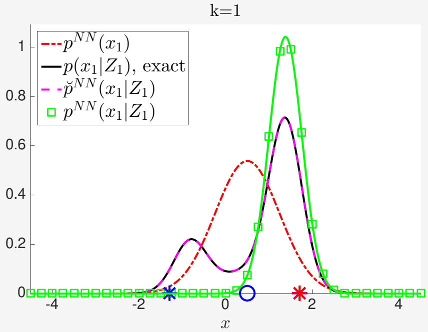
A SECOND NEAREST NEIGHBOUR FILTERING EXAMPLE

Example: NN vs exact posterior

- We compare four densities:

$$p^{\mathrm{NN}}(x_k|Z_{1:k-1}), \quad p(x_k|Z_{1:k})$$
$$\breve{p}^{\mathrm{NN}}(x_k|Z_{1:k}), \quad p^{\mathrm{NN}}(x_k|Z_{1:k})$$

- We visualize the m_k+1 hypotheses, and mark the most probable in red.
- The NN may lose track of the object in complicated scenarios.



NEAREST NEIGHBOUR FILTERING: SUMMARY

Basic idea

- Prune all hypotheses except the most probable one.

Pros and cons

++ A fast algorithm which is simple to implement.

++ Works well in simple scenarios.

−− Ignores uncertainties which increases the risk that we will lose track of the object.

−− Performs poorly in complicated scenarios.

Probabilistic data association filtering

Multi-Object Tracking

Lennart Svensson

PROBLEM SETTING

- PDA approximates posterior and predicted densities as **Gaussian** :

$$p^{\text{PDA}}(x_{k-1}|Z_{1:k-1}) = \mathcal{N}(x_{k-1}; \bar{x}^{\text{PDA}}_{k-1|k-1}, P^{\text{PDA}}_{k-1|k-1}),$$
$$p^{\text{PDA}}(x_{k}|Z_{1:k-1}) = \mathcal{N}(x_{k}; \bar{x}^{\text{PDA}}_{k|k-1}, P^{\text{PDA}}_{k|k-1}).$$

Measurement model (same as for NN)

We assume $P^{\mathrm{D}}(x) = P^{\mathrm{D}}$, $g_k(o|x) = \mathcal{N}(o; H_k x, R_k)$, general $\lambda_c(c)$.

Posterior density, given $p^{\text{PDA}}(x_k|Z_{1:k-1}) = \mathcal{N}(x_k; \bar{x}^{\text{PDA}}_{k|k-1}, P^{\text{PDA}}_{k|k-1})$

Posterior is $\breve{p}^{\text{PDA}}(x_k|Z_{1:k}) = \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} p^{\theta_k}_{k|k}(x_k)$ where $p^{\theta_k}_{k|k}(x_k) = \mathcal{N}(x_k; \hat{x}^{\theta_k}_k, P^{\theta_k}_k)$.

- How can we **approximate** $\breve{p}^{\text{PDA}}(x_k|Z_{1:k})$ as Gaussian?

Basic idea

- Approximate posterior as a Gaussian density with the same mean and covariance as $\breve{p}^{\text{PDA}}(x_k | Z_{1:k})$.

- We set

$$\bar{x}_{k|k}^{\text{PDA}} = \mathbb{E}_{\breve{p}^{\text{PDA}}(x_k|Z_{1:k})}\left[x_k\right]$$

$$P_{k|k}^{\text{PDA}} = \text{Cov}_{\breve{p}^{\text{PDA}}(x_k|Z_{1:k})}\left[x_k\right]$$

- **Note 1:** we then assume that $p^{\text{PDA}}(x_k | Z_{1:k}) = \mathcal{N}(x_k; \bar{x}_{k|k}^{\text{PDA}}, P_{k|k}^{\text{PDA}})$.

PROBABILISTIC DATA ASSOCIATION FILTERING

Basic idea

- Approximate posterior as a Gaussian density with the same mean and covariance as $\breve{p}^{\mathrm{PDA}}(x_k | Z_{1:k})$.

- We set

$$\bar{x}_{k|k}^{\mathrm{PDA}} = \mathbb{E}_{\breve{p}^{\mathrm{PDA}}(x_k|Z_{1:k})}[x_k] = \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} \hat{x}_k^{\theta_k}$$

$$P_{k|k}^{\mathrm{PDA}} = \mathrm{Cov}_{\breve{p}^{\mathrm{PDA}}(x_k|Z_{1:k})}[x_k] = \sum_{\theta_k=0}^{m_k} \underbrace{w_k^{\theta_k} P_k^{\theta_k}}_{\text{average cov.}} + \underbrace{w_k^{\theta_k} \left(\bar{x}_{k|k}^{\mathrm{PDA}} - \hat{x}_k^{\theta_k}\right)\left(\bar{x}_{k|k}^{\mathrm{PDA}} - \hat{x}_k^{\theta_k}\right)^T}_{\text{spread of mean}}.$$

- **Note 2:** this minimizes the Kullback-Leibler divergence

$$\int \breve{p}^{\mathrm{PDA}}(x_k|Z_{1:k}) \log \frac{\breve{p}^{\mathrm{PDA}}(x_k|Z_{1:k})}{\mathcal{N}(x_k; \bar{x}_{k|k}^{\mathrm{PDA}}, P_{k|k}^{\mathrm{PDA}})} \, \mathrm{d}x_k.$$

MOMENTS OF A GAUSSIAN MIXTURE

Moments of a Gaussian mixture

- Suppose

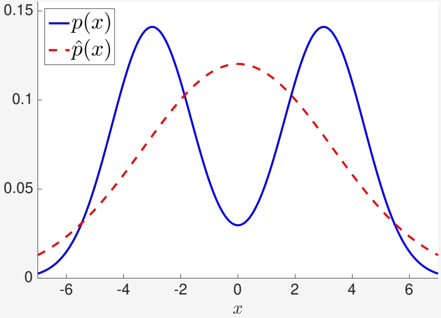
$$p(x) = 0.5\mathcal{N}(x; -3, 2) + 0.5\mathcal{N}(x; 3, 2).$$

- It follows that

$$\begin{aligned} \mathbb{E}_{p(x)}[x] &= 0.5 \times (-3) + 0.5 \times 3 = 0, \\ \mathrm{Cov}_{p(x)}[x] &= \underbrace{0.5 \times 2 + 0.5 \times 2}_{=2} \\ &\quad + \underbrace{0.5 \times 3^2 + 0.5 \times (-3)^2}_{=9} = 11. \end{aligned}$$

- Similar to PDA, we can approximate $p(x)$ using

$$\hat{p}(x) = \mathcal{N}(x; 0, 11).$$



THE PDA FILTERING ALGORITHM

Basic idea

- Approximate posterior as a Gaussian density with the same mean and covariance as $\breve{p}^{\text{PDA}}(x_k \mid Z_{1:k})$.

Algorithm The PDA filtering update.

1: Compute $w_k^{\theta_k}$, $\hat{x}_k^{\theta_k}$ and $P_k^{\theta_k}$, $\theta_k = 0, 1 \ldots, m_k$.

2: Set

$$\bar{x}_{k|k}^{\text{PDA}} = \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} \hat{x}_k^{\theta_k}.$$

3: Compute

$$P_{k|k}^{\text{PDA}} = \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} P_k^{\theta_k} + w_k^{\theta_k} \left(\bar{x}_{k|k}^{\text{PDA}} - \hat{x}_k^{\theta_k}\right) \left(\bar{x}_{k|k}^{\text{PDA}} - \hat{x}_k^{\theta_k}\right)^T.$$

Probabilistic data association filtering – remarks and visualizations

Multi-Object Tracking

Lennart Svensson

PDA FILTERING: POSTERIOR MEAN

- For linear and Gaussian models, and $\bar{z}_{k|k-1} = H_k \bar{x}^{\mathrm{PDA}}_{k|k-1}$,

$$\hat{x}_k^{\theta_k} = \begin{cases} \bar{x}^{\mathrm{PDA}}_{k|k-1} & \text{if } \theta_k = 0 \\ \bar{x}^{\mathrm{PDA}}_{k|k-1} + K_k(z_k^{\theta_k} - \bar{z}_{k|k-1}) & \text{if } \theta_k \in \{1, 2, \dots, m_k\}. \end{cases}$$

- Hence, the **posterior mean** is

$$\begin{aligned} \bar{x}^{\mathrm{PDA}}_{k|k} &= \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} \hat{x}_k^{\theta_k} \\ &= \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} \bar{x}^{\mathrm{PDA}}_{k|k-1} + \sum_{\theta_k=1}^{m_k} w_k^{\theta_k} K_k(z_k^{\theta_k} - \bar{z}_{k|k-1}) \\ &= \bar{x}^{\mathrm{PDA}}_{k|k-1} + K_k \underbrace{\sum_{\theta_k=1}^{m_k} w_k^{\theta_k}(z_k^{\theta_k} - \bar{z}_{k|k-1})}_{\text{expected innovation}}. \end{aligned}$$

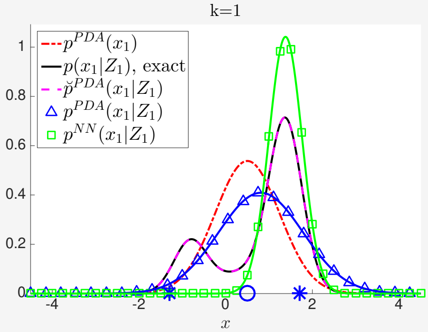
A FIRST PDA FILTERING EXAMPLE

Example: PDA vs NN and exact posterior

- We compare five densities:

$$p^{\text{PDA}}(x_k|Z_{1:k-1}), \quad p(x_k|Z_{1:k}),$$
$$\breve{p}^{\text{PDA}}(x_k|Z_{1:k}), \quad p^{\text{PDA}}(x_k|Z_{1:k}),$$
$$p^{\text{NN}}(x_k|Z_{1:k}).$$

- We visualize the $m_k + 1$ hypotheses.
- PDA yields larger posterior uncertainties than NN filtering.



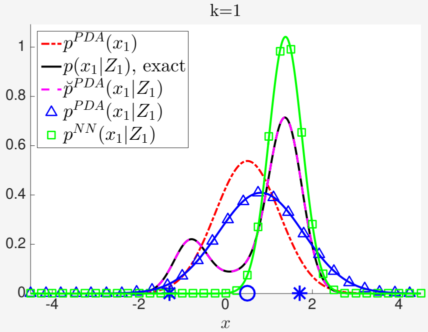
A SECOND PDA FILTERING EXAMPLE

Example: PDA vs NN and exact posterior

- We compare five densities:

$$p^{\text{PDA}}(x_k|Z_{1:k-1}), \quad p(x_k|Z_{1:k}),$$
$$\breve{p}^{\text{PDA}}(x_k|Z_{1:k}), \quad p^{\text{PDA}}(x_k|Z_{1:k}),$$
$$p^{\text{NN}}(x_k|Z_{1:k}).$$

- We visualize the $m_k + 1$ hypotheses.
- In difference to NN, the PDA algorithm did not lose track of object.



PROBABILISTIC DATA ASSOCIATION: SUMMARY

Basic idea

- Approximate posterior as a Gaussian density with the same mean and covariance as $\breve{p}^{\text{PDA}}(x_k \mid Z_{1:k})$.

Pros and cons

++ A fast algorithm which is simple to implement.

++ Works well in simple scenarios.

++ Acknowledges uncertainties slightly better than NN.

−− Performs poorly in complicated scenarios.

Gaussian sum filtering

Multi-Object Tracking

Lennart Svensson

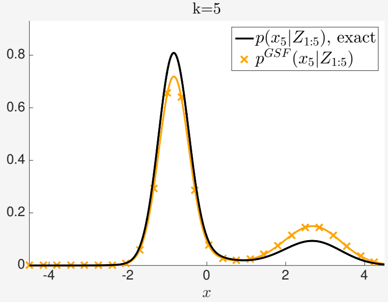
GAUSSIAN SUM FILTERING (1)

Gaussian sum filtering

- **Basic idea:** approximate the posterior as a Gaussian mixture with a few components.

Example

- In the figure to the right, the posterior contains 108 hypotheses.
- We prune all but 5 hypotheses (at all times).
- Approximation is significantly better than PDA and NN.



GAUSSIAN SUM FILTERING (2)

Gaussian sum filtering

- **Basic idea:** approximate the posterior as a Gaussian mixture with a few components.

Prediction and update of a Gaussian mixture

- Suppose $p^{h_{k-1}}_{k-1|k-1}(x_{k-1}) = \mathcal{N}(x_{k-1}; \hat{x}^{h_{k-1}}_{k-1|k-1}, P^{h_{k-1}}_{k-1|k-1})$ and

$$p^{GSF}(x_{k-1} \big| Z_{1:k-1}) = \sum_{h_{k-1}=1}^{\mathcal{H}_{k-1}} w^{h_{k-1}}_{k-1} p^{h_{k-1}}_{k-1|k-1}(x_{k-1}).$$

- Assuming “linear and Gaussian” models, posterior at time k is a Gaussian mixture

$$\breve{p}^{GSF}(x_k \big| Z_{1:k}) = \sum_{h_k=1}^{\mathcal{H}_{k-1}\times(m_k+1)} \breve{w}^{h_k}_k \breve{p}^{h_k}_{k|k}(x_k).$$

- How can we approximate $\breve{p}^{GSF}(x_k \big| Z_{1:k})$ as a Gaussian mixture with **fewer terms**?

PRUNING HYPOTHESES WITH SMALL WEIGHTS

Basic idea

- Prune all hypotheses whose weights are smaller than a threshold γ.

Example

- Suppose
$$p(x) = 0.7\mathcal{N}(x; \hat{x}^1, P^1) + 0.005\mathcal{N}(x; \hat{x}^2, P^2) + 0.295\mathcal{N}(x; \hat{x}^3, P^3)$$
and that $\gamma = 0.01$.

- Pruning then yields
$$\begin{aligned} p(x) \approx \acute{p}(x) &= \frac{0.7}{0.295 + 0.7}\mathcal{N}(x; \hat{x}^1, P^1) + \frac{0.295}{0.295 + 0.7}\mathcal{N}(x; \hat{x}^3, P^3) \\ &= \acute{w}_1\, \mathcal{N}(x; \acute{x}^1, \acute{P}^1) + \acute{w}_2\, \mathcal{N}(x; \acute{x}^2, \acute{P}^2). \end{aligned}$$

MERGING SIMILAR COMPONENTS

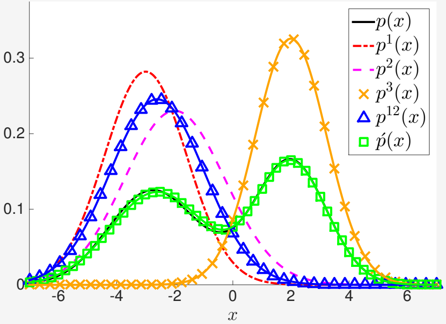
Merging two out of three components

- Suppose $p^1(x)$ and $p^2(x)$ are similar.
- Setting $w^{12} = w^1 + w^2$ we get

$$p(x) = w^1 p^1(x) + w^2 p^2(x) + w^3 p^3(x)$$

$$= w^{12} \underbrace{\left(\frac{w^1 p^1(x)}{w^{12}} + \frac{w^2 p^2(x)}{w^{12}} \right)}_{\approx p^{12}(x), \text{ see PDA}} + w^3 p^3(x)$$

$$\approx w^{12} p^{12}(x) + w^3 p^3(x) = \acute{p}(x)$$



- We select $p^{12}(x)$ to match moments of

$$\frac{w^1 p^1(x)}{w^{12}} + \frac{w^2 p^2(x)}{w^{12}}.$$

CAPPING THE NUMBER OF HYPOTHESES

Basic idea

- Prune hypotheses until we are left with at most $N_{\max}$ hypotheses.

Algorithm Capping the number of hypotheses.

```
Input: N_max, w^i, x̂^i, P^i, i = 1, ..., H > N_max.
Output: ẃ^i, x́^i, Ṕ^i, i = 1, ..., H́ = N_max
[out, ind] = sort([w^1, ..., w^H], 'descend').
% Gives a list 'ind' with indexes w^ind(1) ≥ w^ind(2) ≥ ··· ≥ w^ind(H).
Compute c = sum_{i=1}^{N_max} w_{k|k}^{ind(i)}.
for i = 1 to N_max do
    Set ẃ^i = w^ind(i)/c, x́^i = x̂^ind(i) and Ṕ^i = P^ind(i).
end for
```

Input: $N_{\max}, w^i, \hat{x}^i, P^i$, $i = 1, \ldots, \mathcal{H} > N_{\max}$.

Output: $\acute{w}^i, \acute{x}^i, \acute{P}^i$, $i = 1, \ldots, \acute{\mathcal{H}} = N_{\max}$

$[out, ind] = \text{sort}([w^1, \ldots, w^{\mathcal{H}}], \text{'descend'})$.

% Gives a list 'ind' with indexes $w^{ind(1)} \geq w^{ind(2)} \geq \cdots \geq w^{ind(\mathcal{H})}$.

Compute $c = \sum_{i=1}^{N_{\max}} w_{k|k}^{ind(i)}$.

for $i = 1$ **to** $N_{\max}$ **do**

  Set $\acute{w}^i = w^{ind(i)}/c$, $\acute{x}^i = \hat{x}^{ind(i)}$ and $\acute{P}^i = P^{ind(i)}$.

end for

SUMMARY OF MIXTURE REDUCTION STRATEGIES

- We have described three ways to reduce the number of hypotheses in

$$\breve{p}^{GSF}(x_k|Z_{1:k}) = \sum_{h_k=1}^{\breve{\mathcal{H}}_k} \breve{w}_k^{h_k} \breve{p}_{k|k}^{h_k}(x_k).$$

- These can be combined in different ways, e.g.,
    1. cap the number of hypotheses at $N_{\max}$, or,
    2. we can
        i) remove hypotheses with weights $< \gamma$,
        ii) merge similar components, and then,
        iii) cap the number of hypotheses at $N_{\max}$.

- The resulting Gaussian mixture is the GSF posterior

$$p^{GSF}(x_k|Z_{1:k}) = \sum_{h_k=1}^{\mathcal{H}_k} w_k^{h_k} p_{k|k}^{h_k}(x_k).$$

Gaussian sum filtering – prediction and update equations

Multi-Object Tracking

Lennart Svensson

PREDICTION AND UPDATE EQUATIONS

Prediction and update equations for Gaussian sum filters

- Suppose
$$p^{GSF}(x_{k-1}|Z_{1:k-1}) = \sum_{h_{k-1}=1}^{\mathcal{H}_{k-1}} w_{k-1}^{h_{k-1}} p_{k-1|k-1}^{h_{k-1}}(x_{k-1}).$$

- It then follows that the predicted and updated densities are
$$p^{GSF}(x_k|Z_{1:k-1}) = \sum_{h_{k-1}=1}^{\mathcal{H}_{k-1}} w_{k-1}^{h_{k-1}} p_{k|k-1}^{h_{k-1}}(x_k)$$
$$\breve{p}^{GSF}(x_k|Z_{1:k}) = \sum_{h_k=1}^{\breve{\mathcal{H}}_k} \breve{w}_k^{h_k} \breve{p}_{k|k}^{h_k}(x_k),$$
where $\breve{\mathcal{H}}_k = \mathcal{H}_{k-1} \times (m_k + 1)$.

- In this video, we present equations for computing $p_{k|k-1}^{h_{k-1}}(x_k)$, $\breve{w}_k^{h_k}$ and $\breve{p}_{k|k}^{h_k}(x_k)$.

PREDICTION STEP

Chapman-Kolmogorov for every hypothesis

- For $h_{k-1} = 1, 2, \ldots, \mathcal{H}_{k-1}$,

$$p^{h_{k-1}}_{k|k-1}(x_k) = \int \pi_k(x_k | x_{k-1}) p^{h_{k-1}}_{k-1|k-1}(x_{k-1}) \, \mathrm{d}x_{k-1}.$$

Linear and Gaussian prediction

- If

$$\begin{cases} p^{h_{k-1}}_{k-1|k-1}(x_{k-1}) = \mathcal{N}(x_{k-1}; \hat{x}^{h_{k-1}}_{k-1|k-1}, P^{h_{k-1}}_{k-1|k-1}) \, \mathrm{d}x_{k-1} \\ x_k = F_{k-1} x_{k-1} + q_{k-1}, \qquad q_{k-1} \sim \mathcal{N}(0, Q_{k-1}), \end{cases}$$

  then

$$p^{h_{k-1}}_{k|k-1}(x_k) = \mathcal{N}(x_k; \hat{x}^{h_{k-1}}_{k|k-1}, P^{h_{k-1}}_{k|k-1})$$

  where

$$\hat{x}^{h_{k-1}}_{k|k-1} = F_{k-1} \hat{x}^{h_{k-1}}_{k-1|k-1}, \qquad P^{h_{k-1}}_{k|k-1} = F_{k-1} P^{h_{k-1}}_{k-1|k-1} F^T_{k-1} + Q_{k-1}.$$

UPDATE STEP (1)

Updated weights and densities

- For every pair of hypotheses

$$h_{k-1} \in \{1, 2, \ldots, \mathcal{H}_{k-1}\} \quad \text{and } \theta_k \in \{0, 1, \ldots, m_k\}$$

  we obtain a new hypothesis h_k:

$$\breve{w}_{k|k}^{h_k} \propto \begin{cases} w_{k-1}^{h_{k-1}} \int (1 - P^D(x_k)) p_{k|k-1}^{h_{k-1}}(x_k) \, \mathrm{d}x_k & \text{if } \theta_k = 0 \\ \frac{w_{k-1}^{h_{k-1}}}{\lambda_c(z_k^{\theta_k})} \int p_{k|k-1}^{h_{k-1}}(x_k) P^D(x_k) g_k(z_k^{\theta_k} | x_k) \, \mathrm{d}x_k & \text{if } \theta_k \in \{1, 2, \ldots, m_k\}, \end{cases}$$

$$\breve{p}_{k|k}^{h_k}(x_k) \propto \begin{cases} p_{k|k-1}^{h_{k-1}}(x_k)(1 - P^D(x_k)) & \text{if } \theta_k = 0 \\ p_{k|k-1}^{h_{k-1}}(x_k) P^D(x_k) g_k(z_k^{\theta_k} | x_k) & \text{if } \theta_k \in \{1, 2, \ldots, m_k\}. \end{cases}$$

UPDATE STEP (2)

Updated weights and densities

- When P^{D} is constant and g_k is linear and Gaussian:

$$\breve{w}_{k|k}^{h_k} \propto \begin{cases} w_{k-1}^{h_{k-1}}(1 - P^D(x_k)) & \text{if } \theta_k = 0 \\ \frac{w_{k-1}^{h_{k-1}} P^{\mathrm{D}} \mathcal{N}(z_k^{\theta_k}; \bar{z}_{k|k-1}^{h_{k-1}}, S_{k,h_{k-1}})}{\lambda_c(z_k^{\theta_k})} & \text{if } \theta_k \in \{1, 2, \ldots, m_k\}, \end{cases}$$

$$\breve{p}_{k|k}^{h_k}(x_k) = \mathcal{N}(x_k; \breve{x}_{k|k}^{h_k}, \breve{P}_{k|k}^{h_k}) \text{ where}$$

$$\theta_k = 0 : \qquad \begin{cases} \breve{x}_{k|k}^{h_k} = \hat{x}_{k|k-1}^{h_{k-1}} \\ \breve{P}_{k|k}^{h_k} = P_{k|k-1}^{h_{k-1}}, \end{cases}$$

$$\theta_k \in \{1, 2, \ldots, m_k\} : \qquad \begin{cases} \breve{x}_{k|k}^{h_k} = \hat{x}_{k|k-1}^{h_{k-1}} + K_k^{h_{k-1}} \left(z_k^{\theta_k} - \bar{z}_{k|k-1}^{h_{k-1}} \right) \\ \breve{P}_{k|k}^{h_k} = P_{k|k-1}^{h_{k-1}} - K_k^{h_{k-1}} H_k P_{k|k-1}^{h_{k-1}}. \end{cases}$$

UPDATE STEP (3)

- We obtain a hypothesis h_k for every pair of h_{k-1} and θ_k, but how can we index h_k?
- Two possibilities:

$$
\begin{aligned}
&1) \qquad h_k = h_{k-1} + \mathcal{H}_{k-1}\theta_k \\
&2) \qquad h_k = 1 + \theta_k + \mathcal{H}_{k-1}(h_{k-1} - 1).
\end{aligned}
$$

  Both ensure that we have a one-to-one mapping between (h_{k-1}, θ_k) and h_k.

Indexing four new hypotheses

- If $\mathcal{H}_{k-1} = 2$, $m_k = 1$ and $h_k = h_{k-1} + \mathcal{H}_{k-1}\theta_k$:

$$
\begin{aligned}
h_{k-1} = 1, \theta_k = 0 &\Leftrightarrow h_k = 1 \\
h_{k-1} = 2, \theta_k = 0 &\Leftrightarrow h_k = 2 \\
h_{k-1} = 1, \theta_k = 1 &\Leftrightarrow h_k = 3 \\
h_{k-1} = 2, \theta_k = 1 &\Leftrightarrow h_k = 4.
\end{aligned}
$$

Gaussian sum filtering – estimation and visualizations

Multi-Object Tracking

Lennart Svensson

- If the posterior is a Gaussian mixture, how can we estimate x_k?

Minimum mean square error (MMSE) estimation

- The posterior mean

$$\bar{x}_{k|k} = \mathbb{E}\left[x_k \middle| Z_{1:k}\right] = \sum_{h_k=1}^{\mathcal{H}_k} w_{k|k}^{h_k} \hat{x}_{k|k}^{h_k}$$

minimizes the MMSE, $\mathbb{E}\left[(x_k - \bar{x}_{k|k})^T (x_k - \bar{x}_{k|k}) \middle| Z_{1:k}\right]$.

Most probably hypothesis estimation

- For multi-modal densities, we sometimes prefer

$$h_k^\star = \arg\max_h w_{k|k}^h$$

$$\hat{x}_{k|k} = \hat{x}_{k|k}^{h_k^\star}.$$

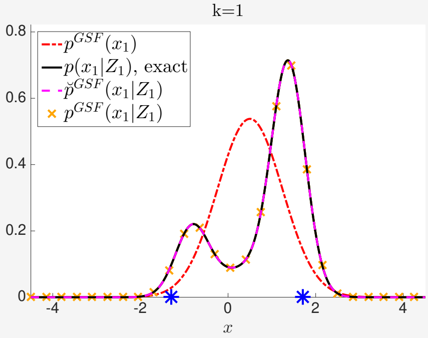
A FIRST GSF FILTERING EXAMPLE

Example: GSF vs exact posterior

- We compare four densities:

$$p^{\text{GSF}}(x_k|Z_{1:k-1}), \quad p(x_k|Z_{1:k}),$$
$$\breve{p}^{\text{GSF}}(x_k|Z_{1:k}), \quad p^{\text{GSF}}(x_k|Z_{1:k}).$$

- The number of hypotheses is capped at $N_{\max} = 5$.
- The GSF filter approximates the posterior well.



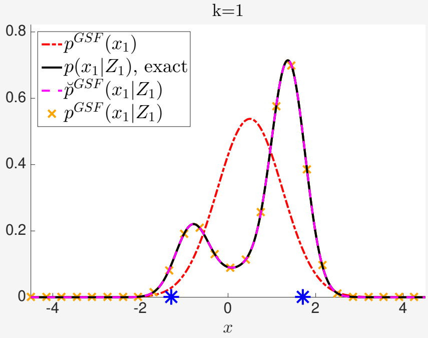
A SECOND GSF FILTERING EXAMPLE

Example: GSF vs exact posterior

- We compare four densities:

$$p^{\text{GSF}}(x_k|Z_{1:k-1}), \quad p(x_k|Z_{1:k}),$$
$$\breve{p}^{\text{GSF}}(x_k|Z_{1:k}), \quad p^{\text{GSF}}(x_k|Z_{1:k}).$$

- The number of hypotheses is capped at $N_{\max} = 5$.
- The GSF approximates the posterior significantly better than NN and PDA.



GAUSSIAN SUM FILTERING – SUMMARY

Basic idea

- Approximate the posterior as a Gaussian mixture with a few components.

Pros and cons

++ Significantly more accurate than NN and PDA.

++ Complexity can be adjusted to computational resources.

– – More complicated to implement than NN/PDA.

– – More computationally demanding to run than NN/PDA.

- **Note:** even though GSFs looks much more accurate than NN and PDA, the difference is mostly noticeable in medium-difficult settings.

Gating to remove unlikely hypotheses

Multi-Object Tracking

Lennart Svensson

MOTIVATION

PDA with large m_k

- Suppose we have an amazing sensor:
  - large P^{D},
  - small λ_c,
  - huge field of view.
- Excellent conditions, but m_k may be **very large**.

- PDA:

$$\bar{x}_{k|k}^{\text{PDA}} = \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} \hat{x}_k^{\theta_k}$$

$$P_{k|k}^{\text{PDA}} = \sum_{\theta_k=0}^{m_k} w_k^{\theta_k} P_k^{\theta_k} + w_k^{\theta_k} \left(\bar{x}_{k|k}^{\text{PDA}} - \hat{x}_k^{\theta_k}\right)\left(\bar{x}_{k|k}^{\text{PDA}} - \hat{x}_k^{\theta_k}\right)^T$$

- Do we have to compute $w_k^{\theta_k}$, $\hat{x}_k^{\theta_k}$ and $P_k^{\theta_k}$ for hypotheses θ_k: $w_k^{\theta_k} \approx 0$?
- **Gating enables us to avoid this**! (Not only for PDA.)

BASIC IDEA

Idea

- Form a gate around the predicted measurement, and only consider detections within the gate.

- Gating leads to much fewer local hypotheses.
- A gate may be designed in many different ways, e.g., rectangular.
- Here, we study one which is natural for Gaussian distributions, namely the **ellipsoidal gate**.

ELLIPSOIDAL GATES: MOTIVATION AND DEFINITION

- We consider $\theta_k > 0$. Recall that

$$\tilde{w}_k^{h_k} = \frac{P^{\mathrm{D}}(x_k)\mathcal{N}(z_k^{\theta_k}; \bar{z}_{k|k}^{h_{k-1}}, S_{k,h_{k-1}})}{\lambda_c(z_k^{\theta_k})}.$$

- We note that $\tilde{w}_k^{h_k}$ is "small" when the distance

$$d^2_{h_{k-1},\theta_k} = (z_k^{\theta_k} - \bar{z}_{k|k}^{h_{k-1}})^T S_{k,h_{k-1}}^{-1}(z_k^{\theta_k} - \bar{z}_{k|k}^{h_{k-1}})$$

is large (if $\lambda_c \approx$ constant).

Ellipsoidal gate

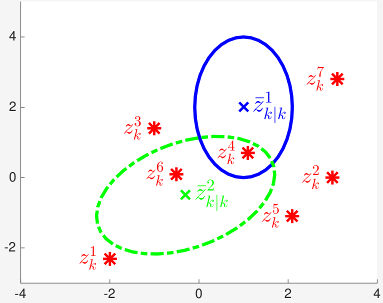
- Disregard $z_k^{\theta_k}$ as a clutter detection under hypothesis h_{k-1}, if

$$d^2_{h_{k-1},\theta_k} > G.$$

VISUALIZING GATING

Example: gating 2D measurements

- We have seven measurements and two predicted hypotheses, $h_{k-1} = 1$ and $h_{k-1} = 2$.
- The ellipsoids illustrate the two gates.
- For $h_{k-1} = 1$ only z_k^4 is inside the gate.
- For $h_{k-1} = 2$ all measurements except z_k^4 and z_k^6 are outside the gate.



SELECTING THE THRESHOLD G

- If G is small, we may have a “large” probability that the object detection is outside the gate.
- Given h_{k-1} and θ_k, where $\theta_k > 0$, the probability that the object measurement is outside the gate is

$$P_G = \Pr\left[d^2_{h_{k-1},\theta_k} > G \middle| h_{k-1}, \theta_k\right].$$

- One can show that

$$d^2_{h_{k-1},\theta_k} \middle| h_{k-1}, \theta_k \sim \chi^2(n_z).$$

- A common strategy is to set a desired value for P_G, say, 99.5%, and then use the cumulative distribution of $\chi^2(n_z)$ to find G.

GATING – A SUMMARY

- Gating is a technique to disregard measurements as clutter (given h_{k-1}) without computing the weights.
- Gating can be combined with all tracking algorithms presented later.
- In Gaussian settings, the ellipsoidal gate is a natural choice which is simple to implement.
- It is important to find a reasonable value for the threshold G.