Jacobs University Bremen

Filtering and fusion

Recap of probability
Bayes theory

Automation
CO23-320203

Recap of filtering so far

- We have developed and analysed first order complementary filter through block diagram and Laplace analysis
- We have equipped it with an integral term which serves as a bias estimator and subjected it to Lyapunov stability analysis
- We looked at a more advanced, non-linear scheme using bias observer and discretised it for efficient implementation
- While elegant and practical, complementary filtering is not the only method of data fusion
- The next subject will introduce broadly used sensor fusion strategy based on the probabilistic/statistical formulation of the filter (=estimator)

Probability basics: Probability distributions

A random variable (RV) X is an event which can take values from domain of X (dom X). p(X = x) ≡ p(x) is the probability that the variable X takes the value x, or alternately the event x occurs.

- If dom X = $\{x_1, \ldots, x_n\}$ is discrete, we have have a **probability mass function** (pmf) p(X = **x**) such that

$$\sum_{\mathbf{x}} p(X = \mathbf{x}) = 1$$

- x If dom X ⊆ R^n then we have a **probability density function** (pdf) ρ(x) such that

$$\int \ldots \int_{dom\ X} \rho(\mathbf{x}) dx_1 \ldots dx_n = 1$$

- If dom X ⊆ R

$$\int_{x_1}^{x_2} \rho(x) dx = p(x_1 \leq x \leq x_2)$$

Probability basics: Conditional probability

p(x | y) is probability p(X = x) given that (conditioned on the fact that) the event Y = y has already occurred

- Given random variables X and Y, their joint probability distribution is

$$p(x, y) \equiv p(X = x \text{ and } Y = y) = p(x|y)p(y) = p(y|x)p(x)$$

- A multivariate pmf or pdf p(x) is actually a joint distribution pmf or pdf $p(x_1, x_2, \ldots, x_n)$. Therefore

$$p(\mathbf{x}) \equiv p(x_1, x_2, \ldots, x_n) =$$
$$= p(x_1|x_2, \ldots, x_n)p(x_2|x_2, \ldots, x_n)\ldots p(x_{n-1}|x_n)p(x_n)$$

- If X and Y are independent p(x, y) = p(x)p(y). If X and Y are mutually exclusive p(x, y) = 0

Probability basics: Marginalisation

Marginalisation w.r.t. y or x_n

- Discrete case:

$$p(x) = \sum_y p(x \mid y)p(y) = \sum_y p(x, y)$$

$$p(x_1, x_2, \ldots, x_{n-1}) = \sum_{x_n} p(x_1, x_2, \ldots, x_{n-1}, x_n)$$

- Continuous case:

$$\rho(x) = \int \rho(x \mid y)\rho(y)dy = \int \rho(x, y)dy$$

$$\rho(x_1, \ldots, x_{n-1}) = \int_{x_n} \rho(x_1, \ldots, x_{n-1}, x_n)dx_n$$

Probability basics: Bayes rule

- Discrete case:

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)} \qquad = \frac{p(y \mid x)p(x)}{\sum_x p(y \mid x)p(x)}$$

- Continuous case:

$$\rho(x \mid y) = \frac{\rho(y \mid x)\rho(x)}{\rho(y)} \qquad = \frac{\rho(y \mid x)\rho(x)}{\int_x \rho(y \mid x)\rho(x)dx}$$

- Related terminology: X is the **hypothesis**, Y is the **data**

Probability basics: More terminology

- Likelihood:

$$p(\text{Data} \mid \text{Hypothesis})$$

- Posterior:

$$p(\text{Hypothesis} \mid \text{Data})$$

- Prior:

$$p(\text{Hypothesis})$$

- Evidence:

$$p(\text{Data})$$

$$\text{Posterior probability} \propto \text{Likelihood} \times \text{Prior probability}$$

Probability basics: Expectation value

- Discrete case:

$$E[X] = \bar{\mathbf{x}} = \sum_{\mathbf{x}} \mathbf{x} p(\mathbf{x}) = \sum_{x_1} \cdots \sum_{x_n} \mathbf{x} p(x_1, \ldots, x_n)$$

- Continuous case:

$$E[X] = \bar{\mathbf{x}} = \int_{x_1} \cdots \int_{x_n} \mathbf{x} \rho(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

- RV transformed by a linear relation, Y: **y** = f(**x**):

$$E[Y] = \bar{\mathbf{y}} = \sum_{\mathbf{x}} \mathbf{f}(\mathbf{x}) p(\mathbf{x}) = \sum_{x_1} \cdots \sum_{x_n} \mathbf{f}(\mathbf{x}) p(x_1, \ldots, x_n)$$

- Continuous case:

$$E[Y] = \bar{\mathbf{y}} = \int_{x_1} \cdots \int_{x_n} \mathbf{f}(\mathbf{x}) \rho(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

Probability basics: Covariance

- For a random variable X:

$$\text{Cov}[X] = \mathbf{C}(\mathbf{x}) = E\left[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T\right]$$

$$C_{ij}(\mathbf{x}) = E\left[(x_i - \bar{x}_i)(x_j - \bar{x}_j)\right]$$

  - C(x) is a positive semidefinite matrix
  - Example:

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} \qquad \mathbf{C}(\mathbf{x}) = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{x\theta} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{y\theta} \\ \sigma_{x\theta} & \sigma_{y\theta} & \sigma_\theta^2 \end{pmatrix}$$

Probability basics: Covariance properties

- If we define two random variables Y: **y** = **Ax** + **b** and Z $\subseteq$ R^n, the covariance will transform in the following way:

$$\mathbf{C}(\mathbf{y}) = \mathbf{A}\mathbf{C}(\mathbf{x})\mathbf{A}^T$$

$$\mathbf{C}(\mathbf{y}, \mathbf{z}) = \mathbf{A}\mathbf{C}(\mathbf{x}, \mathbf{z})$$

$$\mathbf{C}(\mathbf{z}, \mathbf{y}) = \mathbf{C}(\mathbf{z}, \mathbf{x})\mathbf{A}^T$$

- And the expectation:

$$\bar{\mathbf{y}} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{b}$$

Probability basics: Covariance properties

- If we have a RV created by a nonlinear function Y: **y** = f(**x**), we can linearize it about a point x_* by using the Taylor Series:

$$\mathbf{y} \approx \mathbf{f}(\mathbf{x}_\star) + \nabla \mathbf{f}\big|_{\mathbf{x}_\star} (\mathbf{x} - \mathbf{x}_\star) \equiv \mathbf{A}\tilde{\mathbf{x}} + \mathbf{b}, \text{ where,}$$

$$\tilde{\mathbf{x}} \equiv \mathbf{x} - \mathbf{x}_\star, \mathbf{A} \equiv \nabla \mathbf{f}\big|_{\mathbf{x}_\star}, \mathbf{b} \equiv \mathbf{f}(\mathbf{x}_\star)$$

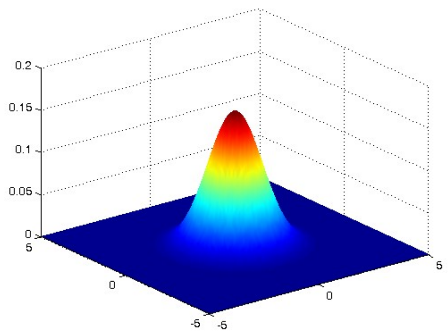
  - All relationships from from the previous slide can be applied to the new variable in an approximate way

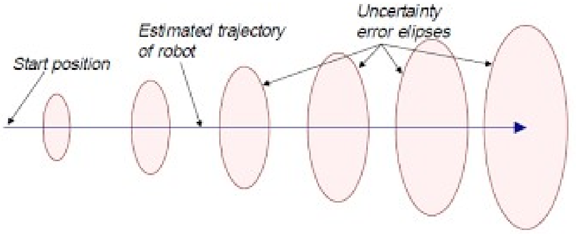
Gaussian distribution

- If $x \in R^n$ is a vector random variable, the normal/Gaussian pdf is defined as:

$$\mathcal{N}(\mathbf{x}\,;\;\bar{\mathbf{x}},\mathbf{C}) \equiv \frac{1}{(2\pi)^{n/2}|\det\mathbf{C}(\mathbf{x})|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\mathbf{C}^{-1}(\mathbf{x})(\mathbf{x}-\bar{\mathbf{x}})\right\}$$

  - The determinant det C(x) of a covariance matrix is a measure of the volume of the uncertainty ellipsoid associated with the state x estimate.





Maximum log-posterior

- Log-posterior is defined as

$$L(\mathbf{x}) \equiv \log[\rho(\mathbf{x} \mid \text{data})]$$

- If at $\mathbf{x} = \mathbf{x}_0$, we have the maximum log-posterior, then
  - $\nabla L\big|_{\mathbf{x}_0} = 0$ (the gradient, $\nabla = \mathbf{i}\frac{d}{dx} + \mathbf{j}\frac{d}{dy} + \mathbf{k}\frac{d}{dz}$)
  - $\left.\frac{\partial^2 L}{\partial \mathbf{x}^2}\right|_{\mathbf{x}_0}$ (the Hessian) is negative definite
- At $\mathbf{x} = \mathbf{x}_0$, the Taylor series expansion till the quadratic term, and then taking exponentials on both sides gives:

$$\rho(\mathbf{x} \mid \text{data}) \propto \exp\left\{\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \left.\frac{\partial^2 L}{\partial \mathbf{x}^2}\right|_{\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0)\right\}$$

Covariance at maximum log-posterior

$$\rho(\mathbf{x} \mid \text{data}) \propto \exp\left\{\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T \left.\frac{\partial^2 L}{\partial \mathbf{x}^2}\right|_{\mathbf{x}_0} (\mathbf{x}-\mathbf{x}_0)\right\}$$

- What are the mean and covariance matrix of this pdf which is the Gaussian approximation of the original pdf?

$$\mathbf{C}(\mathbf{x}) = -\left(\left.\frac{\partial^2 L}{\partial \mathbf{x}^2}\right|_{\mathbf{x}_0}\right)^{-1}$$

- As the Hessian is negative semidefinite, C(x) is positive semi definite

Example of parameter estimation

- Given a normally distributed data-set $\{x_k , k = 1 \ldots N\}$, find $p(\mu \mid \{x_k\})$ and $p(\sigma \mid \{x_k\})$
- Likelihood of a single sample x_i is

$$p(x_i \mid \mu, \sigma) = \mathcal{N}\left(x_i \,;\, \mu, \sigma^2\right)$$

- The likelihood of the whole data-set is

$$p(\{x_k\} \mid \mu, \sigma) = \prod_{k=1}^{N} p(x_k \mid \mu, \sigma)$$

$$= \left(\sigma\sqrt{2\pi}\right)^{-N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k - \mu)^2\right\}$$

Example of parameter estimation

- Given a normally distributed data-set $\{x_k, k = 1 \ldots N\}$, find $p(\mu \mid \{x_k\})$ and $p(\sigma \mid \{x_k\})$
- Assume uniform prior

$$\rho(\mu, \sigma) = \begin{cases} \text{constant} & \sigma > 0, \\ 0 & \text{otherwise} \end{cases}$$

- Posterior joint pmf

$$\rho(\mu, \sigma \mid \{x_k\}) \propto \rho(\{x_k\} \mid \mu, \sigma)\rho(\mu, \sigma)$$

Example of parameter estimation

- Given a normally distributed data-set $\{x_k, k = 1 \ldots N\}$, find $p(\mu \mid \{x_k\})$ and $p(\sigma \mid \{x_k\})$
- Marginalize out σ from the joint distribution

$$
\begin{aligned}
\rho(\mu \mid \{x_k\}) &= \int_0^{\infty} \rho(\mu, \sigma \mid \{x_k\}) d\sigma \\
&\propto \int_0^{\infty} \sigma^{-N} \exp\left\{ \frac{-1}{2\sigma} \sum_{k=1}^{N} (x_k - \mu)^2 \right\} d\sigma \\
&\propto \left(\sum_{k=1}^{N} (x_k - \mu)^2 \right)^{-(N-1)/2}
\end{aligned}
$$

Example of parameter estimation

- Given a normally distributed data-set $\{x_k , k = 1 \dots N\}$, find $p(\mu \mid \{x_k\})$ and $p(\sigma \mid \{x_k\})$

$$p(\mu \mid \{x_k\}) \propto \left(\sum_{k=1}^{N} (x_k - \mu)^2 \right)^{-(N-1)/2}$$

- This is called the Student-t distribution
  - it arises when estimating the mean of a normally distributed population in situations where the sample size is small and population standard deviation is unknown
  - used in Student's t-test for assessing the statistical significance of the difference between two sample means
  - in construction of confidence intervals for the difference between two population means
  - in linear regression analysis

Example of parameter estimation

- Given a normally distributed data-set $\{x_k, k = 1 \ldots N\}$, find $p(\mu \mid \{x_k\})$ and $p(\sigma \mid \{x_k\})$
- Maximize the log-posterior $L(\mu) = \log \rho(\mu \mid \{x_k\})$
- The maximum occurs at μ_0 such that:

$$\frac{dL}{d\mu} = \frac{(N-1)\sum_{k=1}^{N}(x_k - \mu)}{\sum_{k=1}^{N}(x_k - \mu)^2} = 0, \qquad \mu_0 \equiv \frac{1}{N}\sum_{k=1}^{N} x_k$$

Example of parameter estimation

- Given a normally distributed data-set $\{x_k , k = 1 \dots N\}$, find $p(\mu \mid \{x_k\})$ and $p(\sigma \mid \{x_k\})$

- To estimate the variance (square of standard deviation) of our estimation, we use

$$\mathbf{C}(\mathbf{x}) = -\left(\left. \frac{\partial^2 L}{\partial \mathbf{x}^2} \right|_{\mathbf{x}_0} \right)^{-1}$$

- …and obtain

$$S^2 \equiv \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \mu_0)^2, \qquad \left. \frac{d^2 L}{d\mu^2} \right|_{\mu_0} = -\frac{N}{S^2}$$

➔ Therefore

$$\mu = \mu_0 \pm \frac{S}{\sqrt{N}}$$ (within one-sigma)

Example of parameter estimation

- Given a normally distributed data-set $\{x_k , k = 1 \ldots N\}$, find $p(\mu \mid \{x_k\})$ and $p(\sigma \mid \{x_k\})$

- Similarly, if we marginalize away μ from

$$p(\mu, \sigma \mid \{x_k\}) \propto p(\{x_k\} \mid \mu, \sigma) p(\mu, \sigma)$$

- We obtain:

$$p(\sigma \mid \{x_k\}) \propto \frac{1}{\sigma^{N-1}} \exp\left\{ -\frac{\sum_{k=1}^{N}(x_k - \mu_0)^2}{2\sigma^2} \right\}$$

- On maximizing the log-posterior as before, (after many steps), we get the best estimate of σ as $\sigma_0 = S$, where, S was defined earlier. The one-sigma interval for σ is

$$\sigma = \sigma_0 \pm \frac{\sigma_0}{\sqrt{2(N-1)}}$$