TTK4250

Lecture 2

The multivariate Gaussian and the Kalman filter

Edmund Førland Brekke

26. August 2019

Outline

1. The multivariate Gaussian
   - Definition and key concepts
   - Useful rules
   - The product identity
2. State estimation and the Kalman filter
   - General concepts in state estimation
   - From product identity to Kalman filter
   - Example run of a Kalman filter

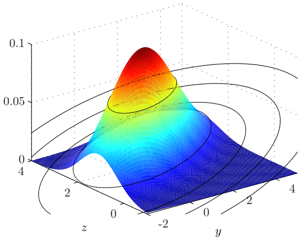
The multivariate Gaussian

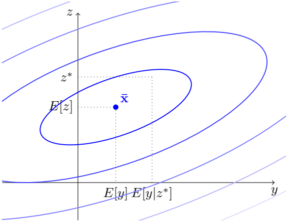
This is the probability distribution given by

$$\mathcal{N}(\mathbf{x}\,;\,\boldsymbol{\mu},\mathbf{P}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{P}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

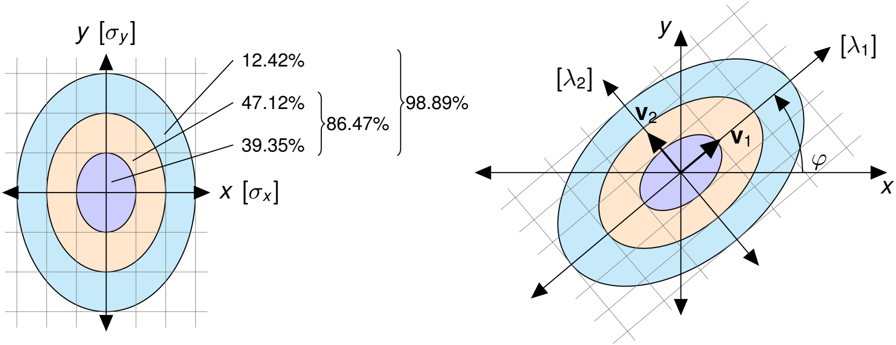
where $\boldsymbol{\mu}$ is the expectation of $\mathbf{x}$ and $\mathbf{P}$ is the covariance of $\mathbf{x}$.

- The matrix $\mathbf{P}$ must be symmetric positive definite (SPD).
- All the dependence on $\mathbf{x}$ is encapsulated by a **quadratic form**.





The shape and probability mass of covariance ellipses

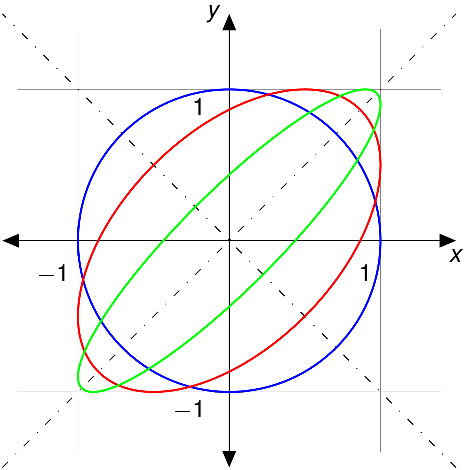


- The probability that $\mathbf{x}$ is within the ellipse $(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = g$ is given by

$$\texttt{chi2cdf}(g^2, n)$$

- The shape of the ellipses is given by the eigenvectors and eigenvalues of $\mathbf{P}$.

The role of correlations



$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \right)$$

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$

$a = 0.0$

$a = 0.5$

$a = 0.9$

- Correlations make the covariance ellipses narrower.

$\Rightarrow$ Correlations can be exploited to achieve accurate state estimation.

Key rules: Independence and Linearity

Independence

Two random vectors $\mathbf{x}$ and $\mathbf{y}$ with probability density functions $\mathcal{N}(\mathbf{x}\,;\,\mathbf{a},\mathbf{A})$ and $\mathcal{N}(\mathbf{y}\,;\,\mathbf{b},\mathbf{B})$ are independent if and only if

$$p(\mathbf{x},\mathbf{y}) = p\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix};\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix},\begin{bmatrix}\mathbf{A} & 0\\ 0 & \mathbf{B}\end{bmatrix}\right).$$

Linearity

If $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}\,;\,\mathbf{a},\mathbf{A})$ and $\mathbf{y} = \mathbf{F}\mathbf{x}$, then $p(\mathbf{y}) = \mathcal{N}(\mathbf{y}\,;\,\mathbf{F}\mathbf{a},\mathbf{F}\mathbf{A}\mathbf{F}^{\mathsf{T}})$.

Example: Cholesky factorization

Let $\mathbf{x} \sim \mathcal{N}(\mathbf{0},\mathbf{A})$. Since $\mathbf{A}$ is symmetric positive definite it has a Cholesky factorization $\mathbf{L}$ so that $\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathsf{T}}$. Define the transformed RV $\mathbf{y} = \mathbf{L}^{-1}\mathbf{x}$. The expectation of $\mathbf{y}$ is then obviously $\mathbf{0}$ and its covariance is

$$\mathrm{Cov}[\mathbf{y}] = \mathbf{L}^{-1}\mathbf{A}(\mathbf{L}^{-1})^{\mathsf{T}} = \mathbf{L}^{-1}\mathbf{L}\mathbf{L}^{\mathsf{T}}(\mathbf{L}^{-1})^{\mathsf{T}} = \mathbf{I}.$$

Key rules: Marginalization and conditioning

Let $\mathbf{x}$ and $\mathbf{y}$ have the joint distribution

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{xx} & \mathbf{P}_{xy} \\ \mathbf{P}_{xy}^{\mathsf{T}} & \mathbf{P}_{yy} \end{bmatrix} \right)$$

Marginalization

The marginal distribution of $\mathbf{y}$ is $p(\mathbf{y}) = \mathcal{N}(\mathbf{y}\,;\, \mathbf{b}, \mathbf{P}_{yy})$.

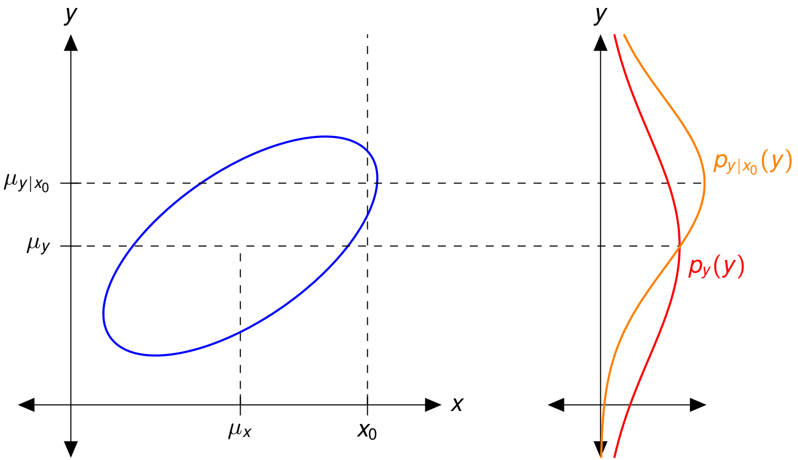
Conditioning

The conditional distribution pf $\mathbf{x}$ given $\mathbf{y}$ is $p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}\,;\, \boldsymbol{\mu}_{x|y}, \mathbf{P}_{x|y})$ where

$$\boldsymbol{\mu}_{x|y} = \mathbf{a} + \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}(\mathbf{y} - \mathbf{b}) \qquad \text{and} \qquad \mathbf{P}_{x|y} = \mathbf{P}_{xx} - \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}\mathbf{P}_{xy}^{\mathsf{T}}.$$

Notice that the formulas for conditioning are very similar to the formulas we derived for the LMMSE estimator.

Illustration: Marginalization and conditioning



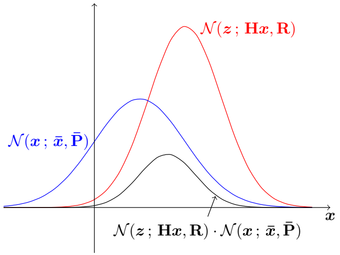
Key rules: The product identity

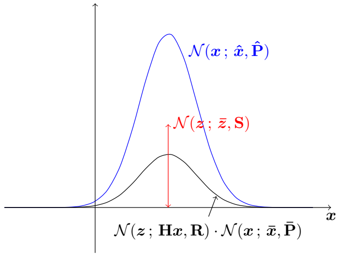
The product of a Gaussian in **x** with a Gaussian that depends linearly on **x** is proportional to another Gaussian in **x**:

$$\mathcal{N}(\mathbf{z}\,;\,\mathbf{Hx},\mathbf{R})\mathcal{N}(\mathbf{x}\,;\,\bar{\mathbf{x}},\bar{\mathbf{P}}) = \mathcal{N}(\mathbf{z}\,;\,\bar{\mathbf{z}},\mathbf{S})\mathcal{N}(\mathbf{x}\,;\,\hat{\mathbf{x}},\hat{\mathbf{P}}).$$

The Gaussians on the right-hand side are given by

$$\begin{aligned}\bar{\mathbf{z}} &= \mathbf{H}\bar{\mathbf{x}} \\ \mathbf{S} &= \mathbf{R} + \mathbf{H}\bar{\mathbf{P}}\mathbf{H}^{\mathsf{T}}\end{aligned} \qquad \begin{aligned}\hat{\mathbf{x}} &= \bar{\mathbf{x}} + \mathbf{W}(\mathbf{z} - \mathbf{H}\bar{\mathbf{x}}) \\ \hat{\mathbf{P}} &= (\mathbf{I} - \mathbf{WH})\bar{\mathbf{P}} \\ \mathbf{W} &= \bar{\mathbf{P}}\mathbf{H}^{\mathsf{T}}\mathbf{S}^{-1}.\end{aligned}$$





Overview of a proof of the product identity

If we assume that the rules for conditioning and marginalization are proved, we can prove the product identity in the following three steps:[1]

1. We construct a joint Gaussian over $\mathbf{z}$ and $\mathbf{x}$ which can be factorized in two manners:

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z}|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z}) \tag{1}$$

   We define $p(\mathbf{z}, \mathbf{x})$ by letting the first factorization in (1) be identical to the left-hand-side of the product identity.

2. The quadratic form in $p(\mathbf{z}, \mathbf{x})$ will then be a sum of two contributions from $p(\mathbf{z}|\mathbf{x})$ and $p(\mathbf{x})$. We manipulate this sum so that it becomes a single quadratic form describing $p(\mathbf{z}, \mathbf{x})$ as a Gaussian in the stacked vector $[\mathbf{z}^\mathsf{T}, \mathbf{x}^\mathsf{T}]^\mathsf{T}$.
3. We obtain the second factorization in (1) by means of the conditioning and marginalization rules. This factorization is identical to the right-hand-side of the product identity.

---

[1] Based on L.-C. Tokle (2018): "Multi target tracking using random finite sets with a hybrid state space and approximations."

Probabilistic state estimation

- So far we have only discussed estimation in static systems.
- In the remainder of the course we want to do estimation in **dynamic systems**.

Continuous time vs discrete time.

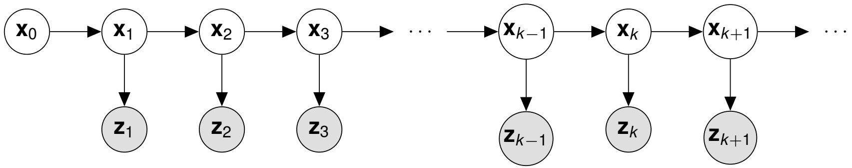
- Continuous time: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{v}), \;\; \mathbf{z} = \mathbf{h}(\mathbf{x}, \mathbf{w})$.
  - Often most closely related to the underlying physics.
  - Conceptually challenging (continuous-time white noise is a mathematical abstraction).
  - Impossible to implement on a computer.
- Discrete time: $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{v}_k), \;\; \mathbf{z} = \mathbf{h}(\mathbf{x}, \mathbf{w})$.

Recursive vs batch.

- In **recursive estimation (filtering),** we go through the steps as new data arrive:
  - Given some information about $\mathbf{x}_{k-1}$ ...
  - we **predict** $\mathbf{x}_k$ ...
  - we adjust our prediction of $\mathbf{x}_k$ based on the data $\mathbf{z}_k$ ...
  - and so on.
- In **batch estimation,** we estimate all the state variables $\mathbf{x}_{1:k} = [\mathbf{x}_1; \ldots; \mathbf{x}_k]$ simultaneously.
- There is also **smoothing,** where one filters both forward and backwards in time, in order to exploit future data to improve past estimates.

Recursive Bayesian estimation: Model and key concepts

We study systems whose structure fits the **graphical model** below:



- The horizonal arrows represent a **process model** of the form $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$
- The horizonal arrows represent a **measurement model** of the form $p(\mathbf{z}_k \mid \mathbf{x}_k)$.

This structure reflects the following **Markov assumptions**

$$p(\mathbf{x}_k \mid \mathbf{x}_1, \ldots, \mathbf{x}_{k-2}, \mathbf{x}_{k-1}, \mathbf{z}_1, \ldots, \mathbf{z}_{k-2}, \mathbf{z}_{k-1}) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \tag{2}$$

$$p(\mathbf{z}_k \mid \mathbf{x}_1, \ldots, \mathbf{x}_{k-2}, \mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{z}_1, \ldots, \mathbf{z}_{k-2}, \mathbf{z}_{k-1}) = p(\mathbf{z}_k \mid \mathbf{x}_k). \tag{3}$$

Recursive Bayesian estimation: The Bayes filter

In the Bayesian philosophy we want a pdf as our solution. This pdf may or may not be given by parameters such as expectation, covariance etc.

What do we know about $\mathbf{x}_k$ after observing $\mathbf{z}_{1:k} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k)$?

- The total probability theorem yields the predicted density

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

- Bayes' rule yields the posterior density

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1})} \propto p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1}).$$

Remark: Violations of the Markov assumptions can be handled by replacing the Markov chain by a higher order Markov chain that models the temporal correlations. We must then extend the state vector with corresponding states.

Linearity, Gaussianity and the Kalman filter

"Everything should be made as simple as possible, but not simpler."

- In general, we cannot find a closed-form solution to the Bayes filter.
- If the posterior can be described with reasonable accuracy by a few parameters (e.g., expectation and covariance), then we should look for a compact representation.

Closed-form solution to the Bayes filter = Kalman filter

When does a closed-form solution to the Bayes filter exist?

- When the initial density is Gaussian $\mathcal{N}(\mathbf{x}_0\,;\,\hat{\mathbf{x}}_0, \mathbf{P}_0)$
- . . . and the Markov model is Gaussian-linear $\mathcal{N}(\mathbf{x}_k\,;\,\mathbf{F}\mathbf{x}_{k-1}, Q)$
- . . . and the likelihood is Gaussian-linear $\mathcal{N}(\mathbf{z}_k\,;\,\mathbf{H}\mathbf{x}_k, \mathbf{R})$
- . . . and standard independence assumptions apply.

The prediction step of the Kalman filter

The predicted density is given by

$$
\begin{aligned}
p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})\mathrm{d}\mathbf{x}_{k-1} \\
&= \int \mathcal{N}(\mathbf{x}_k\,;\,\mathbf{F}\mathbf{x}_{k-1}, Q)\mathcal{N}(\mathbf{x}_{k-1}\,;\,\hat{\mathbf{x}}_{k-1}, \mathbf{P}_{k-1})\mathrm{d}\mathbf{x}_{k-1} \\
&= \mathcal{N}(\mathbf{x}_k\,;\,\mathbf{F}\hat{\mathbf{x}}_{k-1}, \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^{\mathsf{T}} + Q) \\
&\quad \cdot \int \mathcal{N}(\mathbf{x}_{k-1}\,;\,\text{some vector}\,,\ \text{some covariance matrix}\,)\mathrm{d}\mathbf{x}_{k-1} \\
&= \mathcal{N}(\mathbf{x}_k\,;\,\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}).
\end{aligned}
$$

- $\hat{\mathbf{x}}_{k-1}$ is the previous state estimate.
- $\mathbf{P}_{k-1}$ is the previous covariance.
- $\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1}$ is the predicted state estimate.
- $\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^{\mathsf{T}} + Q$ is the predicted covariance.

The update step of the Kalman filter

The posterior density is given by

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{z}_{1:k}) &\propto p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \\ &= \mathcal{N}(\mathbf{z}_k\,;\,\mathbf{H}\mathbf{x}_k, \mathbf{R})\,\mathcal{N}(\mathbf{x}_k\,;\,\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \\ &= \mathcal{N}(\mathbf{z}_k\,;\,\mathbf{H}\hat{\mathbf{x}}_{k|k-1}, \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}} + \mathbf{R})\mathcal{N}(\mathbf{x}_k\,;\,\hat{\mathbf{x}}_k, \mathbf{P}_k) \\ &\propto \mathcal{N}(\mathbf{x}_k\,;\,\hat{\mathbf{x}}_k, \mathbf{P}_k). \end{aligned}$$

- $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1})$ is the posterior state estimate.
- $\mathbf{P}_k = (\mathbf{I} - \mathbf{W}_k\mathbf{H})\mathbf{P}_{k|k-1}$ is the posterior covariance.
- $\mathbf{W}_k = \mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}}(\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^{\mathsf{T}} + \mathbf{R})^{-1}$ is the Kalman gain.

More about the covariance

Joseph form

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{W}_k\mathbf{H})\mathbf{P}_{k|k-1}(\mathbf{I} - \mathbf{W}_k\mathbf{H})^\mathsf{T} + \mathbf{W}\mathbf{R}\mathbf{W}^\mathsf{T}$$

Information form

$$\mathbf{P}_k^{-1} = \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}_{k|k-1}^{-1}$$

Orthogonality properties

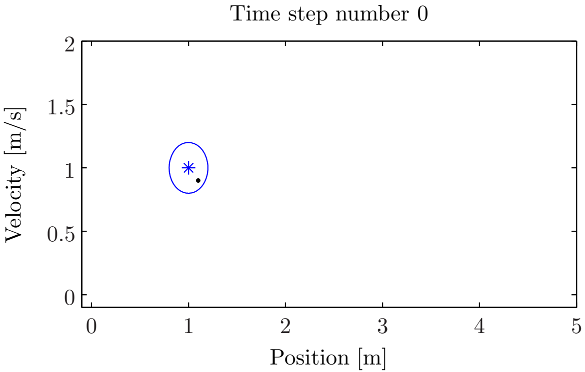
- The estimation errors $\tilde{\mathbf{x}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$ do not constitute a white sequence:

$$E[\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_{k-1}^\mathsf{T}] = (\mathbf{I} - \mathbf{W}_k\mathbf{H})\mathbf{F}\mathbf{P}_k.$$

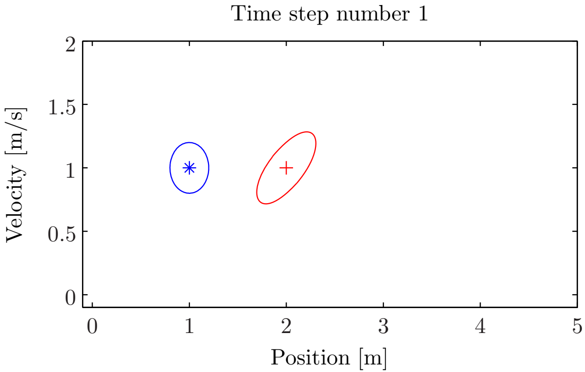
- The innovations on the other hand are a white sequence:

$$E[\boldsymbol{\nu}_k\boldsymbol{\nu}_j^\mathsf{T}] = \mathbf{0} \text{ if } k \neq j \quad \Leftrightarrow \quad p(\mathbf{z}_{1:k}) = \prod_{j=1}^{k} p(\boldsymbol{\nu}_j).$$

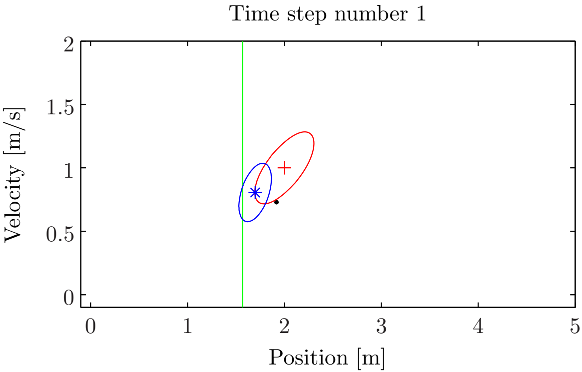
Example run of the Kalman filter



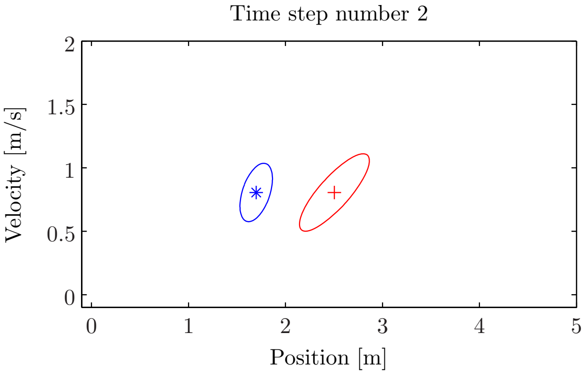
Example run of the Kalman filter



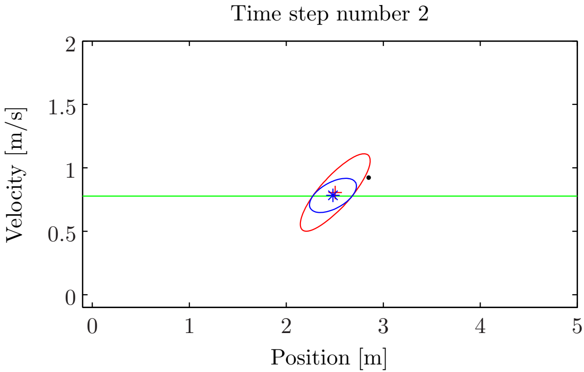
Example run of the Kalman filter



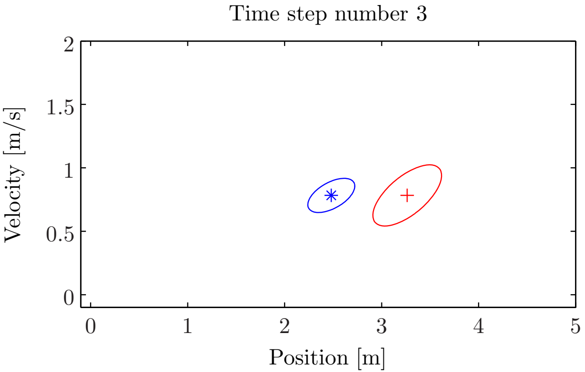
Example run of the Kalman filter



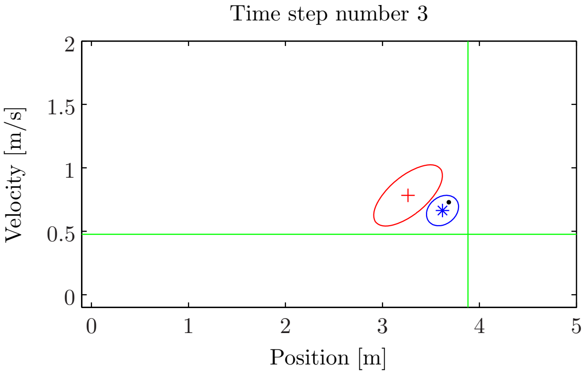
Example run of the Kalman filter



Example run of the Kalman filter

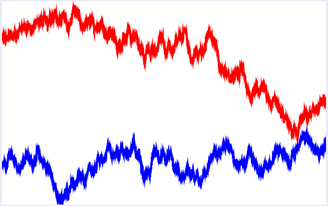


Example run of the Kalman filter

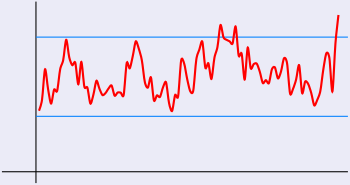


The road ahead

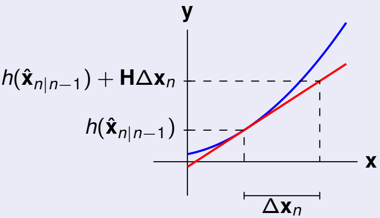
Stochastic processes



Tuning of the Kalman filter



The EKF



Particle filters