TTK4250

Week 3

From the Kalman filter to stochastic processes

Edmund Førland Brekke

5. September 2024

Recap from last week

An estimator is a random variable

- . . . because it depends on the (random) data.
- We can talk about its distribution, expectation and covariance.
- An estimator is unbiased if $E[\mathbf{x} - \hat{\mathbf{x}}] = \mathbf{0}$.

LMMSE estimation

$$\hat{\mathbf{x}} = E[\mathbf{x}] + \mathrm{Cov}(\mathbf{x}, \mathbf{z})\mathrm{Cov}(\mathbf{z})^{-1}(\mathbf{z} - E[\mathbf{z}])$$

is the estimator of the form $\hat{\mathbf{x}} = \mathbf{A}\mathbf{z} + \mathbf{b}$ that minimizes

$$\mathrm{MSE}(\hat{\mathbf{x}}) = E\left[\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2\right].$$

The multivariate Gaussian

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

- Quadratic forms.
- Moment parametrization vs canonical parametrization.

Marginalization and conditioning

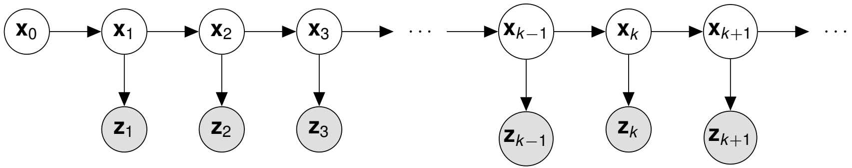
In moment parametrization, conditioning is given by

$$\begin{aligned}\boldsymbol{\mu}_{x|y} &= \mathbf{a} + \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}(\mathbf{y} - \mathbf{b})\\ \mathbf{P}_{x|y} &= \mathbf{P}_{xx} - \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}\mathbf{P}_{xy}^\mathsf{T}.\end{aligned}$$

This leads to the Product Identity.

Recursive Bayesian estimation: Model and key concepts

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



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

This structure reflects the following **Markov assumptions**

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

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

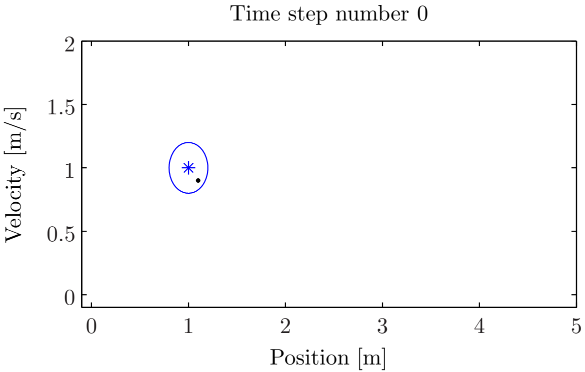
- The estimation errors $\tilde{\mathbf{x}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$ do not constitute a white sequence:

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

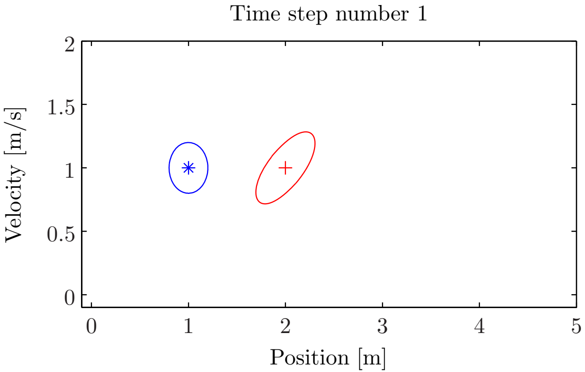
- The innovations $\boldsymbol{\nu}_k = \mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}$ on the other hand are a white sequence:

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

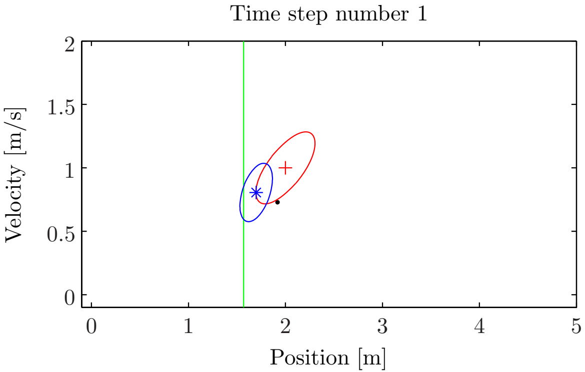
Example run of the Kalman filter



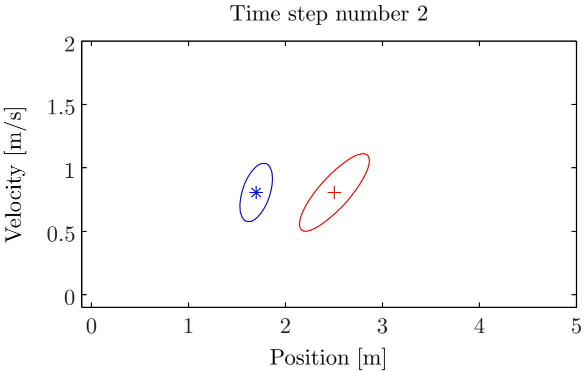
Example run of the Kalman filter



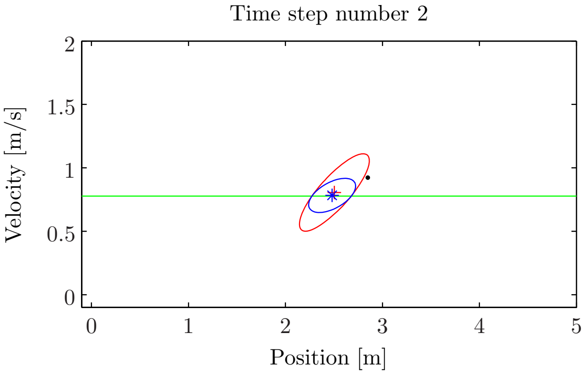
Example run of the Kalman filter



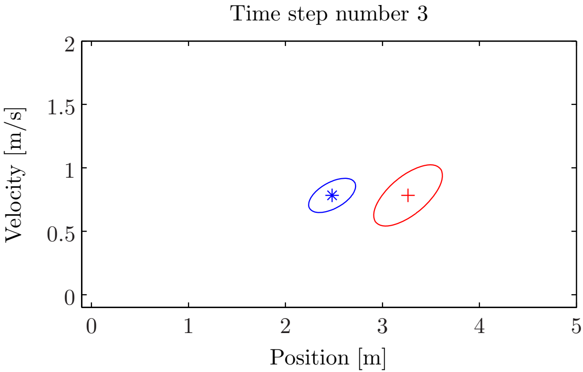
Example run of the Kalman filter



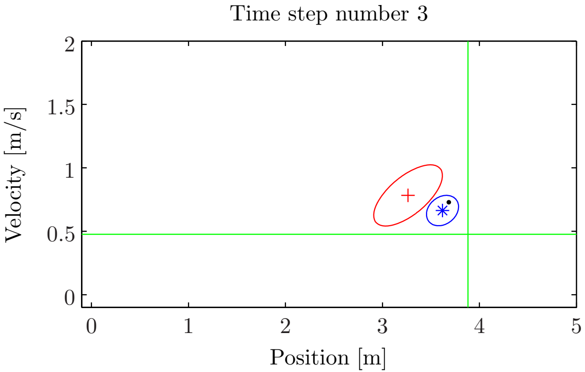
Example run of the Kalman filter



Example run of the Kalman filter



Example run of the Kalman filter



The tuning problem

Given the discrete time model

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{u}_k + \mathbf{v}_k, \;\; \mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k, \;\; \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \;\; \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

we must determine the values of the noise matrices **Q** and **R** that . . .

- faithfully represent the uncertainties of the process and measurement models.
- give the Kalman filter optimal **accuracy** and **robustness**.

The process noise covariance

- The matrix **Q** says something about how the system is expected to evolve between two time steps.
- But the system dynamics are generally modeled in continuous time.
- Therefore we need to relate **Q** to a continuous-time model of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbf{n}$$

The measurement noise covariance

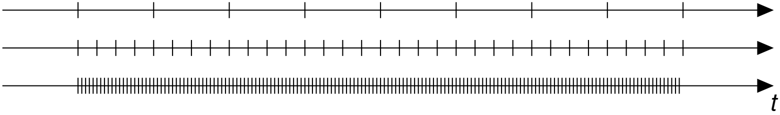
- The matrix **R** says something about how accurate our measurement devices (sensors) are.
- This is fully encapsulated by the discrete-time model.

Stochastic processes

- Consider a stochastic vector

$$\mathbf{x} = \begin{bmatrix} x(t_1) & x(t_2) & \dots & x(t_n) \end{bmatrix}^\mathsf{T}$$

  where $x(t_k)$ is the value of the stochastic variable x at time t_k.
- Let the discretization length $T = t_k - t_{k-1}$ go towards zero.
- Every realization of $\mathbf{x}$ will then be equivalent to a **function** $x(t)$. Such a random function is known as a **stochastic process**.

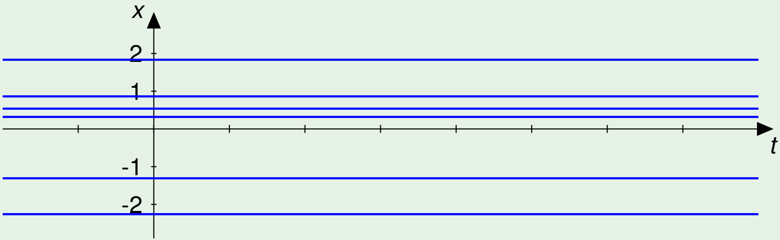


- To fully specify $x(t)$ in the general case we would need the joint distribution of all tuples $x(t_1)\dots, x(t_k)$ for any number k.
- We restrict our attention to stochastic process which can be defined in terms of their construction or in terms of first- and second-order moments.

A very simple stochastic process

A random constant

Let the function $x(t)$ be given by $x(t) = a$ where $a \sim \mathcal{N}(0, 1)$. Different realizations of this stochastic process can be depicted as follows:



Any number that depends on $x(t)$, such as a time integral of $x(t)$, will be a random variable. Let

$$y = \int_0^t x(\tau)\mathrm{d}\tau.$$

Then it can be shown that $y \sim \mathcal{N}(0, t^2)$.

A not so simple stochastic process

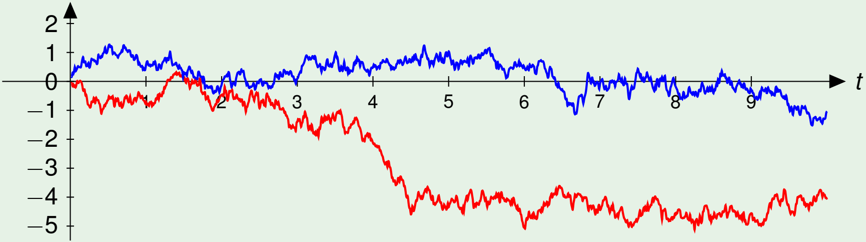
The Wiener process

- Define the stochastic process $x(t)$ by

$$x(nT) = \sum_{i=1}^{n} x_i \quad \text{where} \quad x_i \sim \mathcal{N}(0, T) \quad \text{i.i.d.}$$

- Then we define the Wiener process $b(t)$ as the limit

$$b(t) = \lim_{T \to 0} x(t).$$



More about the Wiener process

Alternative definition

Mathematicians like to define the Wiener process in terms of 4 fundamental properties:

1. $b(0) = 0$.
2. $b(t)$ has independent increments. That is: If $t_1 < t_2$, then $b(t_2) - b(t_1)$ is independent of the past values $b(s)$ for $s < t_1$.
3. $b(t)$ has Gaussian increments: If $t_1 < t_2$ then $b(t_2) - b(t_1) \sim \mathcal{N}(0, t_2 - t_1)$
4. $b(t)$ is continuous in t.

Statistics of the Wiener process

- The expectation of the Wiener process is always 0.
- The variance of the Wiener process at any particular time is

$$E[b(t)^2] = \mathrm{Var}\left[\sum_{i=1}^{n} x_i\right] = nT = \frac{t}{T}T = t.$$

White Gaussian noise

We define **continuous-time white Gaussian noise** as the derivative of the Wiener process

$$n(t) = \lim_{\Delta \to 0} \frac{b(t+\Delta) - b(t)}{\Delta}.$$

- We always use white noise as a driving mechanism in stochastic continuous time models.
- For this to make sense, the contributions from a white noise process over a limited time interval must be finite and non-zero:

$$\Rightarrow \quad 0 < \mathrm{Var}\left[\int_0^s n(t)\mathrm{d}t\right] < \infty.$$

- Making matters complicated, this requirement in turn implies that

$$\mathrm{Var}\left[n(t)\right] = \infty.$$

White noise is a mathematical abstraction because it has infinite energy.

The autocorrelation function

Motivation: We want to have a useful description of important stochastic processes such as white Gaussian noise and its relatives.

Definition: Autocorrelation function (ACF)

The ACF of a stochastic process $\mathbf{x}(t)$ is $R(t_1, t_2) = E[\mathbf{x}(t_1)\mathbf{x}(t_2)^\mathsf{T}]$.

Definition: Wide-sense stationarity

A stochastic process $\mathbf{x}(t)$ is said to be wide-sense stationary if its expectation is constant and its ACF can be written as a function of $\tau = t_2 - t_1$:

$$R(\tau) = E[\mathbf{x}(t)\mathbf{x}(t+\tau)^\mathsf{T}].$$

Example: ACF of white Gaussian noise

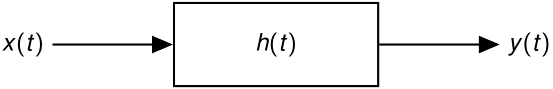
The ACF of the white noise process defined on the previous slide is

$$R(\tau) = \delta(\tau).$$

See the proof of Theorem 4.3.2 in the book for a derivation of this result.

Stochastic linear systems

What happens to white noise (or any other stochastic process) when it is used as input to a system with a given impulse response?



Convolution formulas for the ACF

Let $x(t)$ be a scalar real-valued stochastic process with ACF $R_{xx}(t_1, t_2)$ and let

$$y(t) = \int_{-\infty}^{\infty} h(t-\alpha)x(\alpha)\mathrm{d}\alpha$$

where $h(t)$ also is scalar real-valued. The the ACF of $y(t)$ is given by

$$R_{xy}(t_1, t_2) = \int_{-\infty}^{\infty} R_{xx}(t_1, t_2 - \alpha)h(\alpha)\mathrm{d}\alpha$$

$$R_{yy}(t_1, t_2) = \int_{-\infty}^{\infty} R_{xy}(t_1 - \alpha, t_2)h(\alpha)\mathrm{d}\alpha.$$

The Gauss-Markov process

- Consider a system with impulse response $h(t) = e^{-ct}u(t)$.
- We send white noise $n(t)$ into the system, starting at $t = 0$.
- What is then the ACF of the output?

$$
\begin{aligned}
R_{xy}(t_1, t_2) =& q e^{-c(t_2 - t_1)} u(t_1) u(t_2 - t_1) \\
R_{yy}(t_1, t_2) =& \frac{q}{2c}(1 - e^{-2ct_1}) e^{-c(t_2 - t_1)}.
\end{aligned}
$$

- The formulas are valid if $0 < t_1 < t_2$.
- In the limit as $t_1 \to \infty$ the Gauss-Markov process becomes a stationary process with ACF

$$
\frac{q}{2c} e^{-c|t_2 - t_1|}.
$$

Continuous time modeling: Accelerometer with bias

The Gauss-Markov process can be used to model a slowly varying accelerometer bias.

- Let the state vector be

$$\mathbf{x} = \begin{bmatrix} \text{Position of the vehicle} \\ \text{Velocity of the vehicle} \\ \text{Bias of the accelerometer} \end{bmatrix}$$

- The system is described by a state-space model of the form $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbf{n}$ where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -c \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

  and where

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}\delta(t-\tau)) \text{ where } \mathbf{D} = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}.$$

- Notice that the accelerometer readings are treated as a control input and not as measurements.

Continuous time modeling: The CV model in 2 dimensions

- This is perhaps the most common model used in sensor fusion.
- The model is of the form $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{G}\mathbf{n}$ where the matrices are given by

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

  and the process noise is given by

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}\delta(t-\tau)) \quad \text{where} \quad \mathbf{D} = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_a^2 \end{bmatrix}.$$

- We see that the process noise strength is solely given by the number σ_a, which is a measure of root-mean-square acceleration.
- Since the model essentially integrates white noise the two positional states become independent Wiener processes.

Discretization

Consider the linear continuous-time state space model

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}\delta(t-\tau)).$$

A discrete time solution can be written

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{u}_k + \mathbf{v}_k$$

where

$$\mathbf{F} = e^{\mathbf{A}(t_k - t_{k-1})}, \quad \mathbf{u}_k = \int_{t_{k-1}}^{t_k} e^{\mathbf{A}(t_k-\tau)}\mathbf{B}\mathbf{u}(\tau)\,\mathrm{d}\tau \quad \text{and} \quad \mathbf{v}_k = \int_{t_{k-1}}^{t_k} e^{\mathbf{A}(t_k-\tau)}\mathbf{G}\mathbf{n}(\tau)\,\mathrm{d}\tau.$$

From continuous to discrete time process noise covariance

Let the discretization time be fixed at $T = t_k - t_{k-1}$. The covariance matrix of $\mathbf{v}_k$ in the discrete-time model is then given by

$$\mathbf{Q} = E[\mathbf{v}_k\mathbf{v}_k^\mathsf{T}] = \int_0^T e^{(T-\tau)\mathbf{A}}\mathbf{G}\mathbf{D}\mathbf{G}^\mathsf{T} e^{(T-\tau)\mathbf{A}^\mathsf{T}}\,\mathrm{d}\tau$$

Evaluating the discrete-time process noise matrix

First/nth order approximations

- The simplest possible approximation is $\mathbf{Q} \approx \mathbf{GDG}^{\mathsf{T}} T$.
- Not recommended because it may be singular even if the true $\mathbf{Q}$ is SPD.
- Higher-order approximations are often used.
- Notice the linear dependence on the sampling interval T. This reflects the fact that the variance of the Wiener process increases linearly with time.

Van Loan's formula

Define the matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ according to

$$\exp\left(\begin{bmatrix} -\mathbf{A} & \mathbf{GDG}^{\mathsf{T}} \\ \mathbf{0} & \mathbf{A}^{\mathsf{T}} \end{bmatrix} T\right) = \begin{bmatrix} \times & \mathbf{V}_2 \\ \mathbf{0} & \mathbf{V}_1 \end{bmatrix}.$$

Then we can find $\mathbf{Q}$ according to $\mathbf{Q} = \mathbf{V}_1^{\mathsf{T}} \mathbf{V}_2$.

How to tune the process noise covariance

There are at least 3 approaches that can be used to determine suitable values for the process noise covariance.

Purely physical considerations

- In a CV model we may use the largest accelerations observed as a guideline for how large σ_a needs to be.
- In the accelerometer model we can calculate σ_a as a continuous-time equivalent of of the specified accuracy of the accelerometer.

Consistency analysis

Set the process noise as high as required to make the data or state estimates **plausible**.

Maximum likelihood estimation

Find the **most likely** value of the process noise strength given the data.

The process noise strength should not depend on the measurement model.

The concept of filter consistency

A filter is said to be consistent if

1. The state errors are acceptable as zero mean.
2. The state errors have magnitude commensurate with the state covariance yielded by the filter.
3. The innovations are acceptable as zero mean.
4. The innovations have magnitude commensurate with the innovation covariance yielded by the filter.
5. The innovations are acceptable as white.

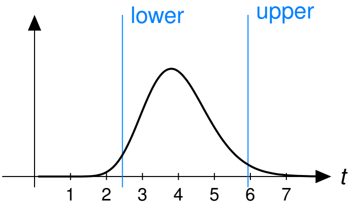
Criteria 2 and 4 are most important, and are tested by means of the normalized estimation error squared (NEES) and the normalized innovations squared (NIS):

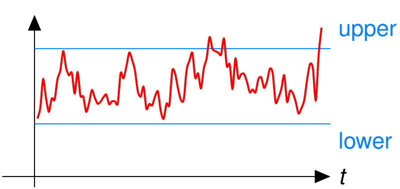
$$\begin{aligned}\epsilon_k &= (\hat{\mathbf{x}}_k - \mathbf{x}_k)^{\mathsf{T}} \mathbf{P}_k^{-1} (\hat{\mathbf{x}}_k - \mathbf{x}_k) \\ \epsilon_k^{\nu} &= \boldsymbol{\nu}_k^{\mathsf{T}} \mathbf{S}_k^{-1} \boldsymbol{\nu}_k = (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^{\mathsf{T}} \mathbf{S}_k^{-1} (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}).\end{aligned}$$

χ^2-test for filter consistency (NEES)

- Suppose that $\mathbf{x} \in \mathbb{R}^d$.
- We perform N Monte-Carlo simulations of our Kalman filter.
- Then a 95% confidence interval for the average (ANEES) value of ϵ_k is given by

$$\texttt{lower} = \texttt{chi2inv}(0.025, Nd)/N \quad \text{and } \texttt{upper} = \texttt{chi2inv}(0.975, Nd)/N$$





- If ANEES lies below the χ^2 limits we can expect the filter to be overly conservative.
- If ANEES lies above the χ^2 limits we can expect the filter to be overconfident and put too little emphasis on the measurements.

Use NIS in a similar manner if working on real data without ground truth.

Example of consistency-based tuning

Finding a suitable value for σ_a in the CV model originally used in the radar-based Autosea tracker.[1]

Table 2 Process noise evaluation via AIS filter consistency. The (r_1, r_2) interval is the two-sided 95% probability concentration region for the χ^2 distribution related to the corresponding NIS. This varies with according to the AIS data record length N. The NIS values that are closest to being covariance-consistent, i.e. closest to the 95% probability region, are emphasised in bold

Name	$\sigma_a = 0.05$		$\sigma_a = 0.5$		(r_1, r_2)	N
	NIS	AI	NIS	AI		
GLUTRA	**4.67**	−0.02	0.90	0.01	(3.47, 4.55)	109
SULA	**3.61**	−0.22	0.51	−0.10	(3.49, 4.56)	106
KORSFJORD	71.8	−1.33	**4.31**	−0.44	(3.52, 4.51)	127
TR.FJORD II	11.3	−0.62	**3.24**	−0.16	(3.76, 4.24)	533
TELEMETRON	371	−0.04	**4.45**	−0.01	(3.77, 4.23)	579

[1] Wilthil et al. (2017): "A target tracking system for ASV collision avoidance based on the PDAF", Springer.

Testing the whiteness of the innovations

Whiteness test in Monte-Carlo simulations

Let ν_k be one of the innovation states in the vector $\boldsymbol{\nu}_k$. Let N be the number of Monte-Carlo simulations. Then the distribution of the **sample autocorrelation**

$$\rho_{kj} = \frac{\sum_{i=1}^{N} \nu_k^{(i)} \nu_j^{(i)}}{\sqrt{\sum_{i=1}^{N} (\nu_k^{(i)})^2 \sum_{i=1}^{N} (\nu_j^{(i)})^2}}$$

should tend to $\mathcal{N}(0, 1/N)$ for all $k \neq j$ when N is large.

Single-run whiteness test

The variance of the **time-average autocorrelation** should tend towards $1/K$:

$$\bar{\rho}_j = \frac{\sum_{k=1}^{K} \nu_k \nu_{k+j}}{\sqrt{\sum_{k=1}^{K} \nu_k^2 \sum_{k=1}^{K} \nu_{k+j}^2}}$$

Tuning the measurement noise covariance

For exteroceptive sensors, the appropriate values in **R** depend on

- The sensor resolution.
- The extent of targets or landmarks.

Example: Point targets with pixellated sensor

- 2-dimensional sensor with square cells of fixed resolution Δx.
- $\mathbf{x} = [x, y, v_x, v_y]^\mathsf{T}$.
- Measurement matrix $\mathbf{H} = [\mathbf{I}_2, \mathbf{0}]$

$$p(\mathbf{z} - \mathbf{H}\mathbf{x} | \mathbf{x}) = \begin{cases} 1/\Delta x^2 & \text{if } \|\mathbf{z} - \mathbf{H}\mathbf{x}\|_\infty < \Delta x/2 \\ 0 & \text{otherwise.} \end{cases}$$

This distribution can be approximated by a Gaussian

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}\,;\, \mathbf{H}\mathbf{x}, \mathbf{R})$$

with the same covariance:

$$\mathbf{R} = \begin{bmatrix} \frac{\Delta x^2}{12} & 0 \\ 0 & \frac{\Delta x^2}{12} \end{bmatrix}.$$

More about the measurement model

Mild nonlinearities in the measurement model can sometimes be removed by converting the measurements. We must then also convert **R** accordingly.

Nonlinear model.

The model is of the form $\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{w}_k$, $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ where

$$\mathbf{h}(\mathbf{x}_k) = \begin{bmatrix} \sqrt{x^2 + y^2} \\ \operatorname{atan2}(y, x) \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix}$$

Converted measurements.

The model is of the form $\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k$, $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_c)$ where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{R}_c = \mathbf{J}\mathbf{R}\mathbf{J}^\mathsf{T}, \quad \mathbf{J} = \frac{\partial}{\partial \mathbf{z}_k}\mathbf{h}^{-1}(\mathbf{z}_k)$$

For more sophisticated conversion techniques see Lerro & Bar-Shalom (1993): “Tracking with debiased consistent converted measurements versus EKF”, IEEE-TAES.

Measurement conversion is generally preferable to EKF techniques because the linearization in an EKF can lead to filter instability.

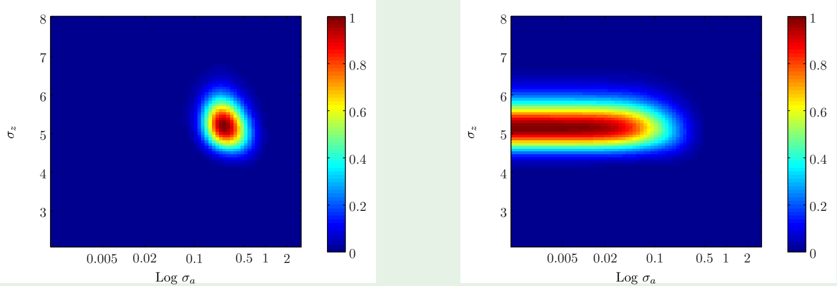
Maximum likelihood estimation of system parameters

Theorem for maximum likelihood estimation

Assume that both **Q** and **R** depend on an unknown parameter vector $\boldsymbol{q}$. The maximum likelihood estimate of $\boldsymbol{q}$, if it exists, can then be found as

$$\boldsymbol{q}_{\mathrm{ML}} = \arg\max_{\boldsymbol{q}} \sum_k \log \mathcal{N}(\mathbf{z}_k\,;\ \mathbf{H}\hat{\mathbf{x}}_{k|k-1}, \mathbf{S}_k)$$

Example: Estimating σ_a and σ_z for CV model



The likelihood function, normalized relative to its maximum, for two realizations of the CV model.

LTV and LTI systems

We point out the main distinction in the continuous case. The discrete case is similar.

Linear time-variant systems

The system can be of the form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u} + \mathbf{G}(t)\mathbf{n} \\ \mathbf{z} &= \mathbf{H}(t)\mathbf{x} + \mathbf{w}\end{aligned}$$

The system matrices are allowed to depend on t.

Linear time-invariant systems

The system can be of the form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbf{n} \\ \mathbf{z} &= \mathbf{H}\mathbf{x} + \mathbf{w}\end{aligned}$$

The system matrices are not allowed to depend on t.

- In both cases, the matrices are assumed known.
- Uncertainty in the matrices can be modeled as a non-linear system.

Observability for LTI systems

Continuous-time observability

Consider a continuous-time LTI system of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \qquad\qquad \mathbf{z} = \mathbf{H}\mathbf{x}.$$

The observability matrix of the system is

$$\mathbf{Q}_O = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}\mathbf{A} \\ \vdots \\ \mathbf{H}\mathbf{A}^d \end{bmatrix}.$$

We say that the pair $[\mathbf{A}, \mathbf{H}]$ is observable if $\mathbf{Q}_O$ is of full rank.

Discrete-time observability

Simply replace **A** and **B** with their discrete-time equivalents.

Example: Observability for CV model

Consider the system model

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad\qquad \mathbf{H} = \begin{bmatrix} h_1 & h_2 \end{bmatrix}.$$

The observability matrix is then

$$\mathbf{Q}_O = \begin{bmatrix} \mathbf{H} \\ \mathbf{HA} \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ 0 & h_1 \end{bmatrix}.$$

Case 1: Only position measurements

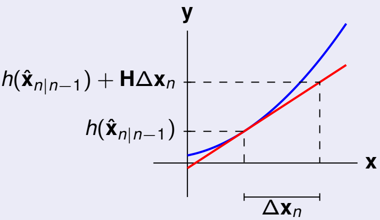
$h_2 = 0 \;\Rightarrow\; \mathbf{Q}_O = \mathbf{I} \;\Rightarrow\; \mathbf{Q}_O$ is of full rank $\;\Rightarrow\;$ The system is observable.

Case 2: Only velocity measurements

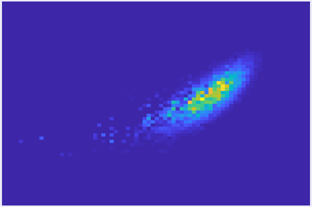
$h_1 = 0 \;\Rightarrow$ Second row of $\mathbf{Q}_O$ is zero $\;\Rightarrow\;$ The system is not observable.

The road ahead

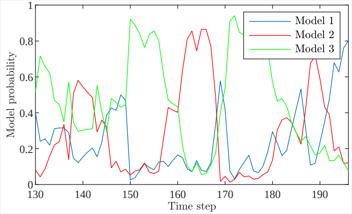
The EKF



Particle filters



Interacting Multiple Models



Target tracking