TTK4115

The Kalman Filter

Morten D. Pedersen

This lecture

1. Random Processes
2. Gaussian noise
3. Optimal estimation
4. Colored noise
5. Diagonalization
6. Discrete time modeling
7. Kalman filtering in discrete time
8. Final notes

Topic

1. Random Processes
2. Gaussian noise
3. Optimal estimation
4. Colored noise
5. Diagonalization
6. Discrete time modeling
7. Kalman filtering in discrete time
8. Final notes

Deterministic state-space model

Assume that $\mathbf{x}_0$ and $\mathbf{u}(t)$ are *known*. Then, the state space model given below is a deterministic process

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \quad \mathbf{x}(0) = \mathbf{x}_0 \\ \mathbf{y} &= \mathbf{C}\mathbf{x}\end{aligned}$$

It is in fact straightforward to compute the deterministic solution which is given by

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}_0 + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau, \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t)$$

Problem

What happens if

- $\mathbf{u}(t)$ is unknown and random? (Denoted $\mathbb{u}(t)$)
- $\mathbf{x}_0$ is unknown and random? (Denoted $\mathbb{x}_0$)

Then it follows that $\mathbf{y}(t)$ and $\mathbf{x}(t)$ must also be random and unknown! Here denoted by the symbols $\mathbb{y}(t)$ and $\mathbb{x}(t)$.

Uncertain state-space model

The state space model given below describes a *random* process

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbb{u}, \quad \mathbb{x}(0) = \mathbb{x}_0$$
$$\mathbb{y} = \mathbf{C}\mathbb{x}$$

The *uncertain* solution follows from

$$\mathbb{x}(t) = e^{\mathbf{A}t}\mathbb{x}_0 + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbb{u}(\tau)\, d\tau, \quad \mathbb{y}(t) = \mathbf{C}\mathbb{x}(t)$$

Rather than attempting to find out what ***will*** *happen, it is possible to find out what is* ***likely*** *to happen.*

What can one *expect* from $\mathbb{x}(t)$, $\mathbb{x}_0$, $\mathbb{y}(t)$, $\mathbb{u}(t)$?

The expectation operator[1] E can be used to identify a series of important quantities at each time t.

Mean : $m_x(t) = \mathrm{E}[\mathbb{x}(t)]$

Variance : $\mathrm{var}[\mathbb{x}(t)] = \mathrm{E}[(\mathbb{x}(t) - m_x(t))^2]$

Covariance : $\mathrm{cov}[\mathbb{x}_1(t), \mathbb{x}_2(t)] = \mathrm{E}[(\mathbb{x}_1(t) - m_{x_1}(t))(\mathbb{x}_2(t) - m_{x_2}(t))^\mathsf{T}]$

[1] A *linear* operator satisfying $\mathrm{E}[\mathbb{x} + c] = \mathrm{E}[\mathbb{x}] + c$, $\quad \mathrm{E}[\mathbb{x}_1 + \mathbb{x}_2] = \mathrm{E}[\mathbb{x}_1] + \mathrm{E}[\mathbb{x}_2]$, $\quad \mathrm{E}[a\mathbb{x}] = a\mathrm{E}[\mathbb{x}]$.

Uncertain state-space model

$$\mathbb{x}(t) = e^{\mathbf{A}t}\mathbb{x}_0 + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbb{u}(\tau)\, d\tau, \quad \mathbb{y}(t) = \mathbf{C}\mathbb{x}(t)$$

Mean of random process

Where to *expect* $\mathbb{x}(t)$ is found in the following manner

$$\mathbf{m}_\mathbf{x}(t) \triangleq \mathrm{E}[\mathbb{x}(t)] = \mathrm{E}\left[e^{\mathbf{A}t}\mathbb{x}_0 + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbb{u}(\tau)\, d\tau, \quad \mathbb{y}(t) = \mathbf{C}\mathbb{x}(t)\right]$$

$$= e^{\mathbf{A}t}\mathrm{E}[\mathbb{x}_0] + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathrm{E}[\mathbb{u}(\tau)]\, d\tau, \quad \mathrm{E}[\mathbb{y}(t)] = \mathbf{C}\mathrm{E}[\mathbb{x}(t)]$$

Alternative representation

Differentiating on both sides produces a simple model for the mean

$$\dot{\mathbf{m}}_\mathbf{x}(t) = \mathbf{A}\left[e^{\mathbf{A}t}\mathrm{E}[\mathbb{x}_0] + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{m}_\mathbf{u}(\tau)\, d\tau\right] + \mathbf{B}\mathbf{m}_\mathbf{u}(t)$$

$$= \mathbf{A}\mathbf{m}_\mathbf{x}(t) + \mathbf{B}\mathbf{m}_\mathbf{u}(t), \quad \mathbf{m}_\mathbf{y}(t) = \mathbf{C}\mathbf{m}_\mathbf{x}(t)$$

Here, $\mathbf{m}_\mathbf{u}(t) \triangleq \mathrm{E}[\mathbb{u}(t)]$ and $\mathbf{m}_{\mathbf{x}_0} \triangleq \mathrm{E}[\mathbb{x}_0]$, whilst $\mathbf{m}_\mathbf{y}(t) \triangleq \mathrm{E}[\mathbb{y}(t)]$.

Uncertain state-space model

The state space model given below describes a *random* process

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbb{u}, \quad \mathbb{x}(0) = \mathbb{x}_0$$
$$\mathbb{y} = \mathbf{C}\mathbb{x}$$

Mean of random process

Its **mean** (expected value) follows from using the expectancy operator on the preceding equation

$$\dot{\mathbf{m}}_{\mathbf{x}} = \mathbf{A}\mathbf{m}_{\mathbf{x}} + \mathbf{B}\mathbf{m}_{\mathbf{u}}, \quad \mathbf{m}_{\mathbf{x}}(0) = \mathbf{m}_{\mathbf{x}_0}$$
$$\mathbf{m}_{\mathbf{y}} = \mathbf{C}\mathbf{m}_{\mathbf{x}}$$

This result implies that deterministic models are found in the limit $\text{var}[\mathbb{x}] \to \mathbf{0}$.

Covariance & variance of random process

The *variance* provides a measure of the spread of the variable in question whereas the *covariance* measures the relation between two random variables. It is customary to collect this information in a *covariance matrix*. A vector $\mathbb{x}(t) \in \mathbb{R}^n$ is thus equipped with the covariance matrix

$$\mathcal{C}_{\mathbf{x}}(t) \triangleq \mathrm{E}\begin{bmatrix} (\mathbb{x}_1 - m_{x_1})(\mathbb{x}_1 - m_{x_1}) & (\mathbb{x}_1 - m_{x_1})(\mathbb{x}_2 - m_{x_2}) & \cdots & (\mathbb{x}_1 - m_{x_1})(\mathbb{x}_n - m_{x_n}) \\ (\mathbb{x}_2 - m_{x_2})(\mathbb{x}_1 - m_{x_1}) & (\mathbb{x}_2 - m_{x_2})(\mathbb{x}_2 - m_{x_2}) & \cdots & (\mathbb{x}_2 - m_{x_2})(\mathbb{x}_n - m_{x_n}) \\ \dots & \dots & \ddots & \vdots \\ (\mathbb{x}_n - m_{x_n})(\mathbb{x}_1 - m_{x_1}) & (\mathbb{x}_n - m_{x_n})(\mathbb{x}_2 - m_{x_2}) & \cdots & (\mathbb{x}_n - m_{x_n})(\mathbb{x}_n - m_{x_n}) \end{bmatrix}$$

A compact vectorial representation is given by

$$\mathcal{C}_{\mathbf{x}}(t) = \mathrm{E}[(\mathbb{x}(t) - \mathbf{m}_{\mathbf{x}}(t))(\mathbb{x}(t) - \mathbf{m}_{\mathbf{x}}(t))^{\mathsf{T}}]$$

Note that variances are located on the *diagonal*.

Uncertain state-space model about the *mean*

$$\dot{\mathbb{x}} - \dot{\mathbf{m}}_\mathbf{x} = \mathbf{A}(\mathbb{x} - \mathbf{m}_\mathbf{x}) + \mathbf{B}(\mathbb{u} - \mathbf{m}_\mathbf{u})$$
$$\mathbb{y} - \mathbf{m}_\mathbf{y} = \mathbf{C}(\mathbb{x} - \mathbf{m}_\mathbf{x})$$

Covariance & variance of random process

The *variance* provides a measure of the spread of the variable in question whereas the *covariance* measures the relation between two random variables. It is customary to collect this information in a *covariance matrix*. A compact vectorial representation is given by

$$\mathcal{C}_\mathbf{x}(t) = \mathrm{E}[(\mathbb{x}(t) - \mathbf{m}_\mathbf{x}(t))(\mathbb{x}(t) - \mathbf{m}_\mathbf{x}(t))^\mathsf{T}]$$

Computation

Direct differentiation yields a *covariance update equation*, viz.

$$\begin{aligned}\dot{\mathcal{C}}_\mathbf{x} &= \mathrm{E}[(\dot{\mathbb{x}} - \dot{\mathbf{m}}_\mathbf{x})(\mathbb{x} - \mathbf{m}_\mathbf{x})^\mathsf{T}] + \mathrm{E}[(\mathbb{x} - \mathbf{m}_\mathbf{x})(\dot{\mathbb{x}} - \dot{\mathbf{m}}_\mathbf{x})^\mathsf{T}] \\ &= \mathrm{E}[(\mathbf{A}(\mathbb{x} - \mathbf{m}_\mathbf{x}) + \mathbf{B}(\mathbb{u} - \mathbf{m}_\mathbf{u}))(\mathbb{x} - \mathbf{m}_\mathbf{x})^\mathsf{T}] + \mathrm{E}[(\mathbb{x} - \mathbf{m}_\mathbf{x})(\mathbf{A}(\mathbb{x} - \mathbf{m}_\mathbf{x}) + \mathbf{B}(\mathbb{u} - \mathbf{m}_\mathbf{u}))^\mathsf{T}] \\ &= \mathbf{A}\mathcal{C}_\mathbf{x} + \mathcal{C}_\mathbf{x}\mathbf{A}^\mathsf{T} + \mathbf{B}\mathrm{E}[(\mathbb{u} - \mathbf{m}_\mathbf{u})(\mathbb{x} - \mathbf{m}_\mathbf{x})^\mathsf{T}] + \mathrm{E}[(\mathbb{x} - \mathbf{m}_\mathbf{x})(\mathbb{u} - \mathbf{m}_\mathbf{u})^\mathsf{T}]\mathbf{B}^\mathsf{T}\end{aligned}$$

But, what is the covariance between $\mathbb{x}$ and $\mathbb{u}$?

Uncertain state-space model about the *mean*

$$\mathbb{x}(t) - \mathbf{m}_{\mathbf{x}}(t) = e^{\mathbf{A}t}(\mathbb{x}_0 - \mathbf{m}_{\mathbf{x}_0}) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}(\mathbb{u}(\tau) - \mathbf{m}_{\mathbf{u}}(\tau))\, d\tau$$

Covariance computation

$$\begin{aligned}
&\mathrm{E}[(\mathbb{x}(t) - \mathbf{m}_{\mathbf{x}}(t))(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}} \\
&= \mathrm{E}[e^{\mathbf{A}t}(\mathbb{x}_0 - \mathbf{m}_{\mathbf{x}_0})(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}\mathbf{B}^{\mathsf{T}}] + \mathrm{E}\left[\left(\int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}(\mathbb{u}(\tau) - \mathbf{m}_{\mathbf{u}}(\tau))\, d\tau\right)(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}\mathbf{B}^{\mathsf{T}}\right] \\
&= e^{\mathbf{A}t}\mathrm{E}[(\mathbb{x}_0 - \mathbf{m}_{\mathbf{x}_0})(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}} + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathrm{E}[(\mathbb{u}(\tau) - \mathbf{m}_{\mathbf{u}}(\tau))(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}}\, d\tau
\end{aligned}$$

Covariance computation

$$\begin{aligned}&\mathsf{E}[(\mathbb{x}(t)-\mathbf{m}_{\mathbf{x}}(t))(\mathbb{u}(t)-\mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}}\\ =\ & e^{\mathbf{A}t}\mathsf{E}[(\mathbb{x}_0-\mathbf{m}_{\mathbf{x}_0})(\mathbb{u}(t)-\mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}}+\int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathsf{E}[(\mathbb{u}(\tau)-\mathbf{m}_{\mathbf{u}}(\tau))(\mathbb{u}(t)-\mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}}\,d\tau\end{aligned}$$

Causality

The input given in the interval $[0, t)$ cannot affect the initial conditions at $t = 0$ by having impacts *backwards* in time. Arguing from causality, one can assume

$$\mathsf{E}[(\mathbb{x}_0-\mathbf{m}_{\mathbf{x}_0})(\mathbb{u}(t)-\mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]=\mathbf{0}$$

Autocovariance

The autocovariance quantifies how a signal **correlates with itself across time**. For the random input used in the present process we have

$$\mathcal{A}_{\mathbf{u}}(t,\tau)\triangleq\mathsf{E}[(\mathbb{u}(\tau)-\mathbf{m}_{\mathbf{u}}(\tau))(\mathbb{u}(t)-\mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]$$

Autocovariance

The autocovariance quantifies how a signal **correlates with itself across time**. For the random input used in the present process we have

$$\mathcal{A}_{\mathbf{u}}(t,\tau) \triangleq \mathsf{E}[(\mathbb{u}(\tau) - \mathbf{m}_{\mathbf{u}}(\tau))(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]$$

If the mean is zero so that $\mathbf{m}_{\mathbf{u}}(t) = \mathbf{0}$ the *autocorrelation* follows as $\mathcal{R}_{\mathbf{u}}(t,\tau) = \mathcal{A}_{\mathbf{u}}(t,\tau)$.

White noise

White noise is a theoretical signal that is *completely uncorrelated* to itself over time. This is to say that knowing the white noise $\mathbb{n}(t)$ at the instant t_1 does not inform us in any way whatsoever about its value at time t_2. Mathematically this is described by

$$\mathcal{A}_n(t,\tau) = \mathsf{E}[(\mathbb{n}(t) - m_n(t))(\mathbb{n}(\tau) - m_n(\tau))] = 0, \quad t \neq \tau$$

At $\tau = t$, the autocovariance reduces to a simple *variance*. This variance is given by

$$\mathcal{A}_n(t,t) = \mathsf{E}[(\mathbb{n}(t) - m_n(t))^2] = \delta(0) q_n(t), \quad t = \tau$$

where $\delta(t)$ represents Dirac's function and $q_n(t) > 0$. The autocovariance of white noise thus follows from

$$\mathcal{A}_n(t,\tau) = \delta(t-\tau) q_n(\tau)$$

White noise is a theoretical construct aimed at simplifying analysis and modeling - No physical signal has infinite variance.

Autocovariance with $\mathbb{u}$ modeled as **white noise**.

The autocovariance quantifies how a signal **correlates with itself across time**. For the random input used in the present process we have

$$\mathcal{A}_{\mathbf{u}}(t,\tau) \triangleq \mathrm{E}[(\mathbb{u}(\tau) - \mathbf{m}_{\mathbf{u}}(\tau))(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]$$

Assuming that $\mathbb{u}(t)$ represents white noise permits the simplification

$$\mathcal{A}_{\mathbf{u}}(t,\tau) = \delta(t-\tau)\mathbf{Q}_{\mathbf{u}}(\tau), \quad \mathbf{Q}_{\mathbf{u}} \succ \mathbf{0}$$

Covariance computation

The particular properties of white noise permit significant simplifications to the analysis. Here we use the half-maximum convention on Heaviside's function $\Theta(0) = 1/2$ to arrive at

$$\mathrm{E}[(\mathbb{x}(t) - \mathbf{m}_{\mathbf{x}}(t))(\mathbb{u}(t) - \mathbf{m}_{\mathbf{u}}(t))^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}} = \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathcal{A}_{\mathbf{u}}(t,\tau)\mathbf{B}^{\mathsf{T}}\, d\tau$$

$$= \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\delta(t-\tau)\mathbf{Q}_{\mathbf{u}}(\tau)\mathbf{B}^{\mathsf{T}}\, d\tau = \int_0^\infty \Theta(t-\tau)e^{\mathbf{A}(t-\tau)}\mathbf{B}\delta(t-\tau)\mathbf{Q}_{\mathbf{u}}(\tau)\mathbf{B}^{\mathsf{T}}\, d\tau = \frac{1}{2}\mathbf{B}\mathbf{Q}_{\mathbf{u}}(t)\mathbf{B}^{\mathsf{T}}$$

Covariance update equation

$$\dot{\mathcal{C}}_{\mathbf{x}} - \mathbf{A}\mathcal{C}_{\mathbf{x}} - \mathcal{C}_{\mathbf{x}}\mathbf{A}^{\mathsf{T}} = \mathbf{B}\mathrm{E}[(\mathbb{u} - \mathbf{m}_{\mathbf{u}})(\mathbb{x} - \mathbf{m}_{\mathbf{x}})^{\mathsf{T}}] + \mathrm{E}[(\mathbb{x} - \mathbf{m}_{\mathbf{x}})(\mathbb{u} - \mathbf{m}_{\mathbf{u}})^{\mathsf{T}}]\mathbf{B}^{\mathsf{T}} = \mathbf{B}\mathbf{Q}_{\mathbf{u}}(t)\mathbf{B}^{\mathsf{T}}$$

It will be assumed in the following that $\mathbf{Q}_{\mathbf{u}}$ is a constant matrix, although this need not be the case.

Uncertain state-space model

It is in fact possible to say quite a lot about what to *expect* from the random process given below, even though both $\mathbb{u}$ and $\mathbb{x}_0$ are *random*.

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbb{u}, \quad \mathbb{y} = \mathbf{C}\mathbb{x}$$

Results

The following quantities are assumed *known*.

Means : $\mathrm{E}[\mathbb{u}(t)] = \mathbf{m_u}(t)$ and $\mathrm{E}[\mathbb{x}_0] = \mathbf{m}_{\mathbf{x}_0}$.

Covariances : $\mathrm{E}[(\mathbb{u} - \mathbf{m_u})(\mathbb{u} - \mathbf{m_u})^\mathsf{T}] = \delta(0)\mathbf{Q_u}$ and $\mathrm{E}[(\mathbb{x}_0 - \mathbf{m}_{\mathbf{x}_0})(\mathbb{x}_0 - \mathbf{m}_{\mathbf{x}_0})^\mathsf{T}] = \mathcal{C}_\mathbf{x}(0)$.

Adopting the assumption that $\mathbb{u}(t)$ is well represented by white noise informs us what to *expect* from the uncertain model. Verify, and note **linearity**, of the following.

The **means** are given by

$$\begin{aligned} \dot{\mathbf{m}}_\mathbf{x} &= \mathbf{A}\mathbf{m}_\mathbf{x} + \mathbf{B}\mathbf{m}_\mathbf{u}, \quad \mathbf{m}_\mathbf{x}(0) = \mathbf{m}_{\mathbf{x}_0} \\ \mathbf{m}_\mathbf{y} &= \mathbf{C}\mathbf{m}_\mathbf{x} \end{aligned}$$

The **covariance matrices** follow from

$$\begin{aligned} \dot{\mathcal{C}}_\mathbf{x} &= \mathbf{A}\mathcal{C}_\mathbf{x} + \mathcal{C}_\mathbf{x}\mathbf{A}^\mathsf{T} + \mathbf{B}\mathbf{Q_u}\mathbf{B}^\mathsf{T}, \quad \mathcal{C}_\mathbf{x}(0) = \mathcal{C}_{\mathbf{x}_0} \\ \mathcal{C}_\mathbf{y} &= \mathbf{C}\mathcal{C}_\mathbf{x}\mathbf{C}^\mathsf{T} \end{aligned}$$

Here $\mathcal{C}_\mathbf{y} \triangleq \mathrm{E}[(\mathbb{y} - \mathbf{m_y})(\mathbb{y} - \mathbf{m_y})^\mathsf{T}] = \mathbf{C}\mathrm{E}[(\mathbb{x} - \mathbf{m_x})(\mathbb{x} - \mathbf{m_x})^\mathsf{T}]\mathbf{C}^\mathsf{T} = \mathbf{C}\mathcal{C}_\mathbf{x}\mathbf{C}^\mathsf{T}$.

Topic

Gaussian distributions

- It is in many cases sufficient to know the means and covariances[2] of a random signal.
- If additional information is sought, a more explicit probability distribution must be found.
- The normal (Gaussian) distribution represents a wide variety of random processes found in nature and technology.
- The multivariable Gaussian probability distribution is completely specified by the mean vector and covariance matrix. *This implies that the model developed above can be used to model the evolution of a complete probability distribution!*

The Gaussian

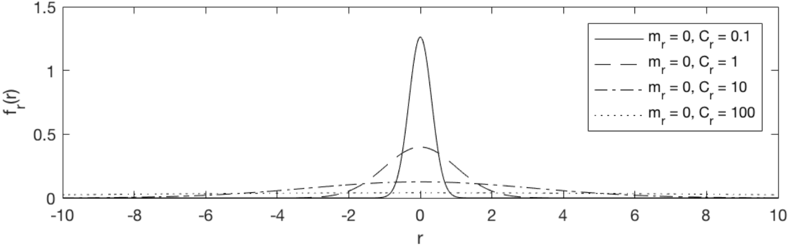
A Gaussian probability distribution can be defined for a random vector $\mathbb{r} \in \mathbb{R}^n$ if its mean $\mathbf{m_r} \in \mathbb{R}^n$ and covariance matrix $\mathcal{C}_\mathbf{r} \in \mathbb{R}^{n\times n}$ are supplied. Then,

$$f_{\mathbb{r}}(\mathbf{r}) = \frac{1}{(2\pi)^{n/2}|\mathcal{C}_\mathbf{r}|^{1/2}} \text{Exp}\left[-\frac{1}{2}(\mathbf{r}-\mathbf{m_r})^\mathsf{T}\mathcal{C}_\mathbf{r}^{-1}(\mathbf{r}-\mathbf{m_r})\right]$$

[2] Understood in the generalized sense including variances on the diagonal.

The Gaussian - scalar case

$$f_{\mathbb{r}}(r) = (2\pi\mathcal{C}_r)^{-1/2}\mathrm{Exp}\left[-(r - m_r)^2/2\mathcal{C}_r\right]$$



Gaussian white noise

The white noise signal $\mathbb{n}(t)$ with variance $\mathcal{C}_n = \delta(0)q$ can be construed as having the probability distribution

$$f_{\mathbb{n}}(n) = \lim_{h\to\infty} \frac{1}{\sqrt{2\pi hq}}\mathrm{Exp}\left[-\frac{(n - m_n)^2}{2qh}\right]$$

Continuous time Gaussian white noise therefore takes on values in the interval $(-\infty, \infty)$ with equal probability.

Low-pass filtering

A first order low-pass filter with a Gaussian white noise on the input is given by

$$\tau \dot{\mathbb{y}}(t) + \mathbb{y}(t) = \mathbb{u}(t)$$

Let the constant covariance and mean of the random input be denoted respectively by $\mathcal{C}_u = q\delta(0)$ and $m_u \neq 0$. *What is the probability distribution of $\mathbb{y}$ as $t \to \infty$?*

Result

The limiting mean response m_y follows from an application of the final value theorem

$$\tau \dot{m}_y(t) + m_y(t) = m_u, \quad \lim_{t\to\infty} m_y(t) = \lim_{s\to 0} s m_y(s) = \lim_{s\to 0} \frac{m_u}{\tau s + 1} = m_u$$

The scalar (co)variance of $\mathbb{y}$ can be given as

$$\dot{\mathcal{C}}_y(t) = -2(1/\tau)\mathcal{C}_y(t) + (1/\tau)^2 q, \quad \lim_{t\to\infty} \mathcal{C}_y(t) = \lim_{s\to 0} s\mathcal{C}_y(s) = \lim_{s\to 0} \frac{q}{\tau^2 s + 2\tau} = \frac{q}{2\tau}$$

The constant mean is not changed under lowpassing, but the variance of the output is *finite* and inversely proportional to the time-constant (alt. proportional to the filter's bandwidth ($\omega_b = 1/\tau$).

The probability distribution of the output can now be supplied as

$$f_{\mathbb{y}}(y) = (\pi q/\tau)^{-1/2} \mathrm{Exp}\left[-\tau(y - m_u)^2/q\right]$$

Topic

Physical model

Let a general plant model be given by a random process

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbb{w}, \quad \mathbb{y} = \mathbf{C}\mathbb{x} + \mathbb{v}$$

The random signals giving rise to the uncertainty are

Noise $\mathbb{v}$: represented by a zero mean white Gaussian signal with autocovariance/autocorrelation $\mathcal{A}_{\mathbf{v}}(t,\tau) = \mathrm{E}[\mathbb{v}(t)\mathbb{v}(\tau)^{\mathsf{T}}] = \delta(t-\tau)\mathbf{R}_{\mathbf{v}}$.

Disturbance $\mathbb{w}$: represented by a zero mean white Gaussian signal with autocovariance/autocorrelation $\mathcal{A}_{\mathbf{w}}(t,\tau) = \mathrm{E}[\mathbb{w}(\tau)\mathbb{w}(t)^{\mathsf{T}}] = \delta(t-\tau)\mathbf{Q}_{\mathbf{w}}$

The noise and disturbance are assumed to be *uncorrelated* implying that $\mathcal{A}_{\mathbf{vw}}(t,\tau) = \mathrm{E}[\mathbb{v}(t)\mathbb{w}(\tau)^{\mathsf{T}}] \equiv \mathbf{0}$.

Luenberger observer

It will be of interest to perform estimation on the random process representing the plant. Let a *Luenberger observer* be given by

$$\dot{\hat{\mathbb{x}}} = \mathbf{A}\hat{\mathbb{x}} + \mathbf{B}\mathbf{u} + \mathbf{L}(t)(\mathbb{y} - \mathbf{C}\hat{\mathbb{x}})$$

Note that the estimate is not deterministic since it is perturbed by the random process $\mathbb{y}$. We let $\mathbf{L}(t)$ be undetermined for now.

Dynamics of the estimation error

The random estimation error is defined by $\mathbb{e} = \mathbb{x} - \hat{\mathbb{x}}$. Verify that

$$\dot{\mathbb{e}} = (\mathbf{A} - \mathbf{L}(t)\mathbf{C})\mathbb{e} + \mathbf{G}\mathbb{w} - \mathbf{L}\mathbb{v}$$

Unbiased estimation

At $t = 0$ the observer is initialized at the *mean* of the true state vector so that $\hat{\mathbf{x}}_0 = \mathrm{E}[\mathbb{x}_0]$. Taking expectations, noting the unbiased noise and disturbance, shows that no mean error is committed

$$\dot{\mathbf{m}}_{\mathbf{e}} = (\mathbf{A} - \mathbf{L}(t)\mathbf{C})\mathbf{m}_{\mathbf{e}}, \quad \mathbf{m}_{\mathbf{e}}(0) = \mathrm{E}[\mathbb{x}_0] - \hat{\mathbf{x}}_0 = \mathbf{0} \quad \Rightarrow \mathbf{m}_{\mathbf{e}}(t) = \mathbf{0}$$

This result implies that the estimate is *unbiased*.

Covariance dynamics

The covariance matrix for the estimation error is equipped with the special notation

$$\mathbf{P}(t) \triangleq \mathrm{E}[\mathbb{e}(t)\mathbb{e}(t)^\mathsf{T}]$$

The matrix $\mathbf{P}$ quantifies the uncertainty in the estimate; low variances (found along the diagonal) imply good estimates! Verify that

$$\dot{\mathbf{P}} = (\mathbf{A} - \mathbf{L}\mathbf{C})\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{L}\mathbf{C})^\mathsf{T} + \mathrm{E}[\mathbb{e}(\mathbb{w}^\mathsf{T}\mathbf{G}^\mathsf{T} - \mathbb{v}^\mathsf{T}\mathbf{L}^\mathsf{T})] + \mathrm{E}[\mathbb{e}(\mathbb{w}^\mathsf{T}\mathbf{G}^\mathsf{T} - \mathbb{v}^\mathsf{T}\mathbf{L}^\mathsf{T})]^\mathsf{T}$$

The covariance matrices $\mathrm{E}[\mathbb{w}\mathbb{e}^\mathsf{T}]$ and $\mathrm{E}[\mathbb{v}\mathbb{e}^\mathsf{T}]$ must now be found.

Dynamics of the estimation error

Since $\mathbf{L}(t)$ is time-varying, a transition matrix[3] satisfying $\dot{\mathbf{\Phi}}(t,\tau) = (\mathbf{A} - \mathbf{L}(t)\mathbf{C})\mathbf{\Phi}(t,\tau)$ and $\mathbf{\Phi}(t,t) = \mathbb{I}$ is used to recover the solution of $\dot{\mathbb{e}} = (\mathbf{A} - \mathbf{L}(t)\mathbf{C})\mathbb{e} + \mathbf{G}\mathbb{w} - \mathbf{L}\mathbb{v}$, viz.

$$\mathbb{e}(t) = \mathbf{\Phi}(t,0)\mathbb{e}_0 + \int_0^t \mathbf{\Phi}(t,\tau)(\mathbf{G}\mathbb{w}(\tau) - \mathbf{L}(\tau)\mathbb{v}(\tau))\,d\tau$$

Covariance dynamics

$$\dot{\mathbf{P}} = (\mathbf{A} - \mathbf{LC})\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{LC})^\mathsf{T} + \mathrm{E}[\mathbb{e}(\mathbb{w}^\mathsf{T}\mathbf{G}^\mathsf{T} - \mathbb{v}^\mathsf{T}\mathbf{L}^\mathsf{T})] + \mathrm{E}[\mathbb{e}(\mathbb{w}^\mathsf{T}\mathbf{G}^\mathsf{T} - \mathbb{v}^\mathsf{T}\mathbf{L}^\mathsf{T})]^\mathsf{T}$$

Computation

Using a similar procedure as before, it follows that

$$\begin{aligned}
&\mathrm{E}[\mathbb{e}(t)(\mathbb{w}(t)^\mathsf{T}\mathbf{G}^\mathsf{T} - \mathbb{v}(t)^\mathsf{T}\mathbf{L}(t)^\mathsf{T})] \\
&\quad = \int_0^\infty \Theta(t-\tau)\mathbf{\Phi}(t,\tau)\mathrm{E}[(\mathbf{G}\mathbb{w}(\tau) - \mathbf{L}(\tau)\mathbb{v}(\tau))(\mathbb{w}(t)\mathbf{G}^\mathsf{T} - \mathbb{v}(t)^\mathsf{T}\mathbf{L}(t)^\mathsf{T}))]\,d\tau \\
&\quad = \int_0^\infty \Theta(t-\tau)\mathbf{\Phi}(t,\tau)(\mathbf{G}\mathcal{A}_\mathbf{w}(t,\tau)\mathbf{G}^\mathsf{T} + \mathbf{L}\mathcal{A}_\mathbf{v}(t,\tau)\mathbf{L}^\mathsf{T})\,d\tau = \frac{1}{2}\mathbf{G}\mathbf{Q}_\mathbf{w}\mathbf{G}^\mathsf{T} + \frac{1}{2}\mathbf{L}\mathbf{R}_\mathbf{v}\mathbf{L}^\mathsf{T}
\end{aligned}$$

Causality justifies the assumptions $\mathrm{E}[\mathbb{w}\mathbb{e}_0^\mathsf{T}] = \mathbf{0}$ and $\mathrm{E}[\mathbb{v}\mathbb{e}_0^\mathsf{T}] = \mathbf{0}$.

[3] Reduces to $\mathbf{\Phi}(t,\tau) = e^{(\mathbf{A}-\mathbf{LC})(t-\tau)}$ for constant $\mathbf{L}$.

Covariance dynamics

The following equation describes the covariance dynamics of the random estimate error $\mathbb{e}$, viz.

$$\dot{\mathbf{P}} = (\mathbf{A} - \mathbf{LC})\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{LC})^{\mathsf{T}} + \mathbf{GQ_wG}^{\mathsf{T}} + \mathbf{LR_vL}^{\mathsf{T}}$$

Estimation performance

It was demonstrated that a low-pass filter reduced the variance in the input signal, hence bringing the expected deviation closer to the mean. The variance of the i'th estimation error at time t is given by

$$\sigma_i^2(t) = \mathrm{E}[\mathbb{e}_i(t)\mathbb{e}_i(t)] = P_{ii}(t)$$

Let the mean-square errors serve as a measure of the overall estimation performance

$$J_{\mathrm{mse}} = \sum_i^n \sigma_i^2 = \mathrm{tr}(\mathbf{P}) > 0$$

Kalman Gain

We now ensure that J_{mse} decreases at the fastest possible rate by optimizing with respect to the observer gain $\mathbf{L}(t)$. A bit of matrix differentiation yields the result

$$\frac{\partial \mathrm{tr}(\dot{\mathbf{P}})}{\partial \mathbf{L}} = \frac{\partial}{\partial \mathbf{L}}\mathrm{tr}((\mathbf{A} - \mathbf{LC})\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{LC})^{\mathsf{T}} + \mathbf{GQ_wG}^{\mathsf{T}} + \mathbf{LR_vL}^{\mathsf{T}}) = -2\mathbf{PC}^{\mathsf{T}} + 2\mathbf{LR_v} = \mathbf{0}$$

The Kalman Gain $\mathbf{L}(t) = \mathbf{P}(t)\mathbf{C}^{\mathsf{T}}\mathbf{R_v}^{-1}$ decreases the uncertainty in the estimate at the fastest rate.

Kalman Gain

We now ensure that J_{mse} decreases at the fastest possible rate by optimizing with respect to the observer gain $\mathbf{L}(t)$. A bit of matrix differentiation yields the result

$$\frac{\partial \text{tr}(\dot{\mathbf{P}})}{\partial \mathbf{L}} = \frac{\partial}{\partial \mathbf{L}} \text{tr}((\mathbf{A} - \mathbf{L}\mathbf{C})\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{L}\mathbf{C})^\mathsf{T} + \mathbf{G}\mathbf{Q}_\mathbf{w}\mathbf{G}^\mathsf{T} + \mathbf{L}\mathbf{R}_\mathbf{v}\mathbf{L}^\mathsf{T}) = -2\mathbf{P}\mathbf{C}^\mathsf{T} + 2\mathbf{L}\mathbf{R}_\mathbf{v} = \mathbf{0}$$

The Kalman Gain $\mathbf{L}(t) = \mathbf{P}(t)\mathbf{C}^\mathsf{T}\mathbf{R}_\mathbf{v}^{-1}$ decreases the uncertainty in the estimate at the fastest rate.

Differentiating yet again with respect to $\mathbf{L}$ shows that a minimum has indeed been found; the Hessian is positive definite.

$$\frac{1}{2}\frac{\partial^2 \text{tr}(\dot{\mathbf{P}})}{\partial \mathbf{L}^2} = \mathbf{R}_\mathbf{v} \succ \mathbf{0}$$

Optimal covariance dynamics

Upon selection of $\mathbf{L}$, the covariance follows from

$$\dot{\mathbf{P}} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\mathsf{T} + \mathbf{G}\mathbf{Q}_\mathbf{w}\mathbf{G}^\mathsf{T} - \mathbf{P}\mathbf{C}^\mathsf{T}\mathbf{R}_\mathbf{v}^{-1}\mathbf{C}\mathbf{P}$$

This is known as the *Matrix Riccati Equation*. The covariance matrix will converge assuming stationarity of the random processes. Hence $\mathbf{P}(t) \to \mathbf{P}_\infty,\ t \to \infty$. An optimal *stationary* observer gain can be obtained by solving

$$\mathbf{A}\mathbf{P}_\infty + \mathbf{P}_\infty\mathbf{A}^\mathsf{T} + \mathbf{G}\mathbf{Q}_\mathbf{w}\mathbf{G}^\mathsf{T} - \mathbf{P}_\infty\mathbf{C}^\mathsf{T}\mathbf{R}_\mathbf{v}^{-1}\mathbf{C}\mathbf{P}_\infty = \mathbf{0}, \quad \mathbf{L}_\infty = \mathbf{P}_\infty\mathbf{C}^\mathsf{T}\mathbf{R}_\mathbf{v}^{-1}$$

LQR

The optimal (output-weighted) feedback gain is well known to be given by

$$\mathbf{A}^\mathsf{T}\mathbf{S} + \mathbf{S}\mathbf{A} + \mathbf{C}^\mathsf{T}\mathbf{Q}_\mathbf{y}\mathbf{C} - \mathbf{S}\mathbf{B}\mathbf{R}_\mathbf{u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S} = \mathbf{0}, \quad \mathbf{K} = \mathbf{R}_\mathbf{u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S}$$

Dual dynamics

Recall the *dual* system

$$\dot{\mathbf{z}} = \mathbf{A}^\mathsf{T}\mathbf{z} + \mathbf{C}^\mathsf{T}\mathbf{u}_{\text{dual}}, \quad \mathbf{y}_{\text{dual}} = \mathbf{B}^\mathsf{T}\mathbf{z}$$

Duality[4]

The stationary *Kalman gain* can be construed as the optimal *feedback gain* for the *dual system*. Letting $\mathbf{G} = \mathbf{B}$ (often a convenient choice for matched disturbances[5]), compare

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\mathsf{T} + \mathbf{B}\mathbf{Q}_\mathbf{w}\mathbf{B}^\mathsf{T} - \mathbf{P}\mathbf{C}^\mathsf{T}\mathbf{R}_\mathbf{v}^{-1}\mathbf{C}\mathbf{P} = \mathbf{0}, \quad \mathbf{L}^\mathsf{T} = \mathbf{R}_\mathbf{v}^{-1}\mathbf{C}\mathbf{P}$$

$$\mathbf{A}^\mathsf{T}\mathbf{S} + \mathbf{S}\mathbf{A} + \mathbf{C}^\mathsf{T}\mathbf{Q}_\mathbf{y}\mathbf{C} - \mathbf{S}\mathbf{B}\mathbf{R}_\mathbf{u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S} = \mathbf{0}, \quad \mathbf{K} = \mathbf{R}_\mathbf{u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S}$$

The LQR problem requires a controllable plant, which must hold for the dual plant. This entails that the pair $(\mathbf{A}, \mathbf{C})$ must be observable in order to permit computation of $\mathbf{L}$.

---

[4] The matlab code is simply `L = (lqr(A',C',B*Qw*B',Rv))'`.

[5] Disturbances that can be canceled directly through control.

LQR **stability**

The closed loop plant is governed by $\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{BK})\mathbf{x}$ where

$$\mathbf{A}^\mathsf{T}\mathbf{S} + \mathbf{SA} + \mathbf{C}^\mathsf{T}\mathbf{Q_y}\mathbf{C} - \mathbf{SBR_u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S} = \mathbf{0}, \quad \mathbf{K} = \mathbf{R_u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S}$$

Let a Lyapunov function be given by $V(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{Sx} > 0$. Differentiation yields

$$\begin{aligned}\dot{V} &= \mathbf{x}^\mathsf{T}[\mathbf{S}(\mathbf{A} - \mathbf{BK}) + (\mathbf{A} - \mathbf{BK})^\mathsf{T}\mathbf{S}]\mathbf{x} \\ &= \mathbf{x}^\mathsf{T}[\mathbf{A}^\mathsf{T}\mathbf{S} + \mathbf{SA} - 2\mathbf{SBR_u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S}]\mathbf{x} \\ &= -\mathbf{x}^\mathsf{T}[\mathbf{C}^\mathsf{T}\mathbf{Q_x}\mathbf{C} + \mathbf{SBR_u}^{-1}\mathbf{B}^\mathsf{T}\mathbf{S}]\mathbf{x} = -[\mathbf{y}^\mathsf{T}\mathbf{Q_y}\mathbf{y} + \mathbf{u}^\mathsf{T}\mathbf{Ru}] < 0\end{aligned}$$

Integration shows that cost function giving rise to the LQR decreases as time proceeds. Stability of the closed loop-plant can be concluded if it is a minimal realization.

$$V(t) = V(0) - \int_0^t \mathbf{y}^\mathsf{T}(\tau)\mathbf{Q_y}\mathbf{y}(\tau) + \mathbf{u}(\tau)^\mathsf{T}\mathbf{Ru}(\tau)\, d\tau$$

KF **stability**

The estimation error of a stationary Kalman filter is governed by

$$\dot{\mathbf{e}} = (\mathbf{A} - \mathbf{LC})\mathbf{e}$$

where

$$\mathbf{AP} + \mathbf{PA}^\mathsf{T} + \mathbf{BQ_wB}^\mathsf{T} - \mathbf{PC}^\mathsf{T}\mathbf{R_v}^{-1}\mathbf{CP} = \mathbf{0}, \quad \mathbf{L} = \mathbf{PC}^\mathsf{T}\mathbf{R_v}^{-1}$$

A deterministic model is used here to avoid the ambiguities of a random Lyapunov function.

Let a Lyapunov function be given by $V(\mathbf{e}) = \mathbf{e}^\mathsf{T}\mathbf{P}^{-1}\mathbf{e} > 0$. Differentiation yields

$$\begin{aligned}
\dot{V} &= \mathbf{e}^\mathsf{T}[\mathbf{P}^{-1}(\mathbf{A} - \mathbf{LC}) + (\mathbf{A} - \mathbf{LC})^\mathsf{T}\mathbf{P}^{-1}]\mathbf{e} \\
&= \mathbf{e}^\mathsf{T}\mathbf{P}^{-1}[\mathbf{AP} + \mathbf{PA} - \mathbf{LCP} - \mathbf{PC}^\mathsf{T}\mathbf{L}^\mathsf{T}]\mathbf{P}^{-1}\mathbf{e} \\
&= \mathbf{e}^\mathsf{T}\mathbf{P}^{-1}[\mathbf{AP} + \mathbf{PA} - 2\mathbf{PC}^\mathsf{T}\mathbf{R_v}^{-1}\mathbf{CP}]\mathbf{P}^{-1}\mathbf{e} \\
&= -\mathbf{e}^\mathsf{T}\mathbf{P}^{-1}[\mathbf{BQ_wB}^\mathsf{T} + \mathbf{PC}^\mathsf{T}\mathbf{R_v}^{-1}\mathbf{CP}]\mathbf{P}^{-1}\mathbf{e} < 0
\end{aligned}$$

This indicates stability since the left-hand side is negative, forcing the Lyapunov function to decrease. (A more detailed analysis is found in Hespanha 2009).

Topic

Colored noise

In the development of the continuous-time Kalman Filter, a crucial assumption was that $\mathbb{v}(t)$ and $\mathbb{w}(t)$ were *white* leading to the simplified autocovariances

$$\mathcal{A}_{\mathbf{v}}(t,\tau) = \mathrm{E}[(\mathbb{v}(t) - \mathbf{m}_{\mathbf{v}}(t))(\mathbb{v}(\tau) - \mathbf{m}_{\mathbf{v}}(\tau))^{\mathsf{T}}] = \mathbf{0}, \quad t \neq \tau$$
$$\mathcal{A}_{\mathbf{w}}(t,\tau) = \mathrm{E}[(\mathbb{w}(t) - \mathbf{m}_{\mathbf{w}}(t))(\mathbb{w}(\tau) - \mathbf{m}_{\mathbf{w}}(\tau))^{\mathsf{T}}] = \mathbf{0}, \quad t \neq \tau$$

What happens if the noise is not white, but colored?

Coloration

Colored noise can be obtained by passing white noise through a linear plant. Let $\mathbb{u}(t) \in \mathbb{R}$ be a white noise with zero mean and autocorrelation $\mathcal{R}_u(\tau) = \mathrm{E}[\mathbb{u}(t)\mathbb{u}(t+\tau)]$. Assuming a stable process initialized a long time ago, the output from a linear filter $H(s)$ is

$$\mathbb{y}(t) = \int_{-\infty}^{\infty} H(t-\tau)\mathbb{u}(\tau)\, d\tau$$

where the causal impulse response is given by

$$H(t) = \Theta(t)\left[\mathbf{c}e^{\mathbf{A}t}\mathbf{b} + d\delta(t)\right]$$

We say that H colors $\mathbb{y}$.

Coloration

Colored noise can be obtained by passing white noise through a linear plant. Let $\mathbb{u}(t)$ be a stationary process with zero mean $m_u = 0$ and autocorrelation $\mathcal{R}_u(\tau) = \mathrm{E}[\mathbb{u}(t)\mathbb{u}(t+\tau)]$. Assuming a stable process initialized a long time ago, the output from a linear filter $H(s)$ is

$$\mathbb{y}(t) = \int_{-\infty}^{\infty} H(t-\alpha)\mathbb{u}(\alpha)\, d\alpha = \int_{-\infty}^{\infty} H(\alpha)\mathbb{u}(t-\alpha)\, d\alpha = H(t) * \mathbb{u}(t)$$

We say that H <u>colors</u> $\mathbb{u}$ to give $\mathbb{y}$.

Stationarity

If the statistics of a random process remain constant over time, it is said to be *stationary*. For some random variable $\mathbb{r}(t)$, this implies that

$$\mathrm{E}[\mathbb{r}(t)] = m_r, \quad \mathrm{E}[\mathbb{r}(t)\mathbb{r}(t+\tau)] = \mathrm{E}[\mathbb{r}(t)\mathbb{r}(t-\tau)] = \mathcal{A}_r(\tau)$$

If the process is zero-mean, the autocovariance reduces to the *autocorrelation*, viz.

$$m_r = 0 \;\Rightarrow \mathcal{A}_r(\tau) = \mathcal{R}_r(\tau)$$

Autocorrelation of $\mathbb{y}$ from $\mathbb{u}$

The autocorrelation of $\mathbb{y}(t)$ can be related to the autocorrelation of $\mathbb{u}(t)$. Verify

$$\mathcal{R}_{uy}(\tau) = \mathsf{E}[\mathbb{y}(t)\mathbb{u}(t+\tau)] = \int_{-\infty}^{\infty} H(\alpha)\underbrace{\mathsf{E}[\mathbb{u}(t-\alpha)\mathbb{u}(t+\tau)]}_{\mathcal{R}_u(\tau+\alpha)}\, d\alpha = H(-\tau) * \mathcal{R}_u(\tau)$$

and

$$\mathcal{R}_{y}(\tau) = \mathsf{E}[\mathbb{y}(t)\mathbb{y}(t-\tau)] = \int_{-\infty}^{\infty} H(\alpha)\underbrace{\mathsf{E}[\mathbb{u}(t-\alpha)\mathbb{y}(t-\tau)]}_{\mathcal{R}_{uy}(\tau-\alpha)}\, d\alpha = H(\tau) * \mathcal{R}_{uy}(\tau)$$

Together, it follows that

$$\mathcal{R}_y(\tau) = H(\tau) * \mathcal{R}_{uy}(\tau) = H(\tau) * [H(-\tau) * \mathcal{R}_u(\tau)] = [H(\tau) * H(-\tau)] * \mathcal{R}_u(\tau)$$

Summary

Assuming stationarity, the filter H produces the autocorrelation $\mathcal{R}_y(\tau)$ from $\mathcal{R}_u(\tau)$ by blending past and present values through convolution

$$\mathcal{R}_y(\tau) = \rho(\tau) * \mathcal{R}_u(\tau), \quad \rho(\tau) \triangleq \int_{-\infty}^{\infty} H(\tau-\alpha)H(-\alpha)\, d\alpha = \int_{-\infty}^{\infty} H(\tau+\beta)H(\beta)\, d\beta$$

Note that $\rho(-\tau) = \rho(\tau)$ since $\rho(\tau) = H(\tau) * H(-\tau)$.

The spectrum of noise

Colors are defined by their spectral power content. Stationary random signals are no different. The **Fourier transform** is the appropriate tool for this analysis.

Fourier transformation

The Fourier transform is defined by

$$\hat{f}(\jmath\omega) = \mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-\jmath\omega t}\,dt, \quad \left(f(s) = \int_{0}^{\infty} f(t)e^{-st}\,dt\right)$$

The Laplace transform for a BIBO plant[6] is shown to the right. If $f(t) = 0$ for $t < 0$ the Fourier transform can be obtained by evaluating the Laplace transform along the imaginary axis.

$$\hat{f}(\jmath\omega) = f(s)|_{s=\jmath\omega}$$

Inverse Fourier transformation

An *inverse* Fourier-transform follows from

$$f(t) = \mathcal{F}^{-1}\{\hat{f}(\jmath\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\jmath\omega)e^{\jmath\omega t}\,d\omega$$

BIBO-stability here forces the *equality* $\mathcal{F}^{-1}\{\hat{f}(\jmath\omega)\} = \mathcal{L}^{-1}\{f(s)\}$.

[6] No poles in the RHP.

The spectrum of noise

Colors are defined by their spectral power content. Stationary random signals are no different. The **Fourier transform** is the appropriate tool for this analysis.

The **Wiener–Khinchin–Einstein**[7] theorem

The following result applies to stationary processes. Since $\mathcal{F}\{f_1(t) * f_2(t)\} = \hat{f}_1(j\omega)\hat{f}_2(j\omega)$ the following result holds for the autocorrelation of $\mathbb{y}$ obtained by filtering $\mathbb{y}$ through H, viz.

$$\mathcal{F}\{\mathcal{R}_y(\tau)\} = \mathcal{F}\{\rho(\tau) * \mathcal{R}_u(\tau)\} = \hat{\rho}(\omega)\mathcal{F}\{\mathcal{R}_u(\tau)\}$$

Furthermore, since $\mathcal{F}\{f(-t)\} = \hat{f}(-\jmath\omega)$, one has

$$\hat{\rho}(\omega) = \mathcal{F}\{H(\tau) * H(-\tau)\} = \hat{H}(\jmath\omega)\hat{H}(-\jmath\omega)$$

Power spectral density

The **spectral density** of a zero-mean stationary random process $\mathbb{r}(t)$ can be *defined* as

$$\mathcal{S}_r(\omega) = \mathcal{F}\{\mathsf{E}[\mathbb{r}(t)\mathbb{r}(t+\tau)]\} = \mathcal{F}\{\mathcal{R}_r(\tau)\}$$

Hence the notation

$$\mathcal{S}_y(\omega) = \hat{H}(\jmath\omega)\hat{H}(-\jmath\omega)\mathcal{S}_u(\omega) = |\hat{H}(\jmath\omega)|^2\mathcal{S}_u(\omega)$$

[7] Einstein was first in 1914!

The spectrum of noise

Colors are defined by their spectral power content. Stationary random signals are no different. The **Fourier transform** is the appropriate tool for this analysis.

Motivating the notion of **white**

White light is special in being made up of a (somewhat) uniform distribution of spectral intensities. Its power spectral density can be seen as *constant*.

$$S_w(\omega) = q$$

The inverse Fourier transform furnishes the autocorrelation of white light as

$$\mathcal{R}_w(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_w(\omega)e^{\jmath\omega\tau}\ d\omega = \frac{q}{2\pi}\int_{-\infty}^{\infty} e^{\jmath\omega\tau}\ d\omega = q\delta(\tau)$$

White noise $\rightarrow \hat{H}(\jmath\omega) \rightarrow$ **Colored noise**

Passing white light through a filter alters its spectral content and gives it color. Let $\mathbb{u}(t)$ be white noise with zero mean $m_u = 0$ and autocorrelation $\mathcal{R}_u(\tau) = q\delta(\tau)$. Then, the output $\mathbb{y}(t)$ filtered by H is equipped with a *colored* spectrum.

$$\mathcal{S}_y(\omega) = |\hat{H}(\jmath\omega)|^2\mathcal{S}_u(\omega) = |\hat{H}(\jmath\omega)|^2 q$$

This operation is referred to as **spectral factorization**.

The spectrum of noise

Colors are defined by their spectral power content. Stationary random signals are no different. The **Fourier transform** is the appropriate tool for this analysis.

Filters & Colors

Color	$H(s)$	$\mathcal{S}(\omega)$	$\mathcal{R}(\tau)$
White	1	1	$\delta(\tau)$
Brown	$\lim_{\epsilon \to 0} \frac{1}{s+\epsilon}$	$\lim_{\epsilon \to 0} \frac{1}{\omega^2+\epsilon^2}$	$\lim_{\epsilon \to 0} \frac{e^{-\epsilon\|\tau\|}}{2\epsilon}$
Violet	$\lim_{\epsilon \to 0} \frac{s}{\epsilon s+1}$	$\lim_{\epsilon \to 0} \frac{\omega^2}{\omega^2\epsilon^2+1}$	$\lim_{\epsilon \to 0} \frac{2\epsilon\delta(\tau)-\Theta(\|\tau\|)e^{-\frac{\|\tau\|}{\epsilon}}}{2\epsilon^3}$
Band-limited	$-$	$\Theta(\omega+\omega_c)-\Theta(\omega-\omega_c)$	$\frac{\sin(\tau\omega_c)}{\pi\tau}$
Low-passed	$\frac{1}{s/\omega_c+1}$	$\frac{1}{(\omega/\omega_c)^2+1}$	$\frac{\omega_c}{2}e^{-\omega_c\|\tau\|}$

Key idea

By passing white noise through one (or more) linear filters, an assortment of colors can be simulated. This technique permits extension of the Kalman filter to cases where the input is not **white** but **colored**.

Model augmentation

The general plant model used by the Kalman filter was given by

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbb{w}, \quad \mathbb{y} = \mathbf{C}\mathbb{x} + \mathbb{v}$$

Here, $\mathbb{v}$ and $\mathbb{w}$ were assumed white. If $\mathbb{w}$ is colored, an augmented state-space can be employed. The notation $\mathbb{v}$ and $\mathbb{w}$ is reserved for white processes. So let the colored disturbance be denoted $\mathbb{d}$, leading to $\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbb{d}$.

Suppose that an element of $\mathbb{d}(t)$ is observed to have the spectrum $\mathcal{S}_d(\omega) = q_w|\hat{H}_d(j\omega)|^2$. The *shaping filter* can be *realized* as

$$\hat{H}_d(\jmath\omega) = H_d(s)|_{s=\jmath\omega}, \quad H_d(s) = \mathbf{c}_d(s\mathbb{I} - \mathbf{A}_d)^{-1}\mathbf{b}_d + d_d$$

In the time-domain, the colored noise is therefore simulated by

$$\dot{\mathbb{x}}_d(t) = \mathbf{A}_d\mathbb{x}_d(t) + \mathbf{b}_d\mathbb{w}(t), \quad \mathbb{d}(t) = \mathbf{c}_d\mathbb{x}_d(t) + d_d\mathbb{w}(t)$$

where $\mathbb{w}$ is a zero-mean white process with variance $\delta(0)q_w$.

Model augmentation

Let the disturbance be modeled (colored) by

$$\dot{\mathbb{x}}_d = \mathbf{A}_d \mathbb{x}_d + \mathbf{B}_d \mathbb{w}, \quad \mathbb{w} = \mathbf{C}_d \mathbb{x}_d + \mathbf{D}_d \mathbb{w}$$

where $\mathbb{w}$ is white with zero mean and variance given by $\mathsf{E}[\mathbb{w}\mathbb{w}^{\mathsf{T}}] = \delta(0)\mathbf{Q}_{\mathbf{w}}$. Then, an augmented state-space model becomes

$$\begin{bmatrix} \dot{\mathbb{x}} \\ \dot{\mathbb{x}}_d \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{G}\mathbf{C}_d \\ \mathbf{0} & \mathbf{A}_d \end{bmatrix} \begin{bmatrix} \mathbb{x} \\ \mathbb{x}_d \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \mathbb{u} + \begin{bmatrix} \mathbf{G}\mathbf{D}_d \\ \mathbf{B}_d \end{bmatrix} \mathbb{w}$$

Note that augmentation need only be done for non-white disturbance processes.

An example

A wind turbine with three blades will experience significant disturbances at the so-called 3P-frequency due to tower-passing. A model of the rotor's velocity around a stable operating point can be furnished by

$$\tau \dot{\mathbb{x}}_1 + \mathbb{x}_1 = \mathbb{Q}$$

The aerodynamic torque driving the plant $\mathbb{Q}$ is here modeled as a random disturbance given by the sum of a slowly-varying component and a periodic disturbance

$$\mathbb{Q} = \mathbb{Q}_0 + \mathbb{Q}_{3P}$$

We wish to estimate $\mathbb{Q}_0$ given the measurement $\mathbb{y} = \mathbb{x} + \mathbb{v}$ where $\mathbb{v}$ is a zero-mean white process of intensity r_v.

The two torque disturbances affect the rotor in the same way, so it would at first glance appear difficult to tease them apart. But, by assuming that they are *shaped* differently[8], progress can be made.

[8] Implying different spectra.

Spectral densities

The slowly varying torque component is well modeled by a random walk $\mathbb{Q}_0 = \mathbb{x}_2$. The random walk can be simulated by

$$\dot{\mathbb{x}}_2 = k_0 \mathbb{w}_1, \quad \mathcal{S}_0(\omega) = \frac{k_0^2}{\omega^2}$$

where $\mathbb{w}_1$ is unbiased white noise of unit intensity (the scaling is done with k_0).

Since the $\mathbb{Q}_{3\text{P}}$ component occurs around the frequency $\omega_{3\text{P}} = 3\Omega$ where Ω represents the rotor's angular velocity, a natural spectrum is furnished by

$$\mathcal{S}_{3\text{P}}(\omega) = \frac{k_{3\text{P}}^2 \omega^2}{\left(\omega^2 - \omega_{3\text{P}}^2\right)^2} = \frac{k_{3\text{P}}^2 \jmath\omega}{(\jmath\omega)^2 + \omega_{3\text{P}}^2} \frac{-k_{3\text{P}}^2 \jmath\omega}{(-\jmath\omega)^2 + \omega_{3\text{P}}^2} = \hat{H}(\jmath\omega)\hat{H}(-\jmath\omega)$$

The appropriate shaping filter is clearly

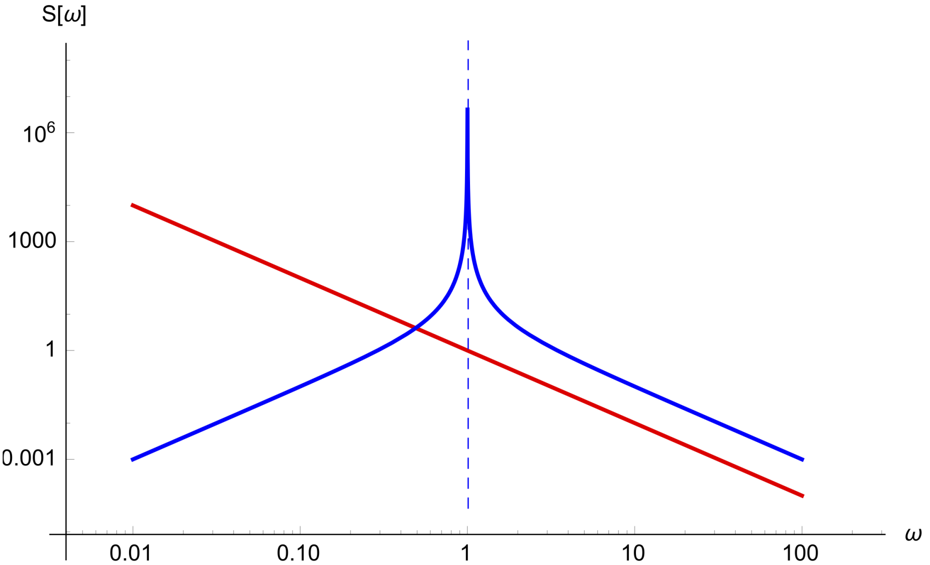
$$H(s) = \frac{k_{3\text{P}} s}{s^2 + \omega_{3\text{P}}^2}$$

Hence the disturbance model

$$\begin{bmatrix} \dot{\mathbb{x}}_3 \\ \dot{\mathbb{x}}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_{3\text{P}}^2 & 0 \end{bmatrix} \begin{bmatrix} \mathbb{x}_3 \\ \mathbb{x}_4 \end{bmatrix} + \begin{bmatrix} 0 \\ k_{3\text{P}} \end{bmatrix} \mathbb{w}_2, \quad \mathbb{Q}_{3\text{P}} = \mathbb{x}_4$$

where $\mathbb{w}_2$ is unbiased white noise of unit intensity (the scaling is done with $k_{3\text{P}}$).

Spectra of brown and monochrome noise



Augmented model

The augmented random process becomes

$$\begin{bmatrix} \dot{\mathbb{x}}_1 \\ \dot{\mathbb{x}}_2 \\ \dot{\mathbb{x}}_3 \\ \dot{\mathbb{x}}_4 \end{bmatrix} = \overbrace{\begin{bmatrix} -\tau^{-1} & \tau^{-1} & 0 & \tau^{-1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega_{3P}^2 & 0 \end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix} \mathbb{x}_1 \\ \mathbb{x}_2 \\ \mathbb{x}_3 \\ \mathbb{x}_4 \end{bmatrix}}^{\mathbb{x}} + \overbrace{\begin{bmatrix} 0 & 0 \\ k_0 & 0 \\ 0 & 0 \\ 0 & k_{3P} \end{bmatrix}}^{\mathbf{G}} \overbrace{\begin{bmatrix} \mathbb{w}_1 \\ \mathbb{w}_2 \end{bmatrix}}^{\mathbb{w}}$$

$$\mathbb{y} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{C}} \begin{bmatrix} \mathbb{x}_1 \\ \mathbb{x}_2 \\ \mathbb{x}_3 \\ \mathbb{x}_4 \end{bmatrix} + \mathbb{v}, \quad \mathbf{Q_w} = \mathbb{I}$$

The pair $(\mathbf{A}, \mathbf{C})$ is required to be observable for the Kalman filter to apply. This is indeed the case.

In order to arrive at the mean torque-component, the following estimate is used

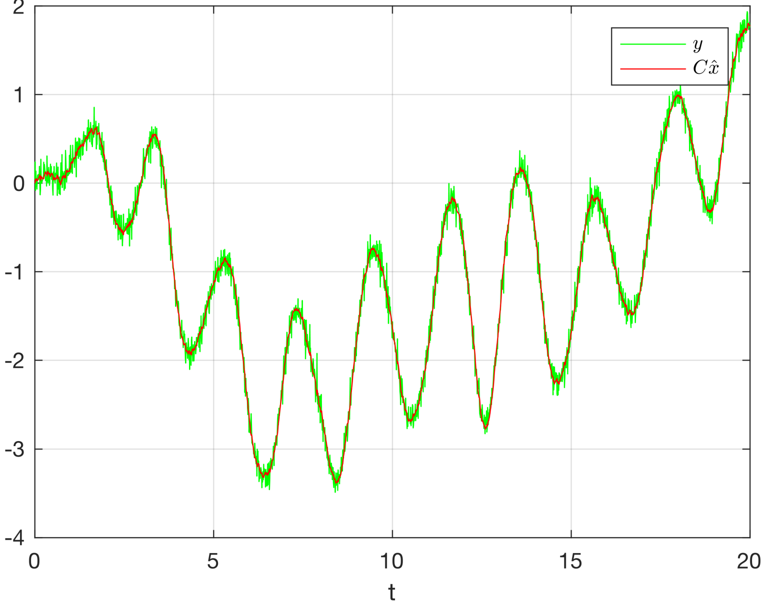
$$\hat{\mathbb{Q}}_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbb{x}}_1 \\ \hat{\mathbb{x}}_2 \\ \hat{\mathbb{x}}_3 \\ \hat{\mathbb{x}}_4 \end{bmatrix}$$

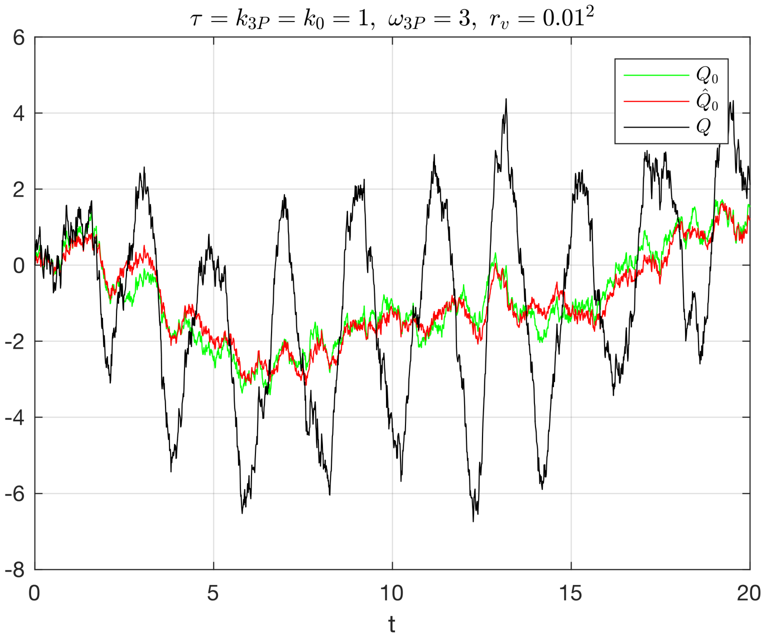
Optimal estimation

An optimal stationary estimator for the states of the uncertain plant can now be obtained from the solution of

$$\mathbf{AP} + \mathbf{PA}^{\mathsf{T}} + \mathbf{GG}^{\mathsf{T}} - \frac{1}{r_v}\mathbf{PC}^{\mathsf{T}}\mathbf{CP} = \mathbf{0}, \quad \mathbf{L} = \frac{1}{r_v}\mathbf{PC}^{\mathsf{T}}$$

$\tau = k_{3P} = k_0 = 1,\ \omega_{3P} = 3,\ r_v = 0.01^2$





$\tau = k_{3P} = k_0 = 1,\ \omega_{3P} = 3,\ r_v = 0.01^2$
6
4
2
0
-2
-4
-6
-8
0
5
10
15
20
t
Q_0
$\hat{Q}_0$
Q

Topic

Physical model

The plant model is described by the random process

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{Bu} + \mathbf{G}\mathbb{w}, \quad \mathbb{y} = \mathbf{C}\mathbb{x} + \mathbb{v}$$

where $\mathcal{C}_{\mathbf{w}} = \delta(0)\mathbf{Q}_{\mathbf{w}}$ and $\mathcal{C}_{\mathbf{v}} = \delta(0)\mathbf{R}_{\mathbf{v}}$. If there are off-diagonal elements in $\mathbf{Q}_{\mathbf{w}}$ and $\mathbf{R}_{\mathbf{v}}$, the elements of the respective random vectors are correlated.

Diagonalization

It is in practice useful to represent the noise $\mathbb{v}$ and disturbance $\mathbb{w}$ in terms of uncorrelated sequences. This is achieved by diagonalizing the covariance matrices. Suppose that $\mathbf{M}$ is a symmetric matrix. Let $\boldsymbol{\Lambda}_M$ denote a diagonal matrix of real[9] eigenvalues and let $\mathbf{E}_M$ describe the corresponding matrix of orthonormal[10] eigenvectors. Then, the matrix can be represented as

$$\mathbf{M} = \mathbf{E}_M \boldsymbol{\Lambda}_M \mathbf{E}_M^\mathsf{T}, \quad \mathbf{E}_M^\mathsf{T} \mathbf{E}_M = \mathbb{I}$$

Diagonalized representation

Let the covariance matrices be diagonalized

$$\mathbf{R}_\mathbf{v} = \mathbf{E}_\mathbf{v} \boldsymbol{\Lambda}_\mathbf{v} \mathbf{E}_\mathbf{v}^\mathsf{T}, \quad \mathbf{Q}_\mathbf{w} = \mathbf{E}_\mathbf{w} \boldsymbol{\Lambda}_\mathbf{w} \mathbf{E}_\mathbf{w}^\mathsf{T}$$

The model can now be simulated with

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbf{E}_\mathbf{w}\mathbb{w}', \quad \mathbb{y} = \mathbf{C}\mathbb{x} + \mathbf{E}_\mathbf{v}\mathbb{v}'$$

where the covariances have been diagonalized so that $\mathcal{C}_{\mathbf{w}'} = \delta(0)\boldsymbol{\Lambda}_\mathbf{w}$ and $\mathcal{C}_{\mathbf{v}'} = \delta(0)\boldsymbol{\Lambda}_\mathbf{v}$. The entries in $\mathbb{v}'$ and $\mathbb{w}'$ now represent *independent* processes. The variance of each entry can be read off the diagonals in the eigenvalue matrices. This permits far easier simulation.

[9] The eigenvalues of a symmetric matrix are always real.

[10] The orthonormal column vectors $\mathbf{e}_i$ making up $\mathbf{E}$ satisfy $\mathbf{e}_i^\mathsf{T}\mathbf{e}_j = \delta[i, j]$. Symmetric matrices always have orthogonal eigenvectors, the rest is a matter of scaling.

Topic

Discrete time Kalman filter

Measurements $\mathbf{y}$ are typically obtained through sampling at discrete intervals in time $t = kT,\ k = 0, 1, 2, \ldots$. Furthermore, estimates $\hat{\mathbf{x}}$ will typically be requested at discrete intervals. For these reasons (and others), the discrete Kalman filter is the version that sees most frequent use (by far).

Discrete time analysis

The passage from continuous to discrete time introduces a range of changes, some of which are quite subtle.

Continuous time random process

The continuous time plant model is given by the random process

$$\dot{\mathbb{x}} = \mathbf{A}\mathbb{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbb{w}, \quad \mathbb{y} = \mathbf{C}\mathbb{x} + \mathbb{v}$$

where the noise and disturbance are unbiased ($\mathbf{m_v} = \mathbf{0},\ \mathbf{m_w} = \mathbf{0}$) and white

$$\mathcal{A}_{\mathbf{v}}(t,\tau) = \mathrm{E}[\mathbb{v}(t)\mathbb{v}(\tau)^{\mathsf{T}}] = \delta(t-\tau)\mathbf{R}, \quad \mathcal{A}_{\mathbf{w}}(t,\tau) = \mathrm{E}[\mathbb{w}(\tau)\mathbb{w}(t)^{\mathsf{T}}] = \delta(t-\tau)\mathbf{Q}$$

Exact solution

Knowing the solution permits exact discretization. For the process model given above, an **exact** solution is furnished by

$$\mathbb{x}(t) = e^{\mathbf{A}t}\mathbb{x}_0 + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\, d\tau + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{G}\mathbb{w}(\tau)\, d\tau, \quad \mathbb{y}(t) = \mathbf{C}\mathbb{x}(t) + \mathbb{v}(t)$$

Exact solution

$$\mathbb{x}(t) = e^{\mathbf{A}t}\mathbb{x}_0 + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\, d\tau + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{G}\mathbb{w}(\tau)\, d\tau, \quad \mathbb{y}(t) = \mathbf{C}\mathbb{x}(t) + \mathbb{v}(t)$$

Discretization

Starting the solution at $t = kT$ and terminating it at $t = (k+1)T$ produces

$$\mathbb{x}[k+1] = e^{\mathbf{A}T}\mathbb{x}[k] + \int_0^T e^{\mathbf{A}\alpha}\mathbf{B}\mathbf{u}((k+1)T - \alpha)\, d\alpha + \int_0^T e^{\mathbf{A}\alpha}\mathbf{G}\mathbb{w}((k+1)T - \alpha)\, d\alpha$$

A calculation variable $\alpha = (k+1)T - \tau$ is here introduced to make life easier.

$$\bar{\mathbf{A}} \triangleq e^{\mathbf{A}T}, \quad \bar{\mathbf{B}} \triangleq \int_0^T e^{\mathbf{A}\alpha}\mathbf{B}\, d\alpha, \quad \bar{\mathbb{w}}[k] \triangleq \int_0^T e^{\mathbf{A}\alpha}\mathbf{G}\mathbb{w}((k+1)T - \alpha)\, d\alpha$$

Assuming that the deterministic input $\mathbf{u}(t)$ varies little over $(k+1)T \leq t \leq (k+1)T$ yields the discretized model

$$\mathbb{x}[k+1] = \bar{\mathbf{A}}\mathbb{x}[k] + \bar{\mathbf{B}}\mathbf{u}[k] + \bar{\mathbb{w}}[k]$$

Note well that the discretized noise contribution is quite different from the continuous time variety, $\bar{\mathbb{w}}[k] \neq \mathbb{w}(kT)$.

Discrete time white **disturbances**

The discrete time white disturbance signal is now subjected to a closer examination.

$$\bar{\mathrm{w}}[k] \triangleq \int_0^T e^{\mathbf{A}\alpha}\mathbf{G}\mathrm{w}((k+1)T-\alpha)\,d\alpha$$

It is straightforward to verify that $\bar{\mathrm{w}}[k]$ inherits the unbiased nature of $\mathrm{w}(t)$. But, the autocovariance (incl. variance) changes in a subtle fashion. The discrete time autocovariance of $\bar{\mathrm{w}}[k]$ is given by

$$\begin{aligned}\bar{\mathcal{A}}_{\mathbf{w}}[k,l] &= \mathrm{E}[\bar{\mathrm{w}}[k]\bar{\mathrm{w}}[l]^{\mathsf{T}}] \\ &= \int_0^T\int_0^T e^{\mathbf{A}\alpha_1}\mathbf{G}\underbrace{\mathrm{E}[\mathrm{w}((k+1)T-\alpha_1)\mathrm{w}((l+1)T-\alpha_2)^{\mathsf{T}}]}_{\mathcal{A}_{\mathbf{w}}((k+1)T-\alpha_1,(l+1)T-\alpha_2)=\delta((l-k)T+\alpha_1-\alpha_2)\mathbf{Q}_{\mathbf{w}}}\mathbf{G}^{\mathsf{T}}e^{\mathbf{A}^{\mathsf{T}}\alpha_2}\,d\alpha_1\,d\alpha_2\end{aligned}$$

Kronecker's δ-function satisfies

$$\delta[k,l] = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

Noting that $\delta((l-k)T+\alpha_1-\alpha_2) = \delta[k,l]\delta(\alpha_1-\alpha_2)$ the result follows

$$\bar{\mathcal{A}}_{\mathbf{w}}[k,l] = \delta[k,l]\bar{\mathbf{Q}}_{\mathbf{w}}, \quad \bar{\mathbf{Q}}_{\mathbf{w}} \triangleq \int_0^T e^{\mathbf{A}\alpha}\mathbf{G}\mathbf{Q}_{\mathbf{w}}\mathbf{G}^{\mathsf{T}}e^{\mathbf{A}^{\mathsf{T}}\alpha}\,d\alpha$$

Exact discretization has rendered the infinite variance of $\mathrm{w}(t)$ *finite* and equal to $\bar{\mathcal{A}}_{\mathbf{w}}[k,k] = \bar{\mathbf{Q}}_{\mathbf{w}}$ in discrete time (this in fact a consequence of the *central limit theorem*).

Discrete time white **noise**

The measurement model in continuous time is given by

$$\mathbb{y} = \mathbf{C}\mathbb{x} + \mathbb{v}, \quad \mathcal{A}_{\mathbf{v}}(t,\tau) = \mathsf{E}[\mathbb{v}(t)\mathbb{v}(\tau)^{\mathsf{T}}] = \delta(t-\tau)\mathbf{R}_{\mathbf{v}}$$

A naïve conversion to discrete time would suggest

$$\mathbb{y}[k] = \mathbf{C}\mathbb{x}[k] + \mathbb{v}[k], \quad \bar{\mathbf{R}}_{\mathbf{v}} = \mathsf{E}[\mathbb{v}[k]\mathbb{v}[k]^{\mathsf{T}}] = \delta(0)\mathbf{R}_{\mathbf{v}}$$

This interpretation leads to extreme exaggerations of noise in discrete time (but is suitable in continuous time).

Averaging convention

Rather than interpreting measurement noise as occuring at the instant of sampling, it can be interpreted in a *averaged* sense. This idea is captured in the convention

$$\bar{\mathbb{v}}[k] \triangleq \frac{1}{T}\int_0^T \mathbb{v}(kT-\alpha)\, d\alpha$$

The discrete time noise vector inherits the unbiased nature of the continuous time signal, whilst the autocovariance transforms to

$$\bar{\mathcal{A}}_{\mathbf{v}}[k,l] = \mathsf{E}[\bar{\mathbb{v}}[k]\bar{\mathbb{v}}[l]^{\mathsf{T}}] = \frac{1}{T^2}\int_0^T\int_0^T \mathsf{E}[\mathbb{v}(kT-\alpha_1)\mathbb{v}(lT-\alpha_2)^{\mathsf{T}}]\, d\alpha_1 d\alpha_2 = \delta[k,l]\bar{\mathbf{R}}_{\mathbf{v}}, \quad \bar{\mathbf{R}}_{\mathbf{v}} \triangleq \mathbf{R}_{\mathbf{v}}/T$$

Discrete time random process

The discrete time plant model is given by the random process

$$\mathbb{x}[k+1] = \bar{\mathbf{A}}\mathbb{x}[k] + \bar{\mathbf{B}}\mathbf{u}[k] + \bar{\mathbb{w}}[k], \quad \mathbb{y}[k] = \mathbf{C}\mathbb{x}[k] + \bar{\mathbb{v}}[k]$$

where the noise and disturbance are unbiased ($\mathbf{m_v} = \mathbf{0},\ \mathbf{m_w} = \mathbf{0}$) and white

$$\bar{\mathcal{A}}_{\mathbf{v}}[k,l] = \mathrm{E}[\bar{\mathbb{v}}[k]\bar{\mathbb{v}}[l]^{\mathsf{T}}] = \delta[k,l]\bar{\mathbf{R}}_{\mathbf{v}}, \quad \bar{\mathcal{A}}_{\mathbf{w}}[k,l] = \mathrm{E}[\bar{\mathbb{w}}[k]\bar{\mathbb{w}}[l]^{\mathsf{T}}] = \delta[k,l]\bar{\mathbf{Q}}_{\mathbf{w}}$$

It will be assumed that the noise and disturbance processes are uncorrelated $\mathrm{E}[\bar{\mathbb{v}}[k]\bar{\mathbb{w}}[l]^{\mathsf{T}}] = \mathbf{0}$.

Continuous to discrete conversion - sampling time T.

Transition matrix: Obtained from exact discretization.

$$\bar{\mathbf{A}} = e^{\mathbf{A}T}$$

Input matrix: Obtained from exact discretization & assumption of constant **u** over sampling period.

$$\bar{\mathbf{B}} = \int_0^T e^{\mathbf{A}\alpha}\mathbf{B}\, d\alpha$$

Disturbance covariance: Obtained from exact discretization.

$$\bar{\mathbf{Q}}_{\mathbf{w}} = \int_0^T e^{\mathbf{A}\alpha}\mathbf{G}\mathbf{Q}_{\mathbf{w}}\mathbf{G}^{\mathsf{T}} e^{\mathbf{A}^{\mathsf{T}}\alpha}\, d\alpha$$

Noise covariance: Obtained through an averaging convention.

$$\bar{\mathbf{R}}_{\mathbf{v}} = \mathbf{R}_{\mathbf{v}}/T$$

Van Loan's method[11]

The integrals in the preceding slide are often quite intractable. It is however possible to arrive at the correct matrices without integrating. This is done with *Van Loan's method*. The key result is

$$\exp\left(\begin{bmatrix} \mathbf{A} & \mathbf{G}\mathbf{Q}_\mathbf{w}\mathbf{G}^\mathsf{T} \\ \mathbf{0} & -\mathbf{A}^\mathsf{T} \end{bmatrix} T\right) = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{0} & \mathbf{M}_{22} \end{bmatrix}, \quad \bar{\mathbf{A}} = \mathbf{M}_{11}, \quad \bar{\mathbf{Q}}_\mathbf{w} = \mathbf{M}_{12}\mathbf{M}_{11}^\mathsf{T}$$

Matrix exponentials are readily computed numerically, obviating the need for integration. The input matrix can be computed from

$$\exp\left(\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} T\right) = \begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{0} & \mathbb{I} \end{bmatrix}, \quad \bar{\mathbf{A}} = \mathbf{N}_{11}, \quad \bar{\mathbf{B}} = \mathbf{N}_{12}$$

[11] Van Loan C.F. (1978), *Computing Integrals Involving the Matrix Exponential*, IEEE Transactions on Automatic Control, Vol. 23, No. 3

Topic

Discrete observer

Discrete time requires a sligthly more explicit observer design. The estimate is generated in two distinct phases.

---

1 - A priori (denoted $\hat{\mathbb{x}}^-[k]$): The best guess for $\mathbb{x}[k]$ **prior** to incorporation of the measurement $\mathbb{y}[k]$. The deterministic model is to arrive at this estimate.

$$\hat{\mathbb{x}}^-[k] = \bar{\mathbf{A}}\hat{\mathbb{x}}[k-1] + \bar{\mathbf{B}}\mathbf{u}[k-1]$$

---

2 - A posteriori (denoted $\hat{\mathbb{x}}[k]$): The best guess for $\mathbb{x}[k]$ **after** incorporation of the measurement $\mathbb{y}[k]$. A linear blend of what the model suggests ($\hat{\mathbb{x}}^-[k]$) and the new measurement $\mathbb{y}[k]$ is used to arrive at this final estimate. The Kalman gain $\mathbf{L}[k]$ serves as the *blending factor*, viz.

$$\hat{\mathbb{x}}[k] = \hat{\mathbb{x}}^-[k] + \mathbf{L}[k](\mathbb{y}[k] - \mathbf{C}\hat{\mathbb{x}}^-[k])$$

Kalman gain

The Kalman gain is (as for the continuous time case) designed to minimize the mean-square error of the estimate at time k.

$$J[k] = \mathrm{tr}(\mathbf{P}[k]), \quad \mathbf{P}[k] \triangleq \mathrm{E}[(\mathbb{x}[k] - \hat{\mathbb{x}}[k])(\mathbb{x}[k] - \hat{\mathbb{x}}[k])^{\mathsf{T}}]$$

A **priori** error and covariance matrix

The a priori and a posteriori estimation errors and covariance matrices are given by

$$\mathbb{e}^-[k] \triangleq \mathbb{x}[k] - \hat{\mathbb{x}}^-[k], \quad \mathbf{P}^-[k] \triangleq \mathsf{E}[\mathbb{e}^-[k]\mathbb{e}^-[k]^\mathsf{T}]$$
$$\mathbb{e}[k] \triangleq \mathbb{x}[k] - \hat{\mathbb{x}}[k], \quad \mathbf{P}[k] \triangleq \mathsf{E}[\mathbb{e}_k \mathbb{e}_k^\mathsf{T}]$$

The process model produces the following state at k

$$\mathbb{x}[k] = \bar{\mathbf{A}}\mathbb{x}[k-1] + \bar{\mathbf{B}}\mathbb{u}[k-1] + \bar{\mathbb{w}}[k-1]$$

whereas the a priori estimate reads as

$$\hat{\mathbb{x}}^-[k] = \bar{\mathbf{A}}\hat{\mathbb{x}}[k-1] + \bar{\mathbf{B}}\mathbf{u}[k-1]$$

This permits the following expression for the **a priori** error

$$\mathbb{e}^-[k] = \bar{\mathbf{A}}\mathbb{e}[k-1] + \bar{\mathbb{w}}[k-1]$$

The **a priori** covariance matrix follows as

$$\mathbf{P}^-[k] = \mathsf{E}[(\bar{\mathbf{A}}\mathbb{e}[k-1] + \bar{\mathbb{w}}[k-1])(\bar{\mathbf{A}}\mathbb{e}[k-1] + \bar{\mathbb{w}}[k-1])^\mathsf{T}] = \bar{\mathbf{A}}\mathbf{P}[k-1]\bar{\mathbf{A}}^\mathsf{T} + \bar{\mathbf{Q}}_\mathbf{w}$$

The disturbance at k is uncorrelated to the a-posteriori estimate at k, hence $\mathsf{E}[\mathbb{e}[k]\bar{\mathbb{w}}[k]^\mathsf{T}] = \mathbf{0}$.

A **posteriori** error and covariance matrix

The a priori and a posteriori estimation errors and covariance matrices are given by

$$\mathbb{e}^-[k] \triangleq \mathbb{x}[k] - \hat{\mathbb{x}}^-[k], \quad \mathbf{P}^-[k] \triangleq \mathsf{E}[\mathbb{e}^-[k]\mathbb{e}^-[k]^\mathsf{T}]$$
$$\mathbb{e}[k] \triangleq \mathbb{x}[k] - \hat{\mathbb{x}}[k], \quad \mathbf{P}[k] \triangleq \mathsf{E}[\mathbb{e}_k\mathbb{e}_k^\mathsf{T}]$$

The a posteriori estimate can be expanded to read

$$\hat{\mathbb{x}}[k] = \hat{\mathbb{x}}^-[k] + \mathbf{L}[k](\mathbb{y}[k] - \mathbf{C}\hat{\mathbb{x}}^-[k]) = \hat{\mathbb{x}}^-[k] + \mathbf{L}[k]\mathbf{C}\mathbb{e}^-[k] + \mathbf{L}[k]\bar{\mathbb{v}}[k]$$

This permits the following expression for the **a posteriori** error

$$\mathbb{e}[k] = (\mathbb{I} - \mathbf{L}[k]\mathbf{C})\mathbb{e}^-[k] + \mathbf{L}[k]\bar{\mathbb{v}}[k]$$

The **a posteriori** covariance matrix follows as

$$\begin{aligned}\mathbf{P}[k] = \mathsf{E}[((\mathbb{I} - \mathbf{L}[k]\mathbf{C})\mathbb{e}^-[k] + \mathbf{L}[k]\bar{\mathbb{v}}[k])((\mathbb{I} - \mathbf{L}[k]\mathbf{C})\mathbb{e}^-[k] + \mathbf{L}[k]\bar{\mathbb{v}}[k])^\mathsf{T}]&\\ = (\mathbb{I} - \mathbf{L}[k]\mathbf{C})\mathbf{P}^-[k](\mathbb{I} - \mathbf{L}[k]\mathbf{C})^\mathsf{T} + \mathbf{L}[k]\bar{\mathbf{R}}_\mathbf{v}[k]\mathbf{L}[k]^\mathsf{T}&\end{aligned}$$

The noise at k is uncorrelated to the a-priori estimate at k, hence $\mathsf{E}[\mathbb{e}^-[k]\bar{\mathbb{v}}[k]^\mathsf{T}] = \mathbf{0}$.

Optimal estimation

The **a posteriori** covariance matrix describes the covariance of the final estimate error $\mathbb{e}[k] = \mathbb{x}[k] - \hat{\mathbb{x}}[k]$. We now seek to minimize the mean-square error

$$J[k] = \text{tr}(\mathbf{P}[k])$$

Differentiation w.r.t. to the Kalman gain and solving for the extremum yields yields

$$\begin{aligned}\frac{\partial \text{tr}(\mathbf{P}[k])}{\partial \mathbf{L}[k]} &= \frac{\partial}{\partial \mathbf{L}[k]} \text{tr}\left((\mathbb{I} - \mathbf{L}[k]\mathbf{C})\mathbf{P}^{-}[k](\mathbb{I} - \mathbf{L}[k]\mathbf{C})^{\mathsf{T}} + \mathbf{L}[k]\bar{\mathbf{R}}_{\mathbf{v}}[k]\mathbf{L}[k]^{\mathsf{T}} \right) \\ &= -2\mathbf{P}^{-}[k]\mathbf{C}^{\mathsf{T}} + 2\mathbf{L}[k](\mathbf{C}\mathbf{P}^{-}[k]\mathbf{C}^{\mathsf{T}} + \bar{\mathbf{R}}_{\mathbf{v}}) = \mathbf{0}\end{aligned}$$

The *Kalman gain* thus follows as

$$\mathbf{L}[k] = \mathbf{P}^{-}[k]\mathbf{C}^{\mathsf{T}}(\mathbf{C}\mathbf{P}^{-}[k]\mathbf{C}^{\mathsf{T}} + \bar{\mathbf{R}}_{\mathbf{v}})^{-1}$$

The filter is initialized at

$$\hat{\mathbf{x}}^-[0] = \mathrm{E}[\mathbb{x}(0)] = \mathbf{m}_{\mathbf{x}_0}$$
$$\mathbf{P}^-[0] = \mathrm{E}[\mathbb{e}^-[0]\mathbb{e}^-[0]^\mathsf{T}] = \mathrm{E}[(\mathbb{x}[0] - \mathbf{m}_{\mathbf{x}_0})(\mathbb{x}[0] - \mathbf{m}_{\mathbf{x}_0})^\mathsf{T}] = \mathcal{C}_{\mathbf{x}_0}$$

The recursive algorithm running over $k = 0 \ldots K$ is summarized by

1 - Compute Kalman gain

$$\mathbf{L}[k] = \mathbf{P}^-[k]\mathbf{C}^\mathsf{T}(\mathbf{C}\mathbf{P}^-[k]\mathbf{C}^\mathsf{T} + \bar{\mathbf{R}}_\mathbf{v})^{-1}$$

---

2 - Update estimate with measurement

$$\hat{\mathbb{x}}[k] = \hat{\mathbb{x}}^-[k] + \mathbf{L}[k](\mathbb{y}[k] - \mathbf{C}\hat{\mathbb{x}}^-[k])$$

---

3 - Update error covariance matrix

$$\mathbf{P}[k] = (\mathbb{I} - \mathbf{L}[k]\mathbf{C})\mathbf{P}^-[k](\mathbb{I} - \mathbf{L}[k]\mathbf{C})^\mathsf{T} + \mathbf{L}[k]\bar{\mathbf{R}}_\mathbf{v}[k]\mathbf{L}[k]^\mathsf{T}$$

---

4 - Project ahead

$$\hat{\mathbb{x}}^-[k+1] = \bar{\mathbf{A}}\hat{\mathbb{x}}[k] + \bar{\mathbf{B}}\mathbf{u}[k]$$
$$\mathbf{P}^-[k+1] = \bar{\mathbf{A}}\mathbf{P}[k]\bar{\mathbf{A}}^\mathsf{T} + \bar{\mathbf{Q}}_\mathbf{w}$$

...repeat with $k = k + 1$...

Handheld GPS

Problem

GPS measurements are typically available at a sample time $T \sim 1[s]$. It is assumed that the horisontal measurements are approximately normally distributed around the true position $\mathbb{p} = \text{col}[\mathbb{p}_1, \mathbb{p}_2]$ with a standard deviation $\sigma_v \sim 5[m]$. A measurement model is thus

$$\mathbb{y}[k] = \mathbb{p}[k] + \mathbb{v}[k], \quad \bar{\mathbf{R}} = \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}$$

---

1. How can one improve upon the direct measurement?
2. Is it possible to obtain velocity estimates?

Solution

The desired improvements can be had be incorporating system knowledge. The position of the handheld GPS unit will change in a manner that cannot be predicted exactly. We assume instead that the user moves in accordance with the random model

$$\tau \ddot{\mathbb{p}}_1 + \dot{\mathbb{p}}_1 = \mathbb{w}_1$$
$$\tau \ddot{\mathbb{p}}_2 + \dot{\mathbb{p}}_2 = \mathbb{w}_2$$

Note that the velocities $\dot{\mathbb{p}}$ enters as states of the model and can therefore be estimated. Physically, this model represents a mass-damper perturbed by an unknown force.

Continuous time random process

The intensities of the disturbance signals and the time-constant τ should be tuned through practical experiments. A useful model structure can however be supplied as

$$\overbrace{\begin{bmatrix} \dot{\mathbb{p}}_1 \\ \dot{\mathbb{p}}_2 \\ \ddot{\mathbb{p}}_1 \\ \ddot{\mathbb{p}}_2 \end{bmatrix}}^{\dot{\mathbb{x}}} = \overbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\tau^{-1} & 0 \\ 0 & 0 & 0 & -\tau^{-1} \end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix} \mathbb{p}_1 \\ \mathbb{p}_2 \\ \dot{\mathbb{p}}_1 \\ \dot{\mathbb{p}}_2 \end{bmatrix}}^{\mathbb{x}} + \overbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \tau^{-1} & 0 \\ 0 & \tau^{-1} \end{bmatrix}}^{\mathbf{G}} \overbrace{\begin{bmatrix} \mathbb{w}_1 \\ \mathbb{w}_2 \end{bmatrix}}^{\mathbb{w}}$$

$$\underbrace{\begin{bmatrix} \mathbb{y}_1 \\ \mathbb{y}_2 \end{bmatrix}}_{\mathbb{y}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{\mathbf{C}} \begin{bmatrix} \mathbb{p}_1 \\ \mathbb{p}_2 \\ \dot{\mathbb{p}}_1 \\ \dot{\mathbb{p}}_2 \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbb{v}_1 \\ \mathbb{v}_2 \end{bmatrix}}_{\mathbb{v}}$$

where

$$\mathbf{Q} = q \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{R} = \sigma_v^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} T$$

Discrete time random process

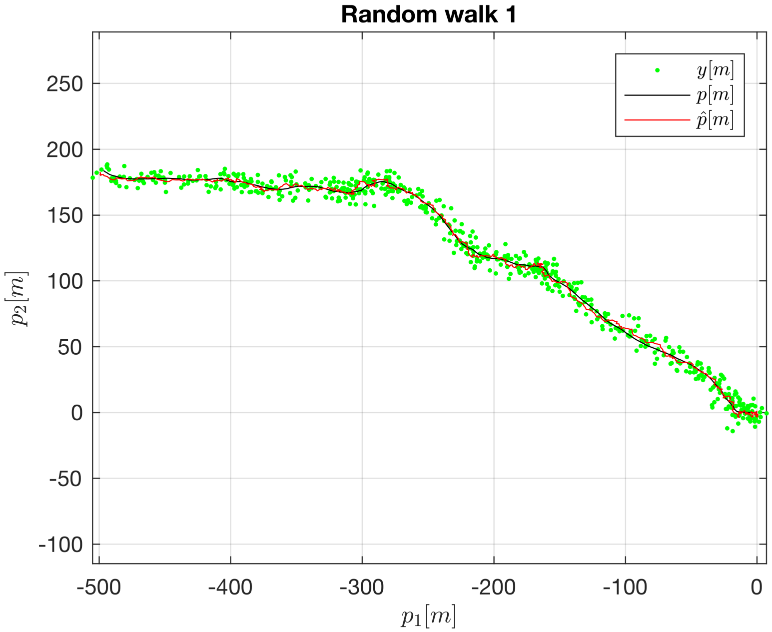
Using Van Loan's method the discrete time system matrices $\bar{\mathbf{A}}$ and $\bar{\mathbf{Q}}$ can be found precisely. The final model reads as

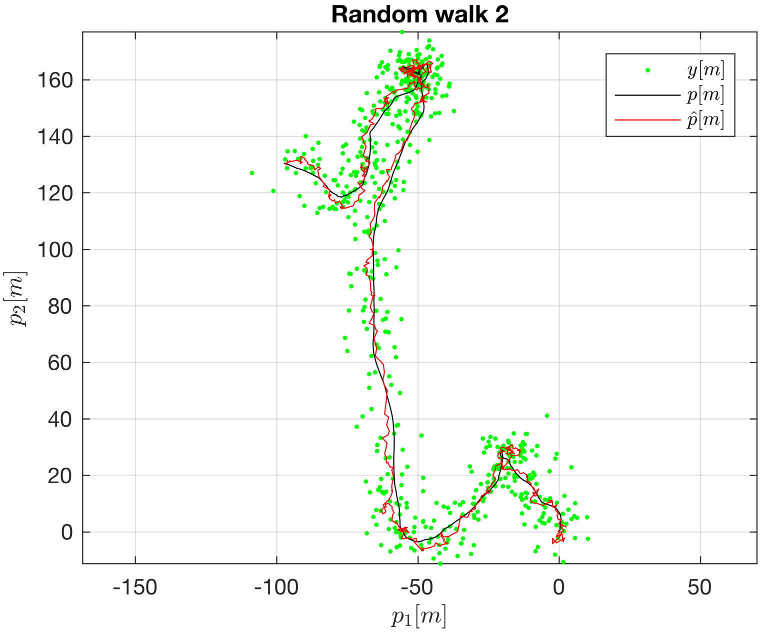
$$\mathbb{x}[k+1] = \bar{\mathbf{A}}\mathbb{x}[k] + \bar{\mathbb{w}}[k], \quad \mathbb{y}[k] = \mathbf{C}\mathbb{x}[k] + \bar{\mathbb{v}}[k]$$

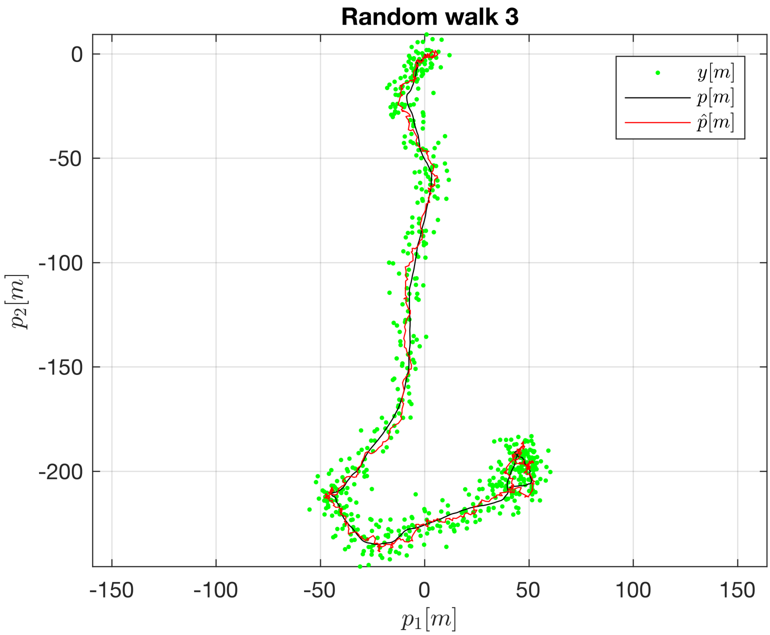
where $\bar{\mathbf{Q}}$ and $\bar{\mathbf{R}}$ describe the respective covariances of the disturbance and noise signals.

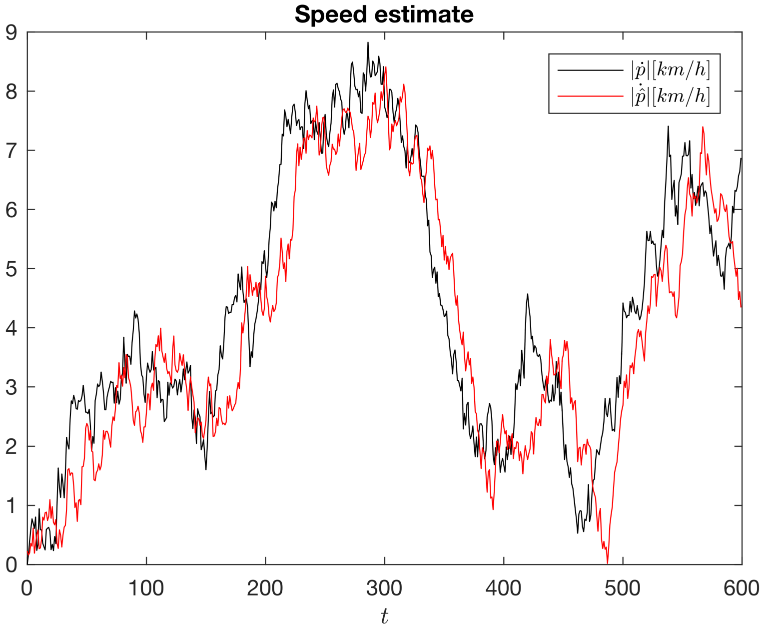
Matlab Demo

The Handheld GPS problem is solved using a discrete time Kalman filter. Tuning constants are chosen as $\tau = 200$ and $q = 25^2$.


Random walk 1
250
200
150
100
50
0
-50
-100
-500
-400
-300
-200
-100
0
$p_1[m]$
$p_2[m]$
$y[m]$
$p[m]$
$\hat{p}[m]$


Random walk 2
$y[m]$
$p[m]$
$\hat{p}[m]$
160
140
120
100
80
60
40
20
0
-150
-100
-50
0
50
$p_1[m]$
$p_2[m]$


Random walk 3
$y[m]$
$p[m]$
$\hat{p}[m]$
0
-50
-100
-150
-200
$p_2[m]$
-150
-100
-50
0
50
100
150
$p_1[m]$


Speed estimate
9
8
7
6
5
4
3
2
1
0
0
100
200
300
400
500
600
t
$|\dot{p}|[km/h]$
$|\dot{\hat{p}}|[km/h]$

Topic

Time varying models

The continuous and discrete time Kalman filters are not limited to time-invariant plants. They can in fact serve as optimal estimators for time-varying systems!

LTV system

Let a linear time-varying random process be given by

$$\dot{\mathbb{x}}(t) = \mathbf{A}(t)\mathbb{x}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{G}(t)\mathbb{w}(t), \quad \mathbb{y}(t) = \mathbf{C}(t)\mathbb{x}(t) + \mathbb{v}(t)$$

where the noise and disturbance are unbiased ($\mathbf{m_v} = \mathbf{0},\ \mathbf{m_w} = \mathbf{0}$) and white

$$\mathcal{A}_\mathbf{v}(t,\tau) = \mathrm{E}[\mathbb{v}(t)\mathbb{v}(\tau)^\mathsf{T}] = \delta(t-\tau)\mathbf{R}(t), \quad \mathcal{A}_\mathbf{w}(t,\tau) = \mathrm{E}[\mathbb{w}(\tau)\mathbb{w}(t)^\mathsf{T}] = \delta(t-\tau)\mathbf{Q}(t)$$

Optimal estimator[12]

An optimal estimator for the LTV process is given by

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}(t)\hat{\mathbf{x}}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{L}(t)(\mathbf{y}(t) - \mathbf{C}(t)\hat{\mathbf{x}}(t)), \quad \mathbf{L}(t) = \mathbf{P}(t)\mathbf{C}^\mathsf{T}(t)\mathbf{R}_\mathbf{v}^{-1}(t)$$

The covariance matrix is here computed by solving the Riccati Equation

$$\dot{\mathbf{P}}(t) = \mathbf{A}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}(t)^\mathsf{T} + \mathbf{G}(t)\mathbf{Q_w}(t)\mathbf{G}(t)^\mathsf{T} - \mathbf{P}(t)\mathbf{C}(t)^\mathsf{T}\mathbf{R_v}(t)^{-1}\mathbf{C}(t)\mathbf{P}(t)$$

[12] Simulation of the continuous time Riccati equation can be challenging. This is one of the reasons that a discrete time formulation is preferred.

Time varying models

The continuous and discrete time Kalman filters are not limited to time-invariant plants. They can in fact serve as optimal estimators for time-varying systems!

DLTV system

Let a discrete time-varying random process plant model be given by

$$\mathbb{x}[k+1] = \bar{\mathbf{A}}[k]\mathbb{x}[k] + \bar{\mathbf{B}}[k]\mathbf{u}[k] + \bar{\mathbb{w}}[k], \quad \mathbb{y}[k] = \mathbf{C}[k]\mathbb{x}[k] + \bar{\mathbb{v}}[k]$$

where the noise and disturbance are unbiased ($\mathbf{m_v} = \mathbf{0},\ \mathbf{m_w} = \mathbf{0}$) and white

$$\bar{\mathcal{A}}_{\mathbf{v}}[k,l] = \mathrm{E}[\bar{\mathbb{v}}[k]\bar{\mathbb{v}}[l]^{\mathsf{T}}] = \delta[k,l]\bar{\mathbf{R}}_{\mathbf{v}}[k], \quad \bar{\mathcal{A}}_{\mathbf{w}}[k,l] = \mathrm{E}[\bar{\mathbb{w}}[k]\bar{\mathbb{w}}[l]^{\mathsf{T}}] = \delta[k,l]\bar{\mathbf{Q}}_{\mathbf{w}}[k]$$

Optimal estimator

The optimal estimator for the preceding system is furnished, quite simply, by letting the matrices in the Kalman filter algorithm be time-varying.

Kalman filter algorithm, general case

The filter is initialized at

$$\hat{\mathbf{x}}^-[0] = \mathrm{E}[\mathbb{x}(0)] = \mathbf{m}_{\mathbf{x}_0}$$
$$\mathbf{P}^-[0] = \mathrm{E}[\mathbb{e}^-[0]\mathbb{e}^-[0]^\mathsf{T}] = \mathrm{E}[(\mathbb{x}[0] - \mathbf{m}_{\mathbf{x}_0})(\mathbb{x}[0] - \mathbf{m}_{\mathbf{x}_0})^\mathsf{T}] = \mathcal{C}_{\mathbf{x}_0}$$

The recursive algorithm running over $k = 0 \ldots K$ is summarized by

1 - Compute Kalman gain

$$\mathbf{L}[k] = \mathbf{P}^-[k]\mathbf{C}[k]^\mathsf{T}(\mathbf{C}[k]\mathbf{P}^-[k]\mathbf{C}[k]^\mathsf{T} + \bar{\mathbf{R}}_\mathbf{v}[k])^{-1}$$

---

2 - Update estimate with measurement

$$\hat{\mathbf{x}}[k] = \hat{\mathbf{x}}^-[k] + \mathbf{L}[k](\mathbf{y}[k] - \mathbf{C}[k]\hat{\mathbf{x}}^-[k])$$

---

3 - Update error covariance matrix

$$\mathbf{P}[k] = (\mathbb{I} - \mathbf{L}[k]\mathbf{C}[k])\mathbf{P}^-[k](\mathbb{I} - \mathbf{L}[k]\mathbf{C}[k])^\mathsf{T} + \mathbf{L}[k]\bar{\mathbf{R}}_\mathbf{v}[k]\mathbf{L}[k]^\mathsf{T}$$

---

4 - Project ahead

$$\hat{\mathbf{x}}^-[k+1] = \bar{\mathbf{A}}[k]\hat{\mathbf{x}}[k] + \bar{\mathbf{B}}[k]\mathbf{u}[k]$$
$$\mathbf{P}^-[k+1] = \bar{\mathbf{A}}[k]\mathbf{P}[k]\bar{\mathbf{A}}[k]^\mathsf{T} + \bar{\mathbf{Q}}_\mathbf{w}[k]$$

...repeat with $k = k + 1$...