TTK4115

# Lecture 11

The Kalman filter part 1

Morten. O. Alver (based on material by Morten D. Pedersen)

# This lecture

1. Kalman filtering in continuous time

2. Colored noise

3. Diagonalization of noise/disturbance terms

# Topic

1. Kalman filtering in continuous time

2. Colored noise

3. Diagonalization of noise/disturbance terms

## Physical model

Let a general plant model be given by a random process

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

---

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 $\mathcal{A}_{\mathbf{vw}}(t,\tau) = \mathrm{E}[\mathbb{v}(t)\mathbb{w}(\tau)^{\mathsf{T}}] \equiv \mathbf{0}$.

## Observer

We want to do state 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 $\mathbb{y}$.

## Dynamics of the estimation error

The random estimation error is defined by $\mathbb{e} = \mathbb{x} - \hat{\mathbb{x}}$. The rate of change of $\mathbb{e}$ is:

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

## 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 **P** quantifies the uncertainty in the estimate; low variances (found along the diagonal) imply good estimates!

## Reminder: covariance of model states

We earlier found that the covariance update equation for this model:

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

Is on this form:

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

## Covariance of estimation error

Now we want to find the covariance update equation for:

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

The rate of change of $\mathbf{P}(t) \triangleq \mathrm{E}[\mathbb{e}(t)\mathbb{e}(t)^{\mathsf{T}}]$ is:

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

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[1] 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

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

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

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

## Covariance dynamics

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

## Covariance dynamics

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

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

## Estimation performance

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 $J_{\mathrm{mse}}$ decreases at the fastest possible rate by optimizing with respect to the observer gain $\mathbf{L}(t)$. We need to calculate:

$$\frac{\partial \mathrm{tr}(\dot{\mathbf{P}})}{\partial \mathbf{L}}$$

which is *the derivative of a scalar (the trace) with respect to a matrix*. Then we need to find the value of $\mathbf{L}$ for which the derivative is 0.

# Kalman gain

## Matrix differentiation rules

These are the differentiation rules that we will use:

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{XA}) = \mathbf{A}^\mathsf{T}$$
$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{AX}^\mathsf{T}) = \mathbf{A}$$
$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{XBX}^\mathsf{T}) = \mathbf{XB}^\mathsf{T} + \mathbf{XB}$$

## Kalman Gain

Using the matrix differentiation rules, we get the result:

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

Then we set $\frac{\partial \mathrm{tr}(\dot{\mathbf{P}})}{\partial \mathbf{L}} = \mathbf{0}$ to find the value of $\mathbf{L}(t)$ that gives the optimum:

$$\mathbf{L}(t) = \mathbf{P}(t)\mathbf{C}^\mathsf{T}\mathbf{R_v}^{-1}$$

This value for the observer feeback gain is called the *Kalman Gain*.

## Kalman Gain

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

Differentiating again with respect to **L** shows that a minimum has indeed been found; the Hessian is positive definite.

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

## Optimal covariance dynamics

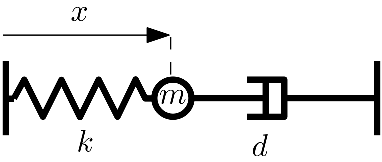
Upon selection of **L**, the covariance follows from

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

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

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

# Example: Mass spring damper




## State space equation

We now study a mass spring damper system with a disturbance and measurement noise:

$$\overbrace{\begin{bmatrix} \dot{\mathbb{x}}_1 \\ \dot{\mathbb{x}}_2 \end{bmatrix}}^{\dot{\mathbb{x}}} = \overbrace{\begin{bmatrix} 0 & 1 \\ -(k/m) & -(d/m) \end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix} \mathbb{x}_1 \\ \mathbb{x}_2 \end{bmatrix}}^{\mathbb{x}} + \overbrace{\begin{bmatrix} 0 \\ 1/m \end{bmatrix}}^{\mathbf{B}} u + \overbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}^{\mathbf{G}} \mathbb{w}$$

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

Noise $\mathbb{v}$ : $\mathcal{A}_{\mathbf{v}}(t,\tau) = \mathrm{E}[\mathbb{v}(t)\mathbb{v}(\tau)^{\mathsf{T}}] = \delta(t-\tau)\mathbf{R_v}$

Disturbance $\mathbb{w}$ : $\mathcal{A}_{\mathbf{w}}(t,\tau) = \mathrm{E}[\mathbb{w}(\tau)\mathbb{w}(t)^{\mathsf{T}}] = \delta(t-\tau)\mathbf{Q_w}$

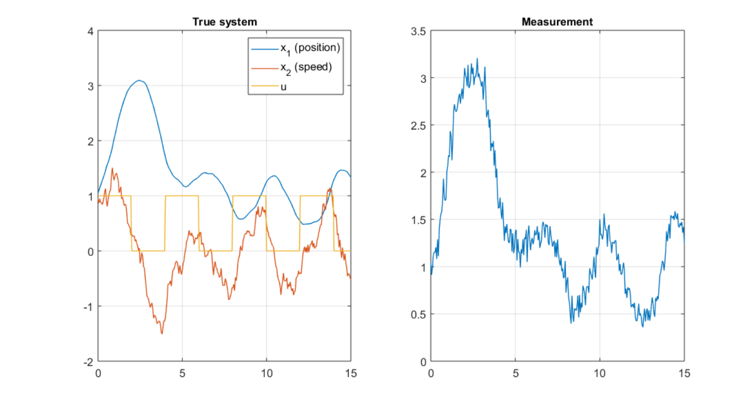
Parameter values : $m = 1$, $k = 0.5$, $d = 0.5$, $\mathbf{R_v} = 0.01$ and $\mathbf{Q_w} = 4$.

# Example: Mass spring damper

## Simulation of system with disturbance and measurement noise

We impose a regular input signal for 2 seconds at a time.

The figure shows one realization of the true system's state values and the noisy measurement:



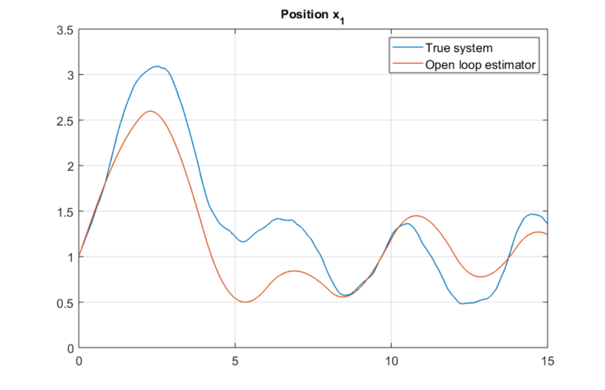
# Example: Mass spring damper

## Open loop estimator

An open loop estimator of the form

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

performs poorly due to the process disturbance:



# Example: Mass spring damper

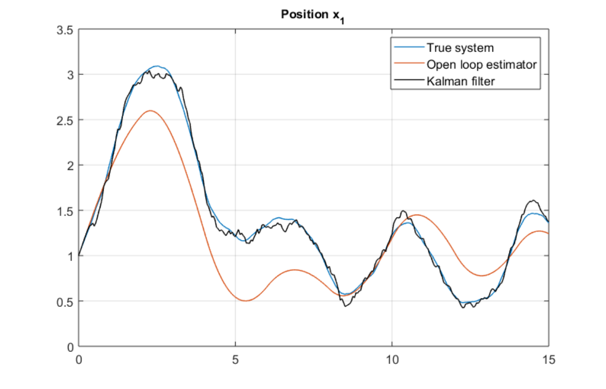
## Kalman filter

We then run an observer:

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

using the Kalman gain: $\mathbf{L}(t) = \mathbf{P}(t)\mathbf{C}^{\mathsf{T}}\mathbf{R}_{\mathbf{v}}^{-1}$, with $\mathbf{P}$ calculated according to:

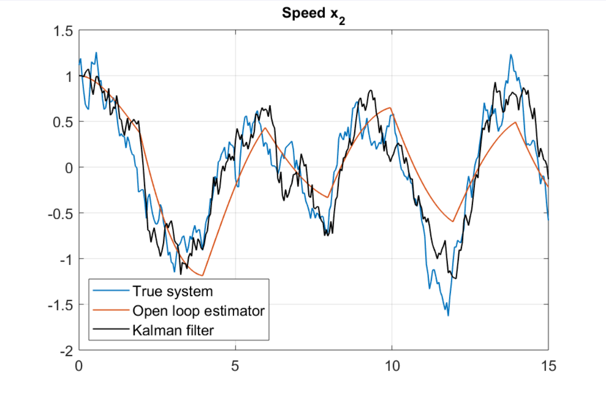
$$\dot{\mathbf{P}} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^{\mathsf{T}} + \mathbf{G}\mathbf{Q}_{\mathbf{w}}\mathbf{G}^{\mathsf{T}} - \mathbf{P}\mathbf{C}^{\mathsf{T}}\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{C}\mathbf{P}$$



# Example: Mass spring damper

## Kalman filter

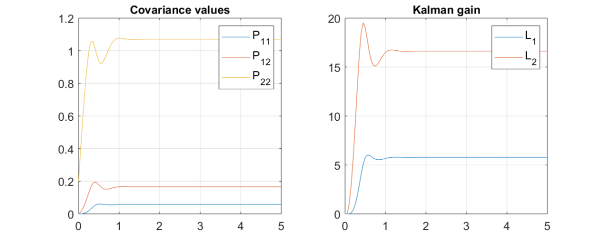
Plotting the true, open loop estimate and Kalman filter estimate of the speed (the unmeasured state), we see again that the observer tracks better:



# Example: Mass spring damper

## Kalman filter

The $\mathbf{P}(t)$ matrix and Kalman gain $\mathbf{L}(t)$ reach constant values after a short time:




The Kalman gain represents the optimal weighting between the uncertainty of the model and of the measurements:

- More noise (higher value of $\mathbf{R_v}$ ) gives lower Kalman gain (less emphasis on measurements).
- Stronger disturbance (higher value of $\mathbf{Q_w}$) gives higher Kalman gain (less emphasis on the model).

# LQR and Kalman filter duality

## Dual dynamics

Recall the *dual* system

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

## LQR

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

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

## Duality[2]

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[3]), 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 <u>observable</u> in order to permit computation of $\mathbf{L}$.

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

[3] 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 $V(\mathbf{x}) = \mathbf{x}^{\mathsf{T}}\mathbf{S}\mathbf{x} > 0$. Then we compute $\dot{V}(\mathbf{x})$:

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

The Lyapunov function is decreasing, indicating stability of the closed-loop plant (provided it is controllable).

## 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 $V(\mathbf{e}) = \mathbf{e}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{e} > 0$. Then we compute $\dot{V}(\mathbf{x})$:

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

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 if the noise affecting our system 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 $\mathbb{y}$.

## 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

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

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 $\mathbb{u}(t)$:

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

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 plant is shown to the right. If the plant is BIBO stable[4] and $f(t) = 0$ 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

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

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

[4] No poles in the RHP.

## The **Wiener–Khinchin–Einstein**[5] 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.

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

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

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

## Power spectral density

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

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

Therefore we have:

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

[5] Einstein was first in 1914!

## 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 gives 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 $m_u = 0$, autocorrelation $\mathcal{R}_u(\tau) = q\delta(\tau)$ and spectral density $\mathcal{S}_u(\omega)$. Let $\mathbb{u}(t)$ be filtered by $H$.
The output $\mathbb{y}(t)$ then has 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**.

## Filters & Colors

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

## 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

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

where $\mathbb{w}$ is white with zero mean and variance given by $\mathrm{E}[\mathbb{w}\mathbb{w}^\mathsf{T}] = \delta(0)\mathbf{Q}_\mathbf{w}$. 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[6], progress can be made.

[6] 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 $\Omega$ 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

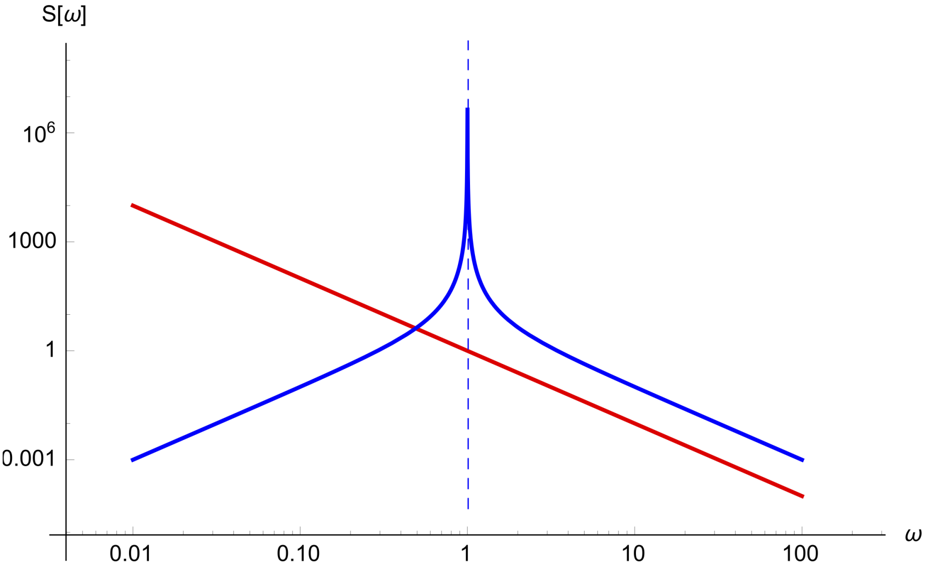
$$H(s) = \frac{k_{3\text{P}} s}{s^2 + \omega_{3\text{P}}^2}$$

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

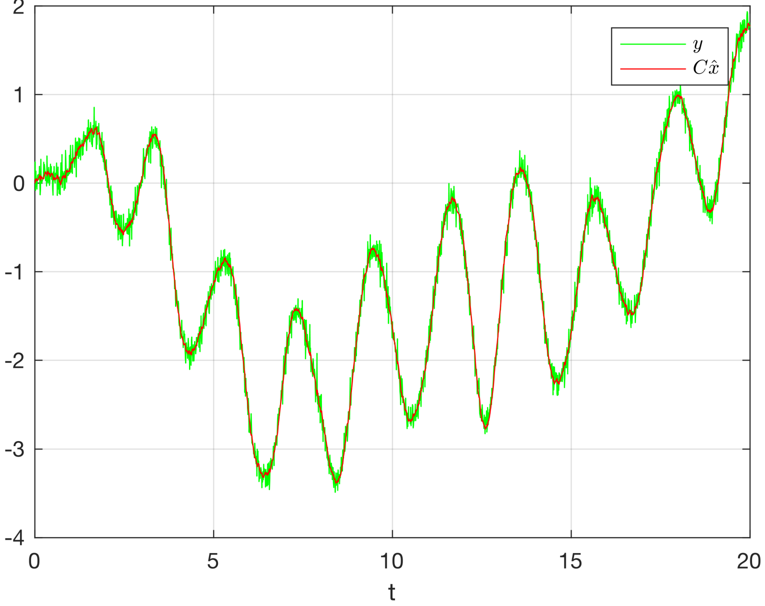
$$\hat{\mathbb{Q}}_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbb{x}}_1 \\ \hat{\mathbb{x}}_2 \\ \hat{\mathbb{x}}_3 \\ \hat{\mathbb{x}}_4 \end{bmatrix}$$

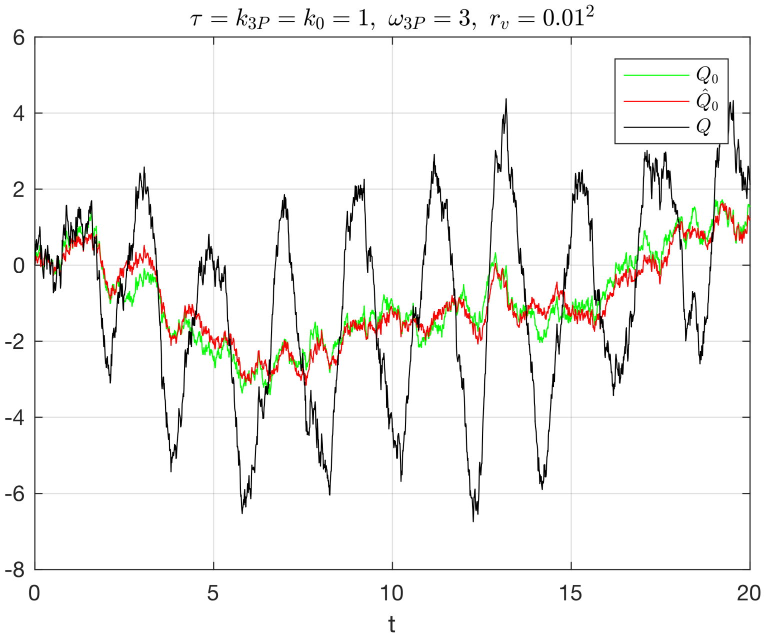
## Optimal estimation

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

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

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





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

# Topic

1. Kalman filtering in continuous time

2. Colored noise

3. Diagonalization of noise/disturbance terms

## Physical model

The plant model is described 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 $\mathcal{C}_{\mathbf{w}} = \delta(0)\mathbf{Q}_{\mathbf{w}}$ 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 <u>correlated</u>.

## 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 $\mathbf{\Lambda}_M$ denote a diagonal matrix of real[7] eigenvalues and let $\mathbf{E}_M$ describe the corresponding matrix of orthonormal[8] eigenvectors. Then, the matrix can be represented as

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

## Diagonalized representation

Let the covariance matrices be diagonalized

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

---

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 $\mathcal{C}_{\mathbf{w}'} = \delta(0)\mathbf{\Lambda}_\mathbf{w}$ and $\mathcal{C}_{\mathbf{v}'} = \delta(0)\mathbf{\Lambda}_\mathbf{v}$. 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.

---

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

[8] 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.