Systems and Control Theory

Lecture Notes

Dr. Claudia Totzeck

July 10, 2019

Contents

Introduction

These lecture notes are based on handwritten notes by Tobias Damm which rely on the books / notes:

- E. Zerz: Introduction to Systems and Control Theory,
- J.W. Polderman, J. Willems: Introduction to Mathematical Systems Theory,
- H.W. Knobloch, H. Kwakernaak: Lineare Kontrolltheorie,
- D. Hinrichsen, A.J. Pritchard: Mathematical Systems Theory I,
- E.D. Sontag: Mathematical Control Theory.

Contents

We discuss basic concepts and ideas of control theory and their applications. In particular we focus on:

- formulation of discrete and continuous time dependent problems of linear and nonlinear dynamical systems,
- stability of dynamical systems,
- reachability, controllability and observability,
- feedback control

Prerequisites

We assume that the contents of the lectures

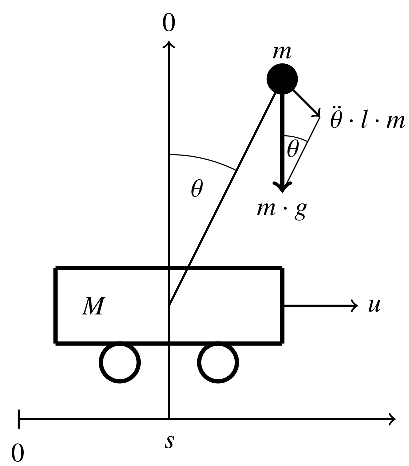
- Introduction to Numerics
- Introduction to differential equations

are well understood.

1 Motivating Examples

1.1 The inverted pendulum on a cart

We consider a pendulum which is assumed to consist of a point mass of mass m affixed to the end of a massless rigid rod of length ℓ that is held by a cart. Clearly, the pendulum is inherently unstable. We want to actively balance the the system to keep the system upright. In order to do so, we are allowed to move the cart to the left and the right. A sketch of the scenario can be found in Figure 1.



m	mass of pendulum
M	mass of cart
s	position of cart, i.e. its center of mass
ℓ	length of rod
θ	displacement angle
u	control (external driving force)
g	standard gravity

Figure 1: Sketch of a inverted pendululm on a cart.

The governing equations for (s, θ) can be derived using Newton's Laws. Indeed, they give the reaction forces at the joint between the pendulum and the cart which result in two equations for each body one in the x-direction and one in the y-direction

$$F - R_x = M\ddot{s}, \tag{1.1}$$
$$F_N - R_y - Mg = 0, \tag{1.2}$$

where R_x and R_y are reaction forces at the joint. F_N is the normal force applied to the cart. The position of the point mass m is given in inertial coordinates $(\hat{x}, \hat{y})$ as

$$r = (s - \ell \sin\theta)\hat{x} + \ell \cos\theta \hat{y}.$$

Taking two time derivatives results in the acceleration vector in the inertial reference frame

$$a = (\ddot{s} + \ell\dot{\theta}^2 \sin\theta - \ell\ddot{\theta}\cos\theta)\hat{x} + (-\ell\dot{\theta}^2\cos\theta - \ell\ddot{\theta}\sin\theta)\hat{y}.$$

Note that the reaction forces are positive as applied to the pendulum and negative when applied to the cart due to Newton's Third Law. Inserting into $R_x = ma_{\hat{x}}$ and $R_y = ma_{\hat{y}}$ yields

$$R_x = m(\ddot{s} + \ell\dot{\theta}^2 \sin\theta - \ell\ddot{\theta}\cos\theta),$$
$$R_y = m(-\ell\dot{\theta}^2\cos\theta - \ell\ddot{\theta}\sin\theta).$$

The unknown force F can be computed using (1.1), in fact

$$F = (M + m)\ddot{s} - m\ell\ddot{\theta}\cos\theta + m\ell\dot{\theta}^2\sin\theta.$$

To derive the equation for θ we employ the equation of motion for the pendulum $\sum F = ma_p$. We need to transform the coordinates to the body frame $\hat{x}_B = \cos\theta\hat{x} + \sin\theta\hat{y}$. We introduce the notation

$$R_p = \sqrt{R_x^2 + R_y^2}$$

and note that the assumption that the rod connecting the point mass and the cart is massless implies that load cannot be transferred perpendicular to the bar. Thus, the rod can only transfer loads along the axis of the rod itself, and hence the inertial frame components of the reaction forces can be simply written as $R_p\hat{y}_B$, where $\hat{y}_B$ is perpendicular to $\hat{x}_B$. Altogether, we get

$$\hat{x}_B \cdot \sum F = \hat{x}_B \cdot (R_x\hat{x} + R_y\hat{y} - mg\hat{y}) = \hat{x}_B \cdot (R_p\hat{y}_B - mg\hat{y}) = -mg\sin\theta.$$

For the left-hand side we obtain

$$m\hat{x}_B \cdot a_p = m\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \cdot \begin{bmatrix}\ddot{s} + \ell\dot{\theta}^2\sin\theta - \ell\ddot{\theta}\cos\theta\\ -\ell\dot{\theta}^2\cos\theta - \ell\ddot{\theta}\sin\theta\end{bmatrix} = m(\ddot{s}\cos\theta - \ell\ddot{\theta}).$$

Combining the two sides we arrive at

$$\ell\ddot{\theta} - g\sin\theta = \ddot{s}\cos\theta.$$

To summarize, the equations governing the motion of the inverted pendulum on a cart are given by the system

$$\begin{aligned} F &= (M+m)\ddot{s} - m\ell\ddot{\theta}\cos\theta + m\ell\dot{\theta}^2\sin\theta + \mu\dot{s},\\ \ell\ddot{\theta} - g\sin\theta &= \ddot{s}\cos\theta, \end{aligned}$$

where we added an friction term with parameter μ. We vectorize the system as

$$\begin{bmatrix}M+m & m\ell\cos\theta\\ \cos\theta & \ell\end{bmatrix}\begin{bmatrix}\ddot{s}\\ \ddot{\theta}\end{bmatrix} = \begin{bmatrix}m\ell\dot{\theta}^2\sin\theta - \mu\dot{s} + F\\ g\sin\theta\end{bmatrix}.$$

Note that this is a nonlinear system of ODEs. As we aim at stabilizing the system, we assume θ to be small and simplify the system. Indeed, a linearization around $[s, \theta, \dot{s}, \dot{\theta}] = [0, 0, 0, 0]$ with $\cos\theta \approx 1, \sin\theta \approx \theta, \dot{\theta}^2 \approx 0$ gives

$$\begin{bmatrix}M+m & m\ell\\ 1 & \ell\end{bmatrix}\begin{bmatrix}\ddot{s}\\ \ddot{\theta}\end{bmatrix} = \begin{bmatrix}-\mu\dot{s} + F\\ g\theta\end{bmatrix} \qquad \Leftrightarrow \qquad \begin{bmatrix}\ddot{s}\\ \ddot{\theta}\end{bmatrix} = \frac{1}{M\ell}\begin{bmatrix}\ell & -m\ell\\ -1 & M+m\end{bmatrix}\begin{bmatrix}-\mu\dot{s} + F\\ g\theta\end{bmatrix}.$$

In the following we assume that we are able to control the force F. We denote our control by $u := F$.
Rewriting as first order system with $x = [s, \theta, \dot{s}, \dot{\theta}]$ we obtain

$$\dot{x} = \begin{bmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & -\frac{mg}{M} & -\frac{\mu}{M} & 0\\ 0 & g\frac{M+m}{M\ell} & \frac{\mu}{M\ell} & 0\end{bmatrix}x + \begin{bmatrix}0\\ 0\\ \frac{1}{M}\\ \frac{1}{M\ell}\end{bmatrix}u =: Ax + Bu.$$

Let $s + \ell\sin\theta \approx s + \ell\theta = [1, L, 0, 0]x =: y$ be the measured output, then the linearized system has the form

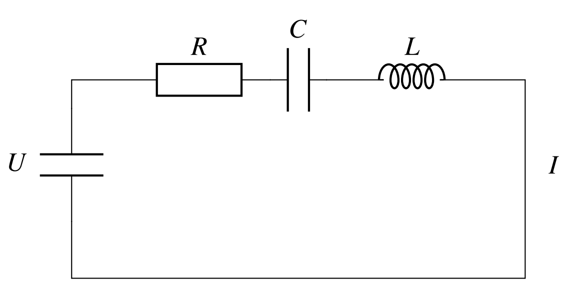
$$\begin{aligned} \dot{x} &= Ax + Bu,\\ y &= Cx. \end{aligned}$$

Questions

- Is there a control law u to stabilize the system in $x = 0$?
- Does a stabilizing u for the linearized system also stabilize the nonlinear system?
- What information does the controller need?

1.2 An electrical circuit

We consider an electrical circuit consisting of a resistor R, a capacity C and an inductor L. We assume the voltage U to be the input and the current I to be the measured output.



R	resistance
C	capacitor
L	inductor
U	voltage
I	current

To get a mathematical model we introduce the auxiliary variables $i_R, u_R, i_C, u_C, i_L, u_L$ and assume the Kirchhoff's Laws to hold:

$$\begin{aligned} I &= i_R = i_C = i_L,\\ U &= u_R + u_C + u_L. \end{aligned}$$

Furthermore, we define the relations

$$\begin{aligned} u_R &= R\, i_R = R\, I, \\ C\dot{u}_C &= i_C = I \qquad (CU = Q = \dot{I}), \\ L\dot{i}_L &= L\dot{I} = u_L \end{aligned}$$

In vector form we obtain

$$\begin{bmatrix} L & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{I} \\ \dot{U} \\ \dot{u}_R \\ \dot{u}_C \\ \dot{u}_L \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 & 0 \\ R & 0 & -1 & 0 & 0 \\ 0 & -1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} I \\ U \\ u_R \\ u_C \\ u_L \end{bmatrix} = 0 \qquad :\Leftrightarrow \quad R_1 \dot{w} + R_0 w = 0,$$

where $w = [I, U, u_R, u_C, u_L]$ is the vector of input, output und auxiliary variables.
Of course this can be generalized as

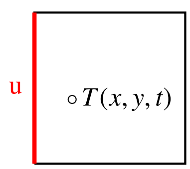
$$\sum_{j=0}^{k} R_j \frac{d^j}{dt^j} w = 0.$$

Questions

- Can we identify free variables?
- How can we eliminate auxiliary variables?
- Is there a description of the input / output behaviour?

1.3 Heat on a square plate

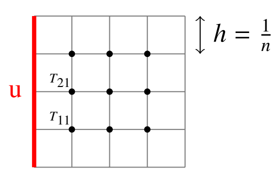
We consider a plate which allows to control the temperature at one side.



$T(x, y, t)$ is the temperature in (x, y) at time t

heat equation: $T_t = T_{xx} + T_{yy}$

We discretize the domain with n cells:



Approximations:

$$\begin{aligned} (T_{ij})_x &\approx \frac{T_{i+1,j} - T_{i,j}}{h}, \\ (T_{i,j})_{xx} &\approx \frac{(T_{i+1,j})_x - (T_{i,j})_x}{h} = \frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{h^2} \end{aligned}$$

Altogether, we get

$$(T_{i,j})_{xx} + (T_{i,j})_{yy} = -\frac{1}{h^2}(4T_{i,j} - T_{i+1,j} - T_{i-1,j} - T_{i,j+1} - T_{i,j-1}) \overset{!}{=} \frac{d}{dt} T_{i,j}$$

which can be written vectorized as

$$\frac{d}{dt}\begin{bmatrix} T_{11} \\ T_{21} \\ T_{31} \\ - \\ \vdots \end{bmatrix} = -\frac{1}{h^2} \left[\begin{array}{cccc|cccc} 4 & -1 & & & -1 & & & \\ -1 & \ddots & \ddots & & & \ddots & & \\ & \ddots & \ddots & -1 & & & \ddots & \\ & & -1 & 4 & & & & -1 \\ \hline -1 & & & & \ddots & & & \\ & \ddots & & & & \ddots & & \\ & & \ddots & & & & & \\ & & & -1 & & & & \end{array}\right] \begin{bmatrix} T_{11} \\ T_{21} \\ T_{31} \\ - \\ \vdots \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ - \\ \vdots \end{bmatrix} u$$

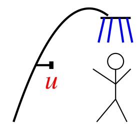
Introducing $x = [T_{11}, T_{21}, T_{31}, \dots]^T$ we obtain the form

$$\dot{x} = Ax + Bu, y = Cx$$

with for example $C = 1/n^2[1 \cdots 1]$. Note that this is the same form as in the first example.

1.4 Thermostatic Problem (Shower)

We consider a discrete time problem $k \in \mathbb{Z}$ and denote by $T(k)$ the temperature of water at time k. We assume that the



water temperature reacts with a delay h to the controller u, i.e. $T(k+h) = u(k)$. This leads to the following system

$$\begin{bmatrix} T(k+1) \\ \vdots \\ \vdots \\ T(k+h) \end{bmatrix} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & & 1 \\ & & & 0 \end{bmatrix} \begin{bmatrix} T(k) \\ \vdots \\ \vdots \\ T(k+h-1) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ u(k) \end{bmatrix}.$$

A typical control in this setting would be

$$u(k) = u(k-1) - \alpha(T(k) - T_{\text{opt}}) =: u(k-1) - \alpha\Delta(k),$$

which can be rewritten as

$$\begin{aligned} T(k+h) &= T(k+h-1) - \alpha(T(k) - T_{\text{opt}}) \\ &= T(k+h-1) - \alpha(T(k) - T_{\text{opt}}) - T_{\text{opt}} \\ &= T(k+h-1) - T_{\text{opt}} - \alpha(T(k) - T_{\text{opt}}) \\ &= \Delta(k+h-1) - \alpha\Delta(k) \end{aligned}$$

$$\Rightarrow \begin{bmatrix} \Delta(k+1) \\ \vdots \\ \vdots \\ \Delta(k+h) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ -\alpha & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta(k) \\ \vdots \\ \vdots \\ \Delta(k+h-1) \end{bmatrix}.$$

A common goal is to *stabilize* the problem, which basically means $\Delta(k) \to 0$ for $k \to \infty$.

2 Polynomial models

In this lecture we focus on systems that can be described by Polynomial models.

2.1 Abstract definition

A dynamical system is a pair $\Sigma = (\mathcal{A}, \mathcal{B})$ consisting of a signal set $\mathcal{A}$ and a behaviour $\mathcal{B}$.
Here $\mathcal{A} \subset \{f : T \to \mathbb{R}^n\}$, where T is the time domain, i.e. $T = \mathbb{R}, \mathbb{Z}, \mathbb{R}_+, \mathbb{N}, [a,b], \dots$. Typically, we consider

$$\mathcal{A} = C^k(\mathbb{R}, \mathbb{R}^n) \quad \text{or} \quad C^\infty(\mathbb{R}_+, \mathbb{R}^n).$$

The behaviour $\mathcal{B} \subset \mathcal{A}$ contains the admissible signals.

In *polynomial models*, the behaviour is defined by polynomial matrices $R \in \mathbb{R}[s]^{m\times n} \cong \mathbb{R}^{m\times n}[s]$, where

$$R(s) = \sum_{j=0}^{k} R_j s^j, \qquad R_0, \dots, R_k \in \mathbb{R}^{m\times n}, \qquad \deg R = k \quad \text{if} \quad k = \min\{\ell : R_p = 0 \text{ for all } p > \ell\}.$$

For $w \in C^k(\mathbb{R}, \mathbb{R}^n)$ we write

$$R\left(\frac{d}{dt}\right) w = \sum_{j=0}^{k} R_j \frac{d^j}{dt^j} w = \sum_{j=0}^{n} R_j w^{(j)}.$$

For sequences $w = (\dots, w(0), w(1), \dots)$ let

$$R(\sigma)w = \sum_{j=0}^{k} R_j \sigma^j w(t) = \sum_{j=0}^{k} R_j w(t+j),$$

where σ denotes the left-shift operator. In more detail, $\sigma(w_0, w_1, w_2, \dots) = (w_1, w_2, \dots)$.

Admissible signals are then defined by

$$R\left(\frac{d}{dt}\right)w = 0 \quad \text{(continuous time)} \qquad \text{or} \qquad R(\sigma)w = 0 \quad \text{(discrete time)}.$$

Examples

- $R_1\dot{w} + R_0 w = 0$ in the example of the electrical circuit in Section 1.2
- The increments in Section 1.4: $\Delta(k+h) = \Delta(k+h-1) - \alpha\Delta(k) \Leftrightarrow R(\sigma)\Delta = 0$ with $R(s) = s^h - s^{h-1} + \alpha$.

Polynomial models describe linear, time-invariant, differential / difference algebraic system, in short LTID.
Other classes of systems are:

-nonlinear	$\dot{x} = x^2 + 2xu$
-time-varying	$\dot{x} = tx$
-stochastic	$dX = A_0 X dt + A_1 X dB_t$ for a Brownian motion B_t
-delay	$\dot{x} = A_0 x(t) + A_1 x(t-T)$
-PDE	$u_t = u_{xx}$

Remark 2.1. Note that we require a high regularity for admissible solutions, $C^k(\mathbb{R}, \mathbb{R}^n)$ or $C^\infty(\mathbb{R}, \mathbb{R}^n)$. In electronics applications with switching behaviour, it is obvious that these requirements are not met by the solutions. Therefore, there exists a theory imposing less assumptions leading wo the notion of *weak solutions* in $\mathcal{L}_1^{\mathrm{loc}}(\mathbb{R}, \mathbb{R}^n)$. Briefly said, those allow for finitely many jumps on compact time intervals. Nevertheless, it can be shown that C^∞-solutions are dense in the set of weak solutions for LTID problems. This justifies that we restrict the considerations of this lecture to these regular solutions.

2.2 The scalar case

We consider the setting $m = 1 = n, R \in \mathbb{R}[s], \deg R = k$. Firstly, we recall some well-known facts:
The scalar linear homogeneous differential equation of k-th order has a k-dimensional solution space $W \subset \mathcal{A}$. All solutions of the inhomogeneous equation

$$R\left(\frac{d}{dt}\right)w = r$$

are given in the form $w_0 + w$, where w_0 is one particular solution and w solves the homogenous problem.

For difference equations essentially the same holds true as the following examples illustrate.

Examples

- $R(s) = s^4 - 1 = (s-1)(s+1)(s-i)(s+i)$, $R(\sigma)w = 0$ for

$$w(k) = \begin{cases} 1^k, & w = (1, 1, 1, \dots) \\ (-1)^k, & w = (-1, 1-, 1, \dots) \\ i^k, & w = (i, -1, -i, 1, i, \dots) \\ (-i)^k, & w = (-i, -1, i, \dots) \end{cases}$$

- $R(s) = (s-2)^2 = (s^2 - 4s + 4)$, $R(\sigma)w = 0$, $w = (2, 4, 8, 16, 32, \dots)$

$$w(k+2) - 4w(k+1) + 4w(k) = 0$$

  and $w = (1 \cdot 2, 2 \cdot 4, 3 \cdot 8, 4 \cdot 16, \dots)$ is solution as well.

If λ is an m-fold zero of $R(s)$, then the signals $w(k) = \lambda^k, w(k) = k\lambda^k, \dots, w(k) = k^{m-1}\lambda^k$ are linearly independent solutions of $R(\sigma)w = 0$ except for $\lambda = 0$. To avoid sophistry, we will assume that $R(0) \neq 0$, if we consider discrete time. Most of the problems we look at deal with the continuous case anyway.

2.3 Smith Normal Form

In this section we define the class of unimodular polynomial models which allow to reduce the considerations to the scalar case. The concept is similar to the singular value decomposition for ordinary matrices. We begin with the notion of equivalence.

Definition 2.2 (Equivalent differential equations)**.** Let $R_i(\xi) \in \mathbb{R}[\xi]^{g_i \times q}$, $i = 1, 2$. The differential equations

$$R_1\left(\frac{d}{dt}\right)w = 0 \quad \text{and} \quad R_2\left(\frac{d}{dt}\right)w = 0$$

are called *equivalent* if they define the same dynamical system. In other words, equivalence means that w is a solution of $R_1\left(\frac{d}{dt}\right)w = 0$ if and only if it is also a solution of $R_2\left(\frac{d}{dt}\right)w = 0$.

Example

Consider the systems given by

$$R_1(\xi) = \begin{bmatrix} \xi^2 + 1 & 0 \\ 0 & \xi^2 - 1 \end{bmatrix}, \qquad R_2(\xi) = \begin{bmatrix} \xi^2 + 1 & 0 \\ \xi^4 + \xi^2 & \xi^2 - 1 \end{bmatrix}.$$

We observe that the first equations is equal in both systems and implies that

$$\frac{d^2}{dt^2}w_1 + \frac{d^4}{dt^4}w_1 = 0.$$

Inserting this into the second equation of R_2 we find

$$\frac{d^2}{dt^2}w_1 + \frac{d^4}{dt^4}w_1 - w_2 + \frac{d^2}{dt^2}w_2 = -w_2 + \frac{d^2}{dt^2}w_2 = 0.$$

Hence, showing the equivalence.

Remark 2.3. Note that equivalent differential equations specify the same behaviour $\mathcal{B}$.

Definition 2.4. A matrix $P \in \mathbb{R}[s]^{n \times n}$ is called

- *nonsingular*, if $\det P(s) \neq 0$, i.e. $\exists s_0 \in \mathbb{C}\colon\ \det P(s_0) \neq 0$.
- *unimodular*, if $\det P(s) = const \neq 0$, i.e. $\forall s \in \mathbb{C}\colon\ \det P(s) \neq 0$.

Here the determinant is defined as the determinant of an ordinary matrix, so $\det P(s)$ is a scalar polynomial.

Remark 2.5. In the unimodular case, we have by Cramer's rule

$$P^{-1}(s) = \frac{1}{\det P(s)}\tilde{P}(s),$$

where $\tilde{P}(s)$ is the matrix of cofactors. Hence, it holds

$$P \in \mathbb{R}[s]^{n \times n} \text{ unimodular} \qquad \Leftrightarrow \qquad P(s)^{-1} \in \mathbb{R}[s]^{n \times n}.$$

The following Lemma shows that transformations by unimodular matrices lead to equivalent systems.

Lemma 2.6. *Let $U \in \mathbb{R}[s]^{m \times m}$, $V \in \mathbb{R}[s]^{n \times n}$ unimodular, $R \in \mathbb{R}[s]^{m \times n}$, $\mathcal{A} = C^\infty(\mathbb{R}, \mathbb{R}^n)$. Then*

$$w \text{ solves } R\left(\frac{d}{dt}\right)w = 0 \qquad \Leftrightarrow \qquad w = V\left(\frac{d}{dt}\right)v \text{ and } v \text{ solves } U\left(\frac{d}{dt}\right)R\left(\frac{d}{dt}\right)V\left(\frac{d}{dt}\right)v = 0$$

Proof. We prove by insertion. For all $w \in \mathcal{A}$ exists $v \in \mathcal{A}$ such that $w = V\left(\frac{d}{dt}\right)v$, indeed, due to the previous remark $v = V\left(\frac{d}{dt}\right)^{-1}w$. Now, let w solves $R\left(\frac{d}{dt}\right)w = 0$ hold, then

$$0 = R\left(\frac{d}{dt}\right)w = R\left(\frac{d}{dt}\right)V\left(\frac{d}{dt}\right)v$$

and hence

$$U\left(\frac{d}{dt}\right)R\left(\frac{d}{dt}\right)V\left(\frac{d}{dt}\right)v = U\left(\frac{d}{dt}\right)0 = 0.$$

On the other hand, let $w = V\left(\frac{d}{dt}\right)v$ and v solves $U\left(\frac{d}{dt}\right)R\left(\frac{d}{dt}\right)V\left(\frac{d}{dt}\right)v = 0$. Then

$$0 = U\left(\frac{d}{dt}\right)R\left(\frac{d}{dt}\right)V\left(\frac{d}{dt}\right)v = U\left(\frac{d}{dt}\right)R\left(\frac{d}{dt}\right)w$$

and since $U\left(\frac{d}{dt}\right)R\left(\frac{d}{dt}\right)$ has full rank due to the unimodularity, we obtain $R\left(\frac{d}{dt}\right)w = 0$. □

We emphasize that unimodularity is essential. Exemplarily, let us consider

$$R(s) = 1, U(s) = s\colon R\left(\frac{d}{dt}\right)w = 0 \quad \Leftrightarrow \quad w = 0,$$

BUT

$$(UR)\left(\frac{d}{dt}\right)w = \frac{d}{dt}v = 0 \Leftrightarrow w = \text{const.}$$

In the following we introduce the Smith Normal Form, its concept is analogous to the singular value decompostion for ordinary matrices.

Theorem 2.7 (Smith Normal Form)**.** *For $R \in \mathbb{R}[s]^{m\times n}$ there exist $U \in \mathbb{R}[s]^{m\times m}$ and $V \in \mathbb{R}[s]^{n\times n}$ unimodular, such that*

$$URV = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \qquad D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_r \end{bmatrix} \in \mathbb{R}[s]^{r\times r} \textit{ nonsingular}$$

with monic scalar polynomials $d_1 | d_2 | \dots | d_r \neq 0$.

Remark 2.8. For the sake of completeness, we recall that a monic polynomial is a single-variable polynomial which leading coefficient is 1.

Proof. We make use of elementary unimodular transformation which include:

- scaling of rows and columns
- row /column permutation
- adding a polynomial multiple of row / columns to another row / column, this can be represented by the matrix

$$\begin{bmatrix} 1 & & & \\ & \ddots & & \\ & p(s) & \ddots & \\ & & & 1 \end{bmatrix}$$

We shall show that using this strategy, R can be transformed to

$$\tilde{R} = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \mathbf{Q} & \\ 0 & & & \end{bmatrix},$$

where d_1 is a monic and divides all entries of Q. Then we obtain the desired result with the help of an induction. We split the proof into different cases.

<u>Case 1:</u> There exists R_{ij} which divides all other entries of R. Without loss of generality $R_{ij} = R_{11}$. We can eliminate $R_{21}, \dots, R_{m_1}, R_{12}, \dots, R_{1n}$.

Suppose the assumption of Case 1 does not hold. Let $\delta(R) = \min\{\deg(R_{ij}) \mid R_{ij} \neq 0\}$, without loss of generality, $\deg R_{11} = \delta(R)$. The idea of the following steps is to transform R to R' with $\delta(R') < \delta(R)$. Then either Case 1 applies or this step can be repeated. Note that this procedure terminates after at most $\delta(R)$ steps.

Case 2a: R_{11} does not divide all R_{1j}, R_{i1}. For example

$$R_{11} \nmid R_{1k} = qR_{11} + p$$

with $\deg p < \deg R_{11} = \delta(R)$. Then subtract q times the 1st column from the k-th column to get

$$\delta(R') \leq \deg p < \delta(R).$$

Case 2b: Suppose the assumption of Case 2a does not hold. We eliminate $R_{1j}, R_{i1}, \dots$ and obtain

$$R' = \begin{bmatrix} d & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \mathbf{Q} & \\ 0 & & & \end{bmatrix}.$$

If $d \nmid Q_{ij}$ then add i-th row to 1st row and go to Case 2a.

□

Consequences of the Smith-form

We note the following implication

$$URV = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},\ D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_r \end{bmatrix} \quad \Rightarrow \quad r = \operatorname{rank} R \ = \#\text{ linear independent columns (rows) of } R \text{ in the sense of } R[s].$$

Furthermore, we note that for

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}, V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}, U_1RV_1 = D, UR = \begin{bmatrix} * \\ 0 \end{bmatrix}, RV = \begin{bmatrix} * & 0 \end{bmatrix}$$

the following holds true:
The rows of U_2 span the left-kernel of R and the columns of V_2 span the right-kernel of R.

2.4 Minimal Representation

We consider $R \in \mathbb{R}[s]^{p \times q}$ such that we have p equations and q variables.

How many equations are needed?

We recall that for U unimodular it holds

$$R\left(\frac{d}{dt}\right)w = 0 \quad \Leftrightarrow \quad UR\left(\frac{d}{dt}\right)w = 0 \quad \Leftrightarrow \quad U_1R\left(\frac{d}{dt}\right)w = 0. \tag{2.9}$$

In particular, we have $\ker R = \ker UR = \ker U_1R$ with $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$.

Definition 2.10. We call $R \in \mathbb{R}[s]^{p \times q}$ *minimal*, if $R \sim \hat{R}$ for $\hat{R} \in \mathbb{R}[s]^{\hat{p} \times q}$ implies $p \leq \hat{p}$.

Corollary 2.11.

$$R \in \mathbb{R}[s]^{p \times q} \text{ is minimal} \qquad \Leftrightarrow \qquad p = \operatorname{rank} R.$$

The minimal representation is directly linked to the 'fundamental principle' of *solvability*.

Corollary 2.12. *Let $R \in \mathbb{R}[s]^{p \times q}$. Then the following statements are equivalent:*

(a) $\operatorname{rank} R = p$

(b) for all $v \in C^\infty(\mathbb{R}, \mathbb{R}^q)$ exists $w \in C^\infty(\mathbb{R}, \mathbb{R}^p)$ such that $R\left(\frac{d}{dt}\right)w = v$.

Proof. Without loss of generality we assume $R = [D, 0]$, where D has p columns and 0 has $q - p$ columns, note that $d_p = 0$ is possible ('Smith'). Then

$$Rw = v \qquad \Leftrightarrow \qquad \begin{cases} d_1\left(\frac{d}{dt}\right)w_1 = v_1 \\ \vdots \\ d_p\left(\frac{d}{dt}\right)w_p = v_p \end{cases}$$

is solvable for all v, which is equivalent to

$$d_p \neq 0 \qquad \Leftrightarrow \qquad \operatorname{rank} R = p.$$

□

2.5 Inputs and outputs

Let $R(s) \in R[s]^{p\times q}$, $q > p, q - p = m$ and consider $R\left(\frac{d}{dt}\right) w = 0$, where $w = \begin{bmatrix} u \\ y \end{bmatrix}$, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^{q-m}$.

Definition 2.13. The variables $u_1, \dots, u_m$ are called *free* or *inputs*, if for all $u \in C^\infty(\mathbb{R}, \mathbb{R}^n)$ exists $y \in C^\infty(\mathbb{R}, \mathbb{R}^{q-m})$ such that

$$R\left(\frac{d}{dt}\right) w = 0.$$

The vector y is called *output*.

Corollary 2.14. *(a) If $R = [-Q, P]$ with $Q \in \mathbb{R}[s]^{m\times p}$, $m = q - p$ and $P \in \mathbb{R}[s]^{(q-m)\times(q-m)} = \mathbb{R}[s]^{p\times p}$ nonsingular and $w = \begin{bmatrix} u \\ y \end{bmatrix}$, then u is free.*

(b) If rank $R = r$*, then $q - r$ variables are free.*

Proof. (a)

$$R\left(\frac{d}{dt}\right) w = 0 \quad \Leftrightarrow \quad Q\left(\frac{d}{dt}\right) u = P\left(\frac{d}{dt}\right) y.$$

Using Corollary 2.11 and the fact that P is nonsingular, we get for all $u \in C^\infty$ the existence of a solution y.

(b) If rank $R = r$ then without loss of generality $R \in \mathbb{R}[s]^{r\times q}$ and without loss of generality $R = [-Q, P]$, $P \in \mathbb{R}[s]^{r\times r}$ nonsingular. Then the result follows by (a).

□

2.5.1 Latent variables and their elimination

Inputs u and outputs y constitute the *manifold variables*, i.e., those we are interested in. Further variables, which are introduced in the course of modelling or perturbations are called *latent* variables ℓ.

$$\xrightarrow{u} \boxed{\Sigma(u, y, \ell)} \xrightarrow{y}$$

Formalization:

We assume $R\left(\frac{d}{dt}\right) w = 0$, $w = \begin{bmatrix} \tilde{w} \\ \ell \end{bmatrix}$ with manifold variables $\tilde{w} \in \mathbb{R}^{\bar{q}}$ and latent variables $\ell \in \mathbb{R}^{q-\bar{q}}$ and $R = [-\tilde{R}, L]$ to be given. Then by definition:

$$\tilde{w} \in C^\infty(\mathbb{R}, \mathbb{R}^{\bar{q}}) \text{ is admissible (as first part of } w) \;\Leftrightarrow\; \text{there exists } \ell \in C^\infty(\mathbb{R}, \mathbb{R}^{q-\bar{q}}) : \; \tilde{R}\left(\frac{d}{dt}\right)\tilde{w} = L\left(\frac{d}{dt}\right)\ell.$$

Sometimes we want to get rid of the latent variables. This can be realized with the help of the Smith form as follows. Assume there exists $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$ unimodular with $UL = \begin{bmatrix} L_1 \\ 0 \end{bmatrix}$ such that

- the rows of U_2 are a basis (over $K[s]$) of the left kernel of L,
- L_1 has full row rank.

If now $U\tilde{R} = \begin{bmatrix} \tilde{R}_1 \\ \tilde{R}_2 \end{bmatrix}$ with $\tilde{R}_2 = U_2\tilde{R}$, then

$$\exists \ell \colon \tilde{R}\left(\frac{d}{dt}\right)\tilde{w} = L\left(\frac{d}{dt}\right)\ell \quad \Leftrightarrow \quad \exists \ell \colon \begin{cases} \tilde{R}_1\left(\frac{d}{dt}\right)\tilde{w} = L_1\left(\frac{d}{dt}\right)\ell \\ \tilde{R}_2\left(\frac{d}{dt}\right)\tilde{w} = 0 \end{cases} \quad \Leftrightarrow \quad \tilde{R}_2\left(\frac{d}{dt}\right)\tilde{w} = 0.$$

Corollary 2.15. *Let $\tilde{R} \in \mathbb{R}[s]^{p\times\bar{q}}$, $L \in \mathbb{R}[s]^{p\times(q-\bar{q})}$ and assume that the rows of $X \in \mathbb{R}[s]^{d\times q}$ form a basis of the left kernel of L. Then there exists ℓ such that it holds:*

$$\tilde{R}\left(\frac{d}{dt}\right)\tilde{w} = L\left(\frac{d}{dt}\right)\ell \quad \Leftrightarrow \quad \left(X\tilde{R}\right)\left(\frac{d}{dt}\right)\tilde{w} = 0.$$

2.5.2 State space representation (introduction of latent variables)

The system $R\left(\frac{d}{dt}\right)w = 0$ with $R(s) = \sum_{j=0}^{k} R_j s^j$ is equivalent to

$$\dot{x} = \begin{bmatrix} 0 & & & \\ I & \ddots & & \\ & & \ddots & \\ & & I & 0 \end{bmatrix} x + \begin{bmatrix} R_0 \\ \vdots \\ R_{k-1} \end{bmatrix} w = Kx + Lw,$$

$$0 = [0, \dots, 0, I]x + R_k w = Mx + Nw$$

See Exercise 3.1 for an explanation. In particular this means

$$w \text{ is admissible} \qquad \Leftrightarrow \qquad \exists x \colon \begin{bmatrix} \frac{d}{dt}I - K \\ -M \end{bmatrix} x = \begin{bmatrix} L \\ N \end{bmatrix} w.$$

If moreover $x = \begin{bmatrix} u \\ y \end{bmatrix} = w$, $R_k = N = [N_1, N_2]$ with N_2 being invertible and $L = [L_1, L_2]$ then

$$y = -N_2^{-1}(Mx + N_1 u) \qquad \Rightarrow \qquad \begin{cases} \dot{x} = Ax + Bu \\ y = Cx + Du \end{cases}$$

with

$$A = (K - L_2 N_2^{-1} M), \quad B = L_1 - L_2 N_2^{-1} N_1, \quad C = -N_2^{-1} M \quad \text{and} \quad D = -N_2^{-1} N_1.$$

This is called a *state space representation* with *state vector* $x(t)$ containing all information we are interested in (i.e. inputs and outputs).

Remark 2.16. If we want to go back to the polynomial model we obtain

$$\begin{bmatrix} sI - A \\ C \end{bmatrix}_{s=\frac{d}{dt}} x = \begin{bmatrix} B & 0 \\ -D & I \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix}$$

which yields

$$\begin{aligned} (sI - A)x &= Bu \qquad \Rightarrow \qquad x = (sI - A)^{-1}Bu, \\ Cx &= -Du + y \end{aligned}$$

Inserting the first into the second equation we get

$$(C(sI - A)^{-1}B + D)u = y$$

We cannot be sure that $(sI - A)^{-1}$ is a polynomial matrix, thus we multiply by the smallest common divisor of the denominators collected in the matrix $\Delta(s)$ to obtain

$$\Delta(s)\left(C(sI - A)^{-1}B + D\right)u = \Delta(s)y$$

where the first matrix on the left-hand side has left kernel $\Delta(s)[C(sI - A)^{-1}, -I]$.

To summarize, depending on the application it is useful to either introduce latent variables or to eliminate them from the system.

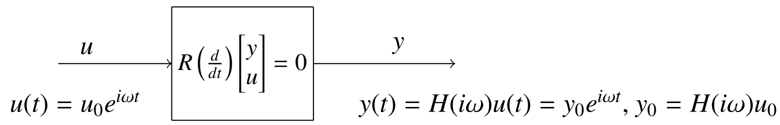
2.6 Transfer function

In this section we are interested in the transformation to the frequency domain.

Definition 2.17. Let $Q \in \mathbb{R}[s]^{m \times p}$, $P \in \mathbb{R}[s]^{m \times m}$ $R = [-Q, P] \in \mathbb{R}[s]^{m \times (m+p)}$ be in input/output form:

$$Q\left(\frac{d}{dt}\right)u(t) = P\left(\frac{d}{dt}\right)y(t), \qquad P \ \text{ nonsingular.}$$

Then the rational function $H(s) = \left(P^{-1}Q\right)(s)$ is called *transfer function*.



As we have seen in the previous remark: for the system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

it takes the form $H(s) = C(s\,\mathrm{I} - A)^{-1}B + D$. Note that we do not need $\Delta(s)$ here as we allow for rational functions!
The transfer function describes the system in the frequency domain. We emphasize this by using the variable s instead of t for time. The value $|H(i\omega)|$ is called *gain* (dt. Verstärkung) and $\arg(H(i\omega))$ is the *phase shift*.

A link between the time domain and the frequency domain is given by the *Laplace transformation* $\mathcal{L}$:
Let $f\colon \mathbb{R}_+ \to \mathbb{R}^n$ with $\|f(t)\| \leq Me^{\alpha t}$ for all t. Then $\hat{f} = F = \mathcal{L}f$ is defined by

$$F(s) = \int_0^\infty e^{-st} f(t)dt \qquad \text{for all} \qquad s \in \mathbb{C},\ \mathrm{Re}s > \alpha.$$

The Laplace transform has the following properties:

- $\mathcal{L}$ is linear
- $(\mathcal{L}\dot{x})(s) = \int_0^\infty e^{-st}\dot{x}(t)dt = e^{-st}x(t)|_{t=0}^{t=\infty} + s\int_0^\infty e^{-st}x(t)dt = s(\mathcal{L}x)(s) - x(0)$
- For the convolution $(f * g)(t) := \int_0^t f(t-\tau)g(\tau)d\tau$, it holds $\mathcal{L}(f * g) = \hat{f}(s) \cdot \hat{g}(s)$.
- Let $f, g\colon \mathbb{R}_+ \to \mathbb{R}$ then
$$(f * g)(t) = \int_0^t f(t-\tau)g(\tau)d\tau = \int_{-\infty}^\infty \tilde{f}(t-\tau)\tilde{g}(\tau)d\tau$$
with
$$\tilde{f} = \begin{cases} f, & t \geq 0, \\ 0, & t < 0 \end{cases} \qquad \tilde{g} = \begin{cases} g, & t \geq 0, \\ 0, & t < 0 \end{cases}.$$

Remark 2.18. In the engineering community the representation with the help of the transfer function is much more common than the dynamical system representation.

Examples

(1) In the case $x(0) = 0$ we have for

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

the time domain the following corresponding system in the frequency domain

$$\begin{aligned} (s\,\mathrm{I} - A)\hat{x}(s) &= \hat{B}u(s), \\ \hat{y}(s) &= C\hat{x}(s) + D\hat{u}(s) \end{aligned}$$

which leads to

$$\hat{y}(s) = C(s\,\mathrm{I} - A)^{-1}B\hat{u}(s) + D\hat{u}(s).$$

On the other hand

$$x(t) = \int_0^t \underbrace{e^{A(t-\tau)}B}_{f(t-\tau)}\ \underbrace{u(\tau)}_{g(\tau)}\ d\tau.$$

Then

$$\hat{f}(s) = (s\,\mathrm{I} - A)^{-1}B, \qquad \hat{x}(s) = (s\,\mathrm{I} - A)^{-1}B\hat{u}(s)$$

leading to

$$\hat{y}(s) = H(s)\hat{u}(s) = (C(s\,\mathrm{I} - A)^{-1}B + D)\hat{u}(s).$$

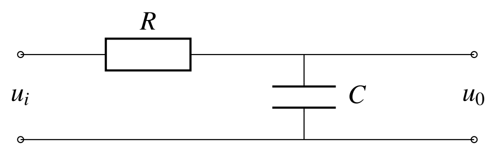


Figure 2: Low pass filter

(2) We consider a low-pass filter
with governing equations

$$\begin{aligned} u_i - u_0 &= u_R, \\ 0 &= -u_R + RI, \\ \dot{u}_0 &= \frac{1}{C} I. \end{aligned}$$

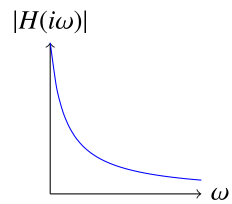
$$\begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & s \end{bmatrix} \begin{bmatrix} u_i \\ u_0 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ -1 & R \\ 0 & \frac{1}{C} \end{bmatrix}}_{[-1,-1,RC] \text{ is the corresponding left kernel}} \begin{bmatrix} u_R \\ I \end{bmatrix}$$

Multiplication with the left kernel, yields

$$[-1, 1 + RCs] \begin{bmatrix} u_i \\ u_0 \end{bmatrix} = 0.$$

Hence, the corresponding transfer function is given by

$$H(s) = \frac{1}{1 + RCs}$$



3 Stability

In this section we will define the different notions of stability and study conditions that tell us the stability properties of a system.

Definition 3.1. Let $P \in \mathbb{R}[s]^{p \times p}$ be nonsingular. The system $P\left(\frac{d}{dt}\right) y = 0$ is called *stable,* if every admissible signal is bounded. It is called *asymptotically stable*, if $y(t) \to 0$ as $t \to \infty$ for every admissible y. For discrete time systems the definition is analogous.

Based on this definition, we have the following observation:
Let u be a given input. If P is (asymptotically) stable and

$$P\left(\frac{d}{dt}\right) y = Q\left(\frac{d}{dt}\right) u,$$

then $\|y_1(t) - y_2(t)\|$ is bounded ($\to 0$) for all possible outputs y_1 and y_2.

3.1 Scalar Case

We consider $P \in \mathbb{R}[s]$. For continuous time we have

$$P\left(\frac{d}{dt}\right) y = 0 \qquad \Rightarrow \qquad y(t) = \sum_{\lambda, p(\lambda)=0} a_\lambda(t) e^{\lambda t}, \qquad a_\lambda \in \mathbb{R}[t]$$

and for discrete time

$$P(\sigma) y = 0 \qquad \Rightarrow \qquad y(k) = \sum_{\lambda, p(\lambda)=0} a_\lambda(k) \lambda^k.$$

Here, it holds $\deg a_\lambda = \mathrm{alg}_p(\lambda) - 1$, where alg_p is the algebraic multiplicity of λ, i.e.

$$p(s) = \prod (s - \lambda_j)^{\theta_j} \quad \Rightarrow \quad \mathrm{alg}_p(\lambda_j) = \theta_j.$$

Lemma 3.2. *Let $P \in \mathbb{R}[s]$.*

- $P\left(\frac{d}{dt}\right)y = 0$ *is stable* $\Leftrightarrow |a_\lambda(t)e^{\lambda t}| \nrightarrow \infty$ *as* $t \to \infty$ *for all* λ, $p(\lambda) = 0 \Leftrightarrow Re\lambda \leq 0$ *and* $\mathrm{alg}_p(\lambda) = 1$ *if* $Re\lambda = 0$.
- $P\left(\frac{d}{dt}\right)y = 0$ *is asymptotically stable* $\Leftrightarrow Re\lambda < 0$ *for all* λ, $p(\lambda) = 0$.

Proof. easy! □

3.2 General Case

Definition 3.3. Let $P \in \mathbb{R}[s]^{p\times p}$ and $\chi_p(s) = \det P(s)$ be the characteristic polynomial. If $\chi_p(\lambda) = 0$, then we call λ an eigenvalue of P, $\mathrm{geom}_p(\lambda) = \dim\ker P(\lambda)$ its geometrical multiplicity. λ is *semisimple*, if $\mathrm{geom}_p(\lambda) = \mathrm{alg}_{\chi_p}(\lambda)$.

The following results provides a link between the eigenvalues and stability.

Theorem 3.4. *Let $P \in \mathbb{R}[s]^{p\times p}$ nonsingular, $\Lambda = \{\lambda \in C\colon\ \det P(\lambda) = 0\}$.*

(i) $P\left(\frac{d}{dt}\right)y = 0$ *stable* $\Leftrightarrow \Lambda \subset \overline{\mathbb{C}}_- = \{s \in C\colon Res \leq 0\}$, *and all* $\lambda \in i\mathbb{R} \cap \Lambda$ *are semisimple.*
$P\left(\frac{d}{dt}\right)y = 0$ *asymptotically stable* $\Leftrightarrow \Lambda \subset \mathbb{C}_- = \{s \in \mathbb{C}\colon Res < 0\}$

(ii) $P(\sigma)\,y = 0$ *stable* $\Leftrightarrow \Lambda \subset \overline{\mathbb{D}} = \{s \in \mathbb{C}\colon |s| \leq 1\}$ *and* $|\lambda| = 1$ *semisimple.*
$P(\sigma)\,y = 0$ *asymptotically stable* $\Leftrightarrow \Lambda \subset \mathbb{D} = \{s \in \mathbb{C}\colon\ s < 1\}$

Proof. Without loss of generality, we assume $P = D = \mathrm{diag}(d_1, \ldots, d_p)$, $d_1|\ldots|d_p$. Then

$$P \text{ (asymptotically) stable } \Leftrightarrow d_j \text{ (asymptotically) stable for all } j.$$

Due to this observation we are left to show the following

$$\lambda \in \Lambda \text{ is semisimple} \Leftrightarrow \left(d_j(\lambda) = 0 \Rightarrow \mathrm{alg}_{d_j}(\lambda) = 1\right).$$

Let $D(\lambda) = \mathrm{diag}(d_1(\lambda), \ldots, d_k(\lambda), 0, \ldots, 0)$ with $d_j(\lambda) \neq 0$ for all $j = 1, \ldots, k$. Then we have

$$\mathrm{alg}_{\chi_p}(\lambda) = \sum_{j=k+1}^{p} \mathrm{alg}_{d_j}(\lambda) \geq p - k, \qquad \mathrm{geom}_p(\lambda) = \mathrm{geom}_D(\lambda) = p - k.$$

Hence, it holds equality if and only if $\mathrm{alg}_{d_j}(\lambda) = 1$ for all $j > k$. □

Remark 3.5. Note that considering only diagonal matrices in the proof above is justified by the transformation using unimodular matrices in order to obtain the Smith Normal Form. Indeed, multiplication with unimodular matrices leads to derivatives in the system. Hence, one obtains a transformed system which consists of linear combinations of solutions and derivatives of solutions of the untransformed system. On the other hand, we can obtain the untransformed system by another transformation with the inverse of the unimodular matrix, which is again polynomial and thus, we obtain the original system again by taking linear combinations of the solutions and the derivatives of solutions of the transformed systen. This basically means that by tranforming the system using a unimodular matrix we only change the constants of the solutions. This will be even more clear, when we look at parametrized solution of an (asymptotically) stable system.

In the case of asympotically stable solutions, we have

$$y(t) = \sum_{\lambda,\xi_p(\lambda)=0} a_\lambda(t)e^{\lambda t}, \mathrm{Re}\lambda < 0.$$

Taking derivatives w.r.t. t we find by product rule, that the degree of a_λ of the transformed system is preserved during the transformation, only constants are changed.

In the case of stable solutions, we have

$$y(t) = \sum_{\lambda,\xi_p(\lambda)=0} c_{1,\lambda}\cos(\lambda t) + ic_{2,\lambda}\sin(\lambda t).$$

Now, we find by chain rule that again only constants are changing when taking derivatives in the transformation process.

Example

We consider a system with discrete time with the matrices

$$P(s) = \begin{bmatrix} 1-s & \\ & 1-s \end{bmatrix}, \qquad \bar{P}(s) = \begin{bmatrix} 1-s & s \\ & 1-s \end{bmatrix}.$$

We have the equivalence

$$\begin{aligned} P(\sigma)\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \quad &\Leftrightarrow \quad \begin{cases} y_1(k) - y_1(k+h) = 0 \\ y_2(k) - y_2(k+h) = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} y_1 = \alpha(1,1,\dots) \\ y_2 = \beta(1,1,\dots) \end{cases}, \\ \bar{P}(\sigma)\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \quad &\Leftrightarrow \quad \begin{cases} y_1(k) - y_1(k+h) + y_2(k+h) = 0 \\ y_2(k) - y_2(k+h) = 0 \end{cases} \end{aligned}$$

Then we obtain, that for example $y_2 = (1,1,\dots), y_1 = (1,2,3,\dots)$ are unstable.

Special Case

We take a closer look now at systems in state space form.

Definition 3.6. The state space system

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx + Du \end{aligned}$$

is called *(asymptotically) stable*, if $\dot{x} = Ax$ is (asymptotically) stable.

Theorem 3.7 (Lyapunov)**.** *Let $A \in \mathbb{R}^{n\times n}$. Then the following statements are equivalent:*

(i) $\sigma(A) \subset \mathbb{C}_-$,

(ii) $\forall Y > 0 \exists X > 0\colon A^T X + XA = -Y$,

(iii) $\exists Y > 0 \exists X > 0\colon A^T X + XA = -Y$.

where $X > 0$ is a short notation for X is symmetric, positive definite.

Proof. We begin with (i)⇒ (ii). Let $\sigma(A) \subset \mathbb{C}_-$ hold. Then there exists $M \geq 1, \beta \geq 0$ such that $\|e^{At}\| \leq Me^{-\beta t}$. Hence it exists

$$X := \int_0^\infty e^{A^T t} Y e^{At} dt > 0, \qquad \text{if } Y > 0.$$

Moreover,

$$-Y = \int_0^\infty \frac{d}{dt}\left(e^{A^T t} Y e^{At}\right) dt = A^T \int_0^\infty e^{A^T t} Y e^{At} dt + \int_0^\infty e^{A^T t} Y e^{At} dt A = A^T X + XA.$$

The implication (ii) ⇒ (iii) is trivial. We consider (iii)⇒(i). Let $\lambda \in \sigma(A), Av = \lambda v, v \neq 0$ and assume it holds

$$A^T X + XA < 0 \quad \text{ with } X > 0.$$

Then it follows $0 > v^*(A^T X + XA)v = (\bar{\lambda} + \lambda)v^* X v$. Due to $v^* X v > 0$ we obtain $\mathrm{Re}\lambda = \frac{1}{2}(\bar{\lambda} + \lambda) < 0$ □

Remark 3.8. a) $X > 0$ defines a new norm $\|v\|_X^2 = v^T X v$. If $\dot{x}(t) = Ax(t)$ holds, then

$$\frac{d}{dt}\|x(t)\|_X^2 = \dot{x}^T X x + x^T X \dot{x} = x^T(A^T X + XA)x < 0,$$

which means strict descent.
b) If $\dot{x} = Ax + r(x)$ with $r(x) = o(\|x\|)$, then the nonlinear system is exponentially stable iff $\sigma(A) \subset \mathbb{C}_-$.

The formal definition of exponential stable is the following:

Definition 3.9. We consider $\dot{x} = f(x)$ with $f(0) = 0$. The system (or the equilibrium $x = 0$) is called *exponentially stable*, if there exist $\delta, M, \beta > 0$ such that

$$\|x(0)\| < \delta \Rightarrow \forall t \geq 0\colon \|x(t)\| \leq Me^{-\beta t}\|x(0)\|.$$

Theorem 3.10. *The system $\dot{x} = Ax + r(x), r(x) \in o(\|x\|)$ is exponentially stable iff $\sigma(A) \subset \mathbb{C}_-$.*

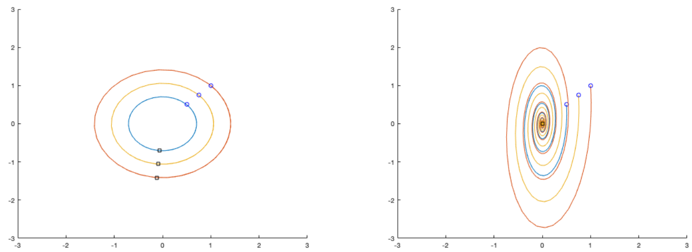


Figure 3: Sketch of stable (left) and asympotitically stable system (right).

Proof. $\Rightarrow$: We denote the solution for given initial data x_0 by $x(t, x_0)$. Then

$$\frac{d}{dt}\frac{\partial}{\partial x_0}x(t,x_0) = \frac{\partial}{\partial x_0}\frac{\partial}{\partial t}x(t,x_0) = \frac{d}{dx_0}f(x(t,x_0)) = f'(x(t,x_0))\frac{\partial}{\partial x_0}x(t,x_0)$$

and $\frac{\partial}{\partial x_0}x(0,x_0) = \mathrm{I}$. In particular, for $x_0 = 0$ we have

$$f'(x(t,x_0)) = f'(0) = A, \quad \text{i.e. } \frac{\partial}{\partial x_0}x(t,x_0) = e^{At}.$$

Altogehter, we obtain

$$x(t) = e^{At}x(0) + \varphi(t,x_0) \qquad \forall t\colon \varphi(t,x_0) = o(\|x_0\|).$$

For $\|x(0)\| < \delta$ we get

$$\frac{\|x(t)\|}{\|x(0)\|} \le Me^{-\beta t} \Rightarrow \|e^{At}\| \le Me^{-\beta t}.$$

$\Leftarrow$: By assumption and Theorem 3.7 there exists $X > 0$ such that $A^T X + XA = -\mathrm{I}$. Futher, we note that there exists $\delta > 0$ with

$$2x^T X r(x) \le \frac{1}{2}x^T x \qquad \text{for } \|x\| \le \delta.$$

Then we have

$$\frac{d}{dt}x(t)^T X x(t) = -x^T x + 2x^T X r(x) \le -\frac{1}{2}x^T x \le -cx^T X x \quad \text{for some } c > 0.$$

Thus,

$$\|x(t)\|_X^2 \;\le\; e^{-ct}\|x(0)\|_X^2$$

if $\|x(0)\|$ is sufficiently small. □

Remark 3.11. Note that the previous proof justifies that Control Theory mostly considers linear systems. The task at hand is to control the system in such a way that it stays close to its equilibrium. Hence, the assumptions of the theorem are satisfied and the stability properties of the nonlinear system coincide with those of the linear system.

3.3 Routh-Hurwitz criterion

Definition 3.12. A polynomial is called a *Hurwitz polynomial* if all its zeros are in $\mathbb{C}_-$.

Some historic remarks

1868 Maxwell 'On Governors': Give easy criterion for a polynomial to be Hurwitz.

1856 problem solved by Hermite, but it was unknown to engineers

1879 solution by Routh ($\rightarrow$ Adams Prize)

189* Stodola: same question as Maxwell

1895 new variant of Routh criterion

We discuss now the different results: Let $p(t) = \sum_{j=0}^n a_j z^j$ with $a_n > 0$.

Proposition 3.13 (Necessary condition, Stodola)**.** *If p is Hurwitz, then $a_{n-1} > 0, \ldots, a_0 > 0$.*

Definition 3.14. The *Hurwitz matrix* corresponding to the polynomial $p(t)$ is given by

$$H_n = \begin{bmatrix} a_{n-1} & a_{n-3} & a_{n-5} & \dots & \dots & \dots & 0 & 0 & 0 \\ a_n & a_{n-2} & a_{n-4} & & & & \vdots & \vdots & \vdots \\ 0 & a_{n-1} & a_{n-3} & & & & \vdots & \vdots & \vdots \\ \vdots & a_n & a_{n-2} & \ddots & & & 0 & \vdots & \vdots \\ \vdots & 0 & a_{n-1} & & \ddots & & a_0 & \vdots & \vdots \\ \vdots & \vdots & a_n & & & \ddots & a_1 & 0 & \vdots \\ \vdots & \vdots & 0 & & & \ddots & a_2 & a_0 & \vdots \\ \vdots & \vdots & \vdots & & & \ddots & a_3 & a_1 & 0 \\ 0 & 0 & 0 & \dots & \dots & \dots & a_4 & a_2 & a_0 \end{bmatrix}$$

Remark 3.15. In the special case $n = 4$, we have

$$H_4 = \begin{bmatrix} a_3 & a_1 & 0 & 0 \\ a_4 & a_2 & a_0 & 0 \\ 0 & a_3 & a_1 & 0 \\ 0 & a_4 & a_2 & a_0 \end{bmatrix}.$$

Proposition 3.16 (Hurwitz-Criterion)**.** *p is Hurwitz $\Leftrightarrow$ all principal minors of H_n are positive, that is*

$$\Delta_1(p) = a_{n-1} \overset{!}{>} 0,\ \Delta_2(p) = \det \begin{bmatrix} a_{n-1} & a_{n-3} \\ a_n & a_{n-2} \end{bmatrix} \overset{!}{>} 0,\ \Delta_3(p) = \det \begin{bmatrix} a_{n-1} & a_{n-3} & a_{n-5} \\ a_n & a_{n-2} & a_{n-4} \\ 0 & a_{n-1} & a_{n-3} \end{bmatrix} \overset{!}{>} 0,\ \dots.$$

Proposition 3.17 (Routh-Criterion)**.** *p is Hurwitz $\Leftrightarrow$ $a_{n-1} > 0$ and $q(z) = p(z) - \frac{a_n}{a_{n-1}}(a_{n-1}z^n + a_{n-3}z^{n-2} + \dots)$ is Hurwitz.*

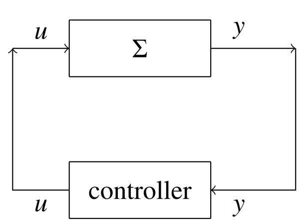
4 Stabilization, Controllability and Observability

We consider the general system Σ given by

$$\dot{x} = Ax + Bu,$$
$$y = Cx + Du$$

with $x \in \mathbb{R}^n, y \in \mathbb{R}^m, u \in \mathbb{R}^p$.

Our **main goal** is to find inputs $u(t)$ based on measurements $y(\tau), 0 \leq \tau \leq t$ such that $x(t) \to 0$ as $t \to \infty$. Basically, we want to stabilize the system. In the following we allow u to be piecewise continuous or L^2, i.e. $u \in U = PC(\mathbb{R}_+, \mathbb{R}^m)$ or $L^2(\mathbb{R}_+, \mathbb{R}^m)$ or $C^k(\mathbb{R}_+ \mathbb{R}^m)$.



The problem will be solved in two steps:

(1) Find $u(t)$ based on the state $x(t)$

(2) Estimate $x(t)$ from the measurements $y(\tau), \tau \in \mathbb{N}$

4.1 Controllability and reachability

We consider $\dot{x} = Ax + Bu$ with the solutions

$$x(t) = x(t, x_0, u) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau, \qquad x(0) = x_0.$$

Definition 4.1. A state $x_1 \in \mathbb{R}^n$ is called *reachable* from $x_0 \in \mathbb{R}^n$ in time t_1 if there exists an input function $u\colon [0, t_1] \to \mathbb{R}^n$ such that $x(t_1, x_0, u) = x_1$. The *reachability space* is denoted by

$$R_t = \{x(t, 0, u) \colon u \in U\} = \{\int_0^t e^{A(t-\tau)} Bu(\tau) d\tau \colon u \in U\}.$$

In this event, x is called *controllable* to x_1 in time t_1. The system is called *controllable*, if for all x_0, x_1 there exists t_1 so that x_0 is controllable to x_1 in time t_1.

Remark 4.2. x_0 is controllable to x_1 in time t_1 iff $x_1 \in e^{At_1} x_0 + R_{t_1}$.

We work with the ansatz

$$x(t_1, x_0, u) = e^{At_1} x_0 + \int_0^{t_1} e^{A(t_1-\tau)} Bu(\tau) d\tau \overset{!}{=} x_1 \qquad \Leftrightarrow \qquad x(t_1, 0, u) = \int_0^{t_1} e^{A(t_1-\tau)} Bu(\tau) d\tau = x_1 - e^{At_1} x_0.$$

and choose $u(\tau) = B^T e^{A^T(t_1-\tau)} v$ such that

$$x(t_1, 0, u) = \underbrace{\int_0^{t_1} e^{A(t_1-\tau)} BB^T e^{A^T(t_1-\tau)} d\tau}_{P_{t_1} \in \mathbb{R}^{n\times n}, P_{t_1} \geq 0} v \overset{!}{=} x_1 - e^{At_1} x_0$$

Note that if $x_1 - e^{At_1} x_0 \in \text{Image}(P_{t_1})$, then such a v can be found.

Definition 4.3. The matrix P_{t_1} is called *controllability Gramian*.

Remark 4.4. One can see via a transformation $s = t_1 - \tau$ that it holds $P_{t_1} = \int_0^{t_1} e^{As} BB^T e^{A^T s} ds$. Thus the integrand is independent of t_1.

Lemma 4.5. *$\forall t_1 > 0$ it holds $R_{t_1} = Image(P_{t_1})$.*

Proof. "$\supseteq$": We argue as above, if $x_1 = P_{t_1} v$, then $x_1 \in R_1$. Thus, we are left to show that "$\subset$" holds. We find

$$R_{t_1} \subset \text{Image}(P_{t_1}) \quad \Leftrightarrow \quad R_{t_1}^{\perp} \supset (\text{Image}(P_{t_1}))^{\perp} = \text{Ker}\, P_{t_1}.$$

And further we observe

$$\begin{aligned} w \in \text{Ker}\, P_{t_1} &\Rightarrow w^T P_{t_1} w = 0 \\ &\Rightarrow \int_0^{t_1} \|w^T e^{A(t_1-\tau)} B\|_2^2 d\tau = 0 \\ &\Rightarrow \forall \tau \in [0, t_1] w^T e^{A(t_1-\tau)} B = 0 \\ &\Rightarrow \forall u \in U \colon \int_0^{t_1} w^T e^{A(t_1-\tau)} Bu(\tau) d\tau = 0 \\ &\Rightarrow w \perp R_{t_1} \end{aligned}$$

□

Definition 4.6. Let $(A, B) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m}$.
The matrix $K(A, B) = [B, AB, A^2B, \ldots, A^{n-1}B] \in \mathbb{R}^{n\times mn}$ is called *Kalman-matrix* or *controllability matrix*.

Theorem 4.7 (Kalman)**.** *For all $t > 0$ it holds $R_t = ImageK(A, B)$.*
In particular, (A, B) controllable $\quad\Leftrightarrow\quad$ rank $K(A, B) = n$.

Proof.

$$\begin{aligned} w \perp R_t &\Leftrightarrow w \perp \text{Image} P_t \Leftrightarrow w^T P_t = 0 \\ &\Leftrightarrow w^T e^{A\tau} B = 0 \text{ for all } \tau \in [0, t] \\ &\Leftrightarrow \sum_{k=0}^{\infty} \tau^k \frac{1}{k!} w^T A^k B = 0 \text{ for all } \tau \in [0, t] \\ &\Leftrightarrow w^T A^k B = 0 \text{ for } k = 0, 1, 2, \ldots \\ &\Leftrightarrow w^T A^k B = 0 \text{ for } k = 0, 1, 2, \ldots, n-1 \\ &\Leftrightarrow w^T K(A, B) = 0 \end{aligned}$$

□

Example (in canonical form)

Let $(A, b) \in \mathbb{R}^{n\times n} \times \mathbb{R}^n$ controllable, then there exists $T \in \mathbb{R}^{n\times n}$ such that

$$T^{-1}AT = \hat{A} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{bmatrix}, \qquad T^{-1}b = \hat{b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

where a_i are the coefficients of the characteristic polynomial $\chi_A(x) = \chi_{\hat{A}}(x)$.

$$\Rightarrow \quad \hat{A}^k\hat{b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ * \\ \vdots \\ * \end{bmatrix} \text{ where the 1 is the } (n-k)\text{th entry} \qquad \rightsquigarrow \qquad K(\hat{A}, \hat{B}) = \begin{bmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & * \end{bmatrix} \text{ is a lower triangular matrix.}$$

Theorem 4.8 (minimal energy control)**.** *Let $w = -e^{At_1}x_o + x_1 \in \text{Image}P_{t_1}$ with $P_{t_1}v = w$. Then $u\colon [0, t_1] \to \mathbb{R}^m, u(\tau) = B^Te^{A^T(t_1-\tau)}v$ is the unique control satisfying $x(t_1, x_0, u) = x_1$ and minimizing the energy functional*

$$\int_0^{t_1} \|u(\tau)\|_2^2 d\tau = \|u\|_{L^2}^2.$$

Proof. We know $x(t_1, x_0, u) = e^{At_1}x_0 + \int_0^{t_1} e^{A(t_1-\tau)}B\underbrace{B^Te^{A^T(t_1-\tau)}v}_{=u}d\tau = e^{At_1}x_0 + w = x_1$. Assume that also $x(t_1, x_0, \bar{u}) = x_1$. Then

$$0 = \int_0^{t_1} e^{A(t_1-\tau)}B\Big(u(\tau) - \bar{u}(\tau)\Big)d\tau$$

and multiplication with v^T from the left yields

$$\begin{aligned} & 0 = v^T\int_0^{t_1} e^{A(t_1-\tau)}B\Big(u(\tau) - \bar{u}(\tau)\Big)d\tau = \int_0^{t_1} u(\tau)^T\Big(u(\tau) - \bar{u}(\tau)\Big)d\tau \\ \Rightarrow \quad & u \perp u - \bar{u} \text{ in } L^2([0, t_1], \mathbb{R}^m) \\ \Rightarrow \quad & \|\bar{u}\|_{L^2}^2 = \|u\|_{L^2}^2 + \|\bar{u} - u\|_{L^2}^2 \geq \|u\|_{L^2}^2 \end{aligned}$$

which equality only for $u = \bar{u}$. □

Remark 4.9. If (A, B) is controllable, then

$$E_C(x_1, t_1) = \min_{x(t_1,0,u)=x_1} \|u\|_{L^2}^2 = x_1^TP_{t_1}^{-1}x_1.$$

Indeed, in this case we get $v = P_{t_1}^{-1}x_1$ which leads to

$$u(\tau) = B^Te^{A^T(t_1-\tau)}P_{t_1}^{-1}x_1$$

and further

$$\begin{aligned} \|u\|_{L^2}^2 &= \int_0^{t_1} u(\tau)^Tu(\tau)d\tau \\ &= \int_0^{t_1} x_1^T(P_{t_1}^{-1})^Te^{A(t_1-\tau)}BB^Te^{A^T(t_1-\tau)}d\tau P_{t_1}^{-1}x_1 \\ &= x_1^T(P_{t_1}^{-1})^TP_{t_1}^TP_{t_1}^{-1}x_1 \\ &= x_1^T(P_{t_1}^{-1}P_{t_1})^TP_{t_1}^{-1}x_1 \\ &= x_1^TP_{t_1}^{-1}x_1. \end{aligned}$$

For $0 < s < t$ we get $P_t - P_s > 0$ and $P_s^{-1} - P_t^{-1} > 0$. Thus

$$E_c(s, x_1) = x_1^TP_s^{-1}x_1 > x_1^TP_t^{-1}x_1 = E_c(t, x_1).$$

This has a physical interpretation saying: if we have less time available to steer the system, we need more energy. Moreover, we can look at $t_1 \to \infty$: if $\sigma(A) \subset \mathbb{C}_-$, then $P_{t_1} \to P = \int_0^\infty e^{A\tau} BB^T e^{A^T\tau} d\tau$, i.e. $AP + PA^T = -BB^T$, where

$$E_C(x_1) = x_1^T P^{-1} x_1.$$

This is an important application of the Lyapunov equation!

Theorem 4.10 (Hautus)**.** *The following statements are equivalent:*

(i) (A, B) *is controllable*

(ii) *If* $A^T v = \lambda v$ *and* $B^T v = 0$ *then* $v = 0$

(iii) *For all* $\lambda \in \mathbb{C}$ *it holds* $\operatorname{rank}(\lambda \mathrm{I} - A, B) = n$*.*

Proof. We begin to show $(\neg(ii)) \Rightarrow (\neg(i))$: Let $A^T v = \lambda v$, $B^T v = 0$. Then

$$K(A,B)^T v = \begin{bmatrix} B^T \\ B^T A^T \\ \vdots \\ B^T (A^{n-1})^T \end{bmatrix} v = \mathbf{0} \qquad \Rightarrow \operatorname{rank} K(A,B) < n.$$

Now we show $(\neg(i)) \Rightarrow (\neg(ii))$: Let $\operatorname{rank} K(A,B) < n$. Then

$$\exists v \in \ker K(A,B)^T \qquad \Rightarrow B^T (A^T)^k v = 0 \text{ for all } k \in \mathbb{N} \qquad \Rightarrow A^T v \in \ker K(A,B)^T.$$

Due to the A^T-invariance of $\ker K(A,B)^T$, $\ker K(A,B)^T$ contains an eigenvector of A^T. Indeed,

$$\exists v \neq 0, A^T v = \lambda v, K(A,B)^T v = 0.$$

In particular, $B^T v = 0$. Finally, we observe that (ii)⇔(iii) is easy. □

4.2 Feedback stabilization

If (A, B) is controllable, then every x_0 is controllable to 0 in arbitrary time t_1.

Example

We consider $\dot{x} = x + u,\ \ x_0 = 1, u(t) = \begin{cases} -2e^{2t_1 - t}(e^{2t_1} - 1)^{-1}, & t \le t_1 \\ 0, & t > t_1 \end{cases}$. We compute

$$\begin{aligned} x(t_1) &= e^{t_1} + \int_0^{t_1} e^{t_1 - \tau}(-2e^{2t_1 - \tau}(e^{2t_1} - 1)^{-1} + \epsilon) d\tau \\ &= e^{t_1} + \int_0^{t_1} -2e^{3t_1 - 2\tau}(e^{2t_1} - 1)^{-1} + e^{t_1 - \tau}\epsilon) d\tau \\ &= e^{t_1} + \left[e^{3t_1 - 2\tau}(e^{2t_1} - 1)^{-1} - \epsilon e^{t_1 - \tau} \right]_{\tau = 0}^{\tau = t_1} \\ &= 0 + \epsilon(e^{t_1} - 1) \to \infty(\text{ as } t_1 \to \infty) \end{aligned}$$

Thus, arbitrarily small perturbations destabilize the system.
Workaround: Feedback mechanism!

Definition 4.11. The system $\dot{x} = Ax + Bu$ is called *stabilizable*, if for all $x_0 \in \mathbb{R}^n$ exists $u = u_{x_0} \in U$ such that

$$x(t, x_0, u_{x_0}) \to 0 \quad \text{as } t \to \infty.$$

It is called *feedback-stabilizable*, if there exists $F \in \mathbb{R}^{m \times n}$ such that $\sigma(A + BF) \subset \mathbb{C}_-$.

Remark 4.12. The following implications are obvious:

$$\begin{aligned} \text{controllable} &\Rightarrow \text{stabilizable} \\ \text{feedback stabilizable} &\Rightarrow \text{stabilizable} \end{aligned}$$

Theorem 4.13 (Wonham)**.** *(A, B) controllable if and only if for all $p \in \mathbb{R}[s]$, $p(s) = s^n + p_{n-1}s^{n-1} + \cdots + p_0$ exists $F \in \mathbb{R}^{m\times n}$ such that $\chi_{A+BF} = p$.*

For the proof we recall the canonical form after state space transformation: Let $x = S\tilde{x}$, $\det S \neq 0$. Then

$$\dot{\tilde{x}} = S^{-1}\dot{x} = S^{-1}(Ax + Bu) = \underbrace{S^{-1}AS}_{\tilde{A}}\,\tilde{x} + \underbrace{S^{-1}B}_{\tilde{B}}\,u, \qquad y = \underbrace{CS}_{\tilde{C}}\,\tilde{x} + Du.$$

In this case we have

$$\begin{aligned} (A, B) \text{ controllable} &\quad\Leftrightarrow\quad (S^{-1}AS, S^{-1}B) \text{ controllable},\\ \sigma(A + BF) &\quad=\quad \sigma(\underbrace{S^{-1}AS}_{\tilde{A}}, S^{-1}B\underbrace{FS}_{\tilde{F}}). \end{aligned}$$

Definition 4.14. $(A, B, C, D) \mapsto (S^{-1}AS, S^{-1}B, CS, D)$ is called *state space transformation*.
The two systems are called *similar*.

proof of Theorem 4.13. Case $m = 1$: "$\Rightarrow$": Let $m = 1, B = b \in \mathbb{R}^m$. Then (A, b) is controllable iff $K(A, b) \in \mathbb{R}^{n\times n}$ is nonsingular. Let $\chi_A(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$. Then

$$AK(A, b) = [Ab, A^2b, \ldots, A^{n-1}b, -\sum_{j=0}^{n-1} a_j A^j b] = K(A, b)\begin{bmatrix} 0 & & & -a_0\\ 1 & & & \vdots\\ & \ddots & & \vdots\\ & & 1 & -a_{n-1} \end{bmatrix} =: K(A, b)M(\chi_A),$$

where $M(\chi_A)$ is uniquely defined by χ_A. For the following arguments we recall that

$$\hat{A} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1\\ -a_0 & -a_1 & \ldots & -a_{n-1} \end{bmatrix}, \qquad \hat{b} = \begin{bmatrix} 0\\ \vdots\\ 0\\ 1 \end{bmatrix} \quad \text{is controllable}$$

and $\chi_A = \chi_{\hat{A}} \Rightarrow \hat{A}K(\hat{A}, \hat{B}) = K(\hat{A}, \hat{b})M(\chi_A)$. We define the short hand notation $K = K(A, B)$, $\hat{K} = K(\hat{A}, \hat{b})$. Then

$$\hat{A} = \hat{K}M(\chi_A)\hat{K}^{-1} = \hat{K}K^{-1}AKK^{-1} = S^{-1}AS, \qquad \hat{b} = \hat{K}K^{-1}b,$$

since $\hat{K}^{-1}\hat{b} = K^{-1}b = e_1$. This shows that (A, b) and $(\hat{A}, \hat{b})$ are similar. Hence it suffices to consider $(\hat{A}, \hat{b})$ for the remainder of the proof. For $\hat{F} = \hat{f} = [\hat{f}_0, \ldots, \hat{f}_{n-1}] \in \mathbb{R}^{1\times n}$ we have

$$\hat{A} + \hat{b}\hat{f} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1\\ \hat{f}_0 - a_0 & \hat{f}_1 - a_1 & \ldots & \hat{f}_{n-1} - a_{n-1} \end{bmatrix} \Rightarrow \chi_{\hat{A}+\hat{b}\hat{f}} = s^n + \sum_{j=0}^{n-1}(\hat{f}_j - a_j)\, s^j \overset{!}{=} p(s).$$

Now, choose $\hat{f}_j = p_j + a_j$, $f = [p_0 - a_0, \ldots, p_{n-1} - a_{n-1}]\hat{K}K^{-1}$, which is the unique solution and proves the result for $m = 1$. □

Before we proceed with the case $m > 1$ we prove **Ackermann's formula:**

$$f = -e_n^T K(A, b)^{-1} p(A).$$

Proof.

$$p(A + bf) = (A + bf)^n + p_{n-1}(A + bf)^{n-1} + \cdots + p_0\,\mathrm{I} \overset{!}{=} 0 \quad \text{(Cayley-Hamilton)}$$

Here it holds

$$(A + bf)^k = A^k + A^{k-1}bf + \sum_{j=0}^{k-2} A^j b z_j$$

for suitable row vectors z_j, which can be proven by an easy induction. Putting this together we obtain

$$-p(A) = A^{n-1}bf + \sum_{j=0}^{n-2} A^j b\tilde{z}_j = K(A, b)[\tilde{z}_0, \ldots, \tilde{z}_{n-1}, f]^T \qquad\Rightarrow\qquad f = -e_n^T K(A, b)^{-1} p(A).$$

□

proof of Theorem 4.13, case $m > 1$. We use the idea of "Heyman's Lemma": Choose $v \in \mathbb{R}^m$ such that $b = Bv \neq 0$ and find $F_1 \in \mathbb{R}^{m\times n}$ such that $(A + BF_1, b)$ is controllable. Due to case $m = 1$ there exists f such that

$$p = \chi_{A+BF_1+bf} = \chi_{A+B(F_1+vf)}.$$

To show the existence of F_1 we construct a maximal number of linearly independent vectors x_j of the form

$$\begin{aligned} x_1 &= b, \\ x_2 &= Ax_1 + Bu_1 = Ax_1 + BF_1x_1 \\ &\vdots \\ x_k &= Ax_{k-1} + Bu_{k-1} = Ax_{k-1} + BF_1x_{k-1}, \end{aligned}$$

with $u_1, \dots, u_{k-1} \in \mathbb{R}^m$ arbitrary to maximize k. We claim that $n = k$. To see this, let $V = \operatorname{span}\{x_1, \dots, x_k\}$. Then the maximality of k yields

$$\begin{aligned} & Ax_k + \operatorname{image} B \subset V \\ \Rightarrow\ & Ax_k \in V, \operatorname{image} B \subset V \\ \Rightarrow\ & Ax_1, \dots, Ax_k \in V, \text{ i.e., } V \text{ is } A\text{-invariant}, \operatorname{image} B \subset V \\ \Rightarrow\ & \operatorname{image} K(A, B) \subset V \\ \Rightarrow\ & \dim V = n. \end{aligned}$$

Hence, we can define F_1 by the equation

$$F_1[x_1, \dots, x_n] = [u_1, \dots, u_{n-1}, 0]$$

such that it holds

$$K(A + BF_1, b) = [b, (A + BF_1)b, \dots] = [x_1, x_2, \dots, x_n], \text{ rank } n \Rightarrow (A + BF_1, b) \text{ is controllable.}$$

□

What happens if (A, B) is not controllable?

Lemma 4.15. *Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ with* $\operatorname{rank} K(A, B) = r < n$. *Then (A, B) is similar to a system $(\tilde{A}, \tilde{B})$ with*

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}, \tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ 0 \end{bmatrix}, \tilde{A}_{11} \in \mathbb{R}^{r\times r}, \tilde{B}_1 \in \mathbb{R}^{r\times m}, (\tilde{A}_{11}, \tilde{B}_1) \textit{ controllable.}$$

Proof. Let $V_1 \in \mathbb{R}^{n\times r}$, $\operatorname{image} V_1 = \operatorname{image} K(A, B)$, $V = [V_1, V_2]$ nonsingular. Then

$$AV = [AV_1, AV_2] = [V_1, V_2]\begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}, \quad B = [V_1, V_2]\begin{bmatrix} \tilde{B}_1 \\ 0 \end{bmatrix}, \quad K(\tilde{A}, \tilde{B}) = \begin{bmatrix} \tilde{B}_1 & \tilde{A}_{11}\tilde{B}_1 & \cdots & \tilde{A}_{11}^{r-1}\tilde{B}_1 & \cdots \\ 0 & 0 & & \cdots & 0 \end{bmatrix}$$

Hence, $\operatorname{rank} K(\tilde{A}_{11}, \tilde{B}_1) = r$. □

We note the following: If $\dot{x} = Ax + Bu, \xi = V^{-1}x = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$, then

$$\begin{aligned} \dot{\xi}_1 &= \tilde{A}_{11}\xi_1 + \tilde{A}_{12}\xi_2 + \tilde{B}_1 u && \text{controllable part} \\ \dot{\xi}_2 &= \tilde{A}_{22}\xi_2 && \text{(completely) uncontrollable part.} \end{aligned}$$

This is called a *Kalman-decomposition*. $\sigma(\tilde{A}_{22}) \subset \mathbb{C}_-$ necessary for stabilizability.

Corollary 4.16. *Let (A, B) have a Kalman decomposition $(\tilde{A}, \tilde{B})$. For $F = \tilde{F}V = [\tilde{F}_1, \tilde{F}_2]V$ we have*

$$\sigma(A + BF) = \sigma(\tilde{A} + \tilde{B}\tilde{F}) = \sigma\left(\begin{bmatrix} \tilde{A}_{11} + \tilde{B}_1\tilde{F}_1 & \tilde{A}_{12} + \tilde{B}_1\tilde{F}_1 \\ 0 & \tilde{A}_{22} \end{bmatrix}\right) = \underbrace{\sigma(\tilde{A}_{11} + \tilde{B}_1\tilde{F}_1)}_{\textit{arbitrary assignment}} \cup \underbrace{\sigma(\tilde{A}_{22})}_{\textit{fixed}}.$$

The system is stabilizable iff $\sigma(\tilde{A}_{22}) \subset \mathbb{C}_-$.

Main observation: General pole placement is only possible if (A, B) is controllable!

Theorem 4.17. *The following statements are equivalent:*

(i) (A, B) *stabilizable*

(ii) $\exists F$ *such that* $\sigma(A + BF) \subset \mathbb{C}_-$

(iii) *in the Kalman form holds* $\sigma(\tilde{A}_{22}) \subset \mathbb{C}_-$

(iv) *if* $A^T v = \lambda v, \lambda \in \bar{\mathbb{C}}_+$ *and* $B^T v = 0$, *then* $v = 0$

(v) $\forall \lambda \in \bar{\mathbb{C}}_+$ *holds* $\operatorname{rank}[\lambda \mathrm{I} - A, B] = n$

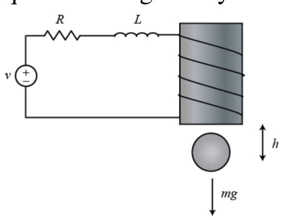
(vi) $\exists P > 0$ *with* $AP + PA^T \leq -BB^T$

Proof. We obtain (i)⇔(ii)⇔(iii)⇔(iv)⇔(v) directly from the previous discussion.
For (vi)⇒(ii) choose $F = B^T P^{-1}$ and the proof of (ii)⇒(vi) is omitted. □

MATLAB-Example: magnetically suspended ball

Idea taken from:
`http://ctms.engin.umich.edu/CTMS/index.php?example=Introduction&section=ControlStateSpace`

We consider the state-space model for a magnetically suspended ball. The current through the coils induces a megnetic force which can balance the force of gravity anc cause the ball (made of magnetic material) to be suspended in mid-air. The equations are given by



$$m\frac{d^2}{dt^2}h = mg - \frac{Ki^2}{h},$$
$$V = L\frac{di}{dt} + iR$$

Here h is the vertical position of the ball, i is the current through the electromagnet, V is the applied voltage, $m = 0.05kg$ is the mass of the ball, $g = 9.81m/s^2$ is the acceleration due to gravity, $L = 0.01$ is the inductance, $R = 1$ is the resistance and $K = 0.0001$ is a coefficient that determines the magnetic force exerted on the ball. We define that the ball is at equilibrium, if the ball is suspended in mid-air, i.e., whenever $h = Ki^2/mg$ at which $dh/dt = 0$. We linearize the equations around the point $h = 0.01m$ to obtain a linear state-space model

$$\frac{d}{dt}x = Ax + Bu,$$
$$y = Cx + Du.$$

where $x = [\Delta h, \Delta\dot{h}, \Delta i]^T$. Play with the corresponding Code in OLAT.

4.3 Observability and detectability

We motivate this section with the linearized inverted pendulum. Let

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{m_g}{M} & -\frac{\mu}{M} & 0 \\ 0 & g\frac{M+m}{Mn} & \frac{\mu}{ML} & 0 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 1/M \\ -1/(ML) \end{bmatrix}, C = [1, L, 0, 0].$$

We have $\det K(A, b) = -\frac{g^2}{L^4 M^4} \rightsquigarrow$ controllable $\rightsquigarrow$ state feedback stabilizable.

Recall: $x = \begin{bmatrix} s \\ \theta \\ \dot{s} \\ \dot{\theta} \end{bmatrix}$ only, $s + L\theta = Cx$ is measured. Is there a stabilizing output feedback $u = Fy, F \in \mathbb{R}$?

$$\dot{x} = Ax + BFy = (A + BFC)x, \quad \sigma(A + BFC) \overset{!}{\subset} \mathbb{C}_-$$

But $\chi_{A+BFC} = \lambda^4 + \frac{\mu}{M}\lambda^3 - \frac{M+m}{LM}\lambda^2 - g\mu\lambda + Fg$ can never be Hurwitz!

Remark 4.18. This is not surprising, since there are 4 eigenvalues and only one parameter F. For $B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}$ with $pm > n$ the general pole-placement problem is generically solvable, but this is quite difficult.

We try a different approach: Estimate $x(t)$ from the measurements $y(\tau), \tau \in [0, t]$ using an observer.

Definition 4.19. An *observer* for a given system Σ is a system $\hat{\Sigma}$ that reconstructs the state of Σ from its input and output signals. The *Luenberger-observer* of

$$\Sigma = \begin{cases} \dot{x} = Ax + Bu, \\ y = Cx + Du \end{cases}$$

has the form

$$\hat{\Sigma} = \begin{cases} \dot{\hat{x}} = A\hat{x} + Bu + K(\hat{y} - y), \\ \hat{y} = C\hat{x} + Du \end{cases}$$

Remark 4.20. The Luenberger observer is a copy of the original system with additional input $\hat{y} - y$. The observer matrix $K \in \mathbb{R}^{n \times p}$ has to be found.

Goal: $\hat{x}(t) - x(t) =: e(t) \to 0, \quad \text{as } t \to \infty.$

We have

$$\dot{e} = \dot{\hat{x}} - \dot{x} = A(\hat{x} - x) + KC(\hat{x} - x) = (A + KC)e.$$

Thus, the task is to find K such that $\sigma(A+KC) \subset \mathbb{C}_-$! We note that $\sigma(A+KC) = \sigma(A^T + C^T K^T) \quad \rightsquigarrow \quad$ feedback stabilization.

Conclusion: There exists $K \in \mathbb{R}^{n \times p}, \sigma(A + KC) \subset \mathbb{C}_- \quad \Leftrightarrow \quad (A^T, C^T)$ stabilizable $\quad \Leftrightarrow: \quad (A, C)$ detectable

Definition 4.21. Let $x(t)$ denote an arbitrary solution of $\dot{x} = Ax$. The pair (A, C) is called *observable*, if $Cx = 0$ for all t implies $x(t) = 0$ for all t. It is called *detectable*, if $Cx = 0$ for all t implies $x(t) \to 0$ as $t \to \infty$.

Lemma 4.22. *(i) (A, C) is observable $\quad \Leftrightarrow \quad (A^T, C^T)$ is controllable*

(ii) (A, C) is detectable $\quad \Leftrightarrow \quad (A^T, C^T)$ is stabilizable.
This is called the duality principle.

Proof. (i) It holds $Cx(t) = Ce^{At}x_0 = \sum_{k=0}^{\infty} \frac{t^k}{k!} CA^k x_0$. Thus

$$Cx \equiv 0 \quad \Leftrightarrow \quad 0 = Cx_0 = CAx_0 = \cdots = CA^{n-1}x_0 \quad \Leftrightarrow \quad x_0 \in \ker \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}.$$

Hence: (A, C) observable iff $\ker K(A^T, C^T)^T = \{0\} \quad \Leftrightarrow \quad (A^T, C^T)$ controllable.

(ii) Using the Kalman decomposition we obtain

$$V^{-1}A^T V = \begin{bmatrix} \tilde{A}_{11}^T & \tilde{A}_{12}^T \\ 0 & \tilde{A}_{22}^T \end{bmatrix} \quad \Leftrightarrow \quad V^T A V^{-T} = \begin{bmatrix} \tilde{A}_{11} & 0 \\ \tilde{A}_{12} & \tilde{A}_{22} \end{bmatrix},$$
$$V^{-1}C^T = \begin{bmatrix} \tilde{C}_1^T \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad CV^{-T} = \begin{bmatrix} \tilde{C}_1 & 0 \end{bmatrix},$$

which implies $(\tilde{A}_{11}, \tilde{C}_1)$ is observable.
For $\xi = V^T x, \dot{\xi} = \tilde{A}\xi, \tilde{C}\xi = 0$, we have

$$\dot{\xi}_1 = \tilde{A}_{11}\xi_1, \tilde{C}_1\xi_1 \equiv 0 \quad \Rightarrow \xi_1 \equiv 0$$
$$\dot{\xi}_2 = \tilde{A}_{22}\xi_2 \text{ unobservable}$$

Altogether, (A, C) detectable $\Leftrightarrow \sigma(\tilde{A}_{22}) \subset \mathbb{C}_- \Leftrightarrow (A^T, C^T)$ stabilizable.

□

Definition 4.23. $O(A, C) = K(A^T, C^T)$ is called *observability matrix*. $Q(t) = \int_0^t e^{A^T\tau} C^T C e^{A\tau} d\tau$ is the *observability Gramian*. $\ker O(A, C) \overset{t>0}{=} \ker Q(t)$ is the *unobservable subspace*.

Corollary 4.24. *(A, C) observable $\Leftrightarrow Q(t) > 0$ for arbitrary $t > 0$. In this case*

$$x_0 = Q(t)^{-1} \int_0^t e^{A^T\tau} C^T y(\tau) d\tau,$$

i.e. the state can be reconstructed from measurements.

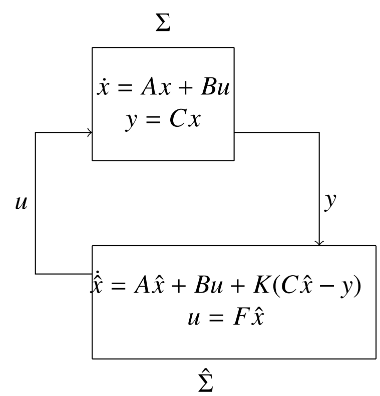
Theorem 4.25. *The following statements are equivalent:*

1) (A, C) *observable*	*1')* (A, C) *detectable*
2) $Av = \lambda v, Cv = 0 \Rightarrow v = 0$	*2')* $Av = \lambda v, \lambda \in \bar{\mathbb{C}}_+, Cv = 0 \Rightarrow v = 0$
3) $\operatorname{rank}\begin{bmatrix} A - \lambda \mathrm{I} \\ C \end{bmatrix} = n$ *for all* $\lambda \in \mathbb{C}$	*3')* $\operatorname{rank}\begin{bmatrix} A - \lambda \mathrm{I} \\ C \end{bmatrix} = n$ *for all* $\lambda \in \bar{\mathbb{C}}_+$

If (A, C) *is detectable, then the state can be reconstructed asymptotically.*

4.4 Dynamic output feedback (dynamic compensation)

We consider the following system with $\sigma(A + BF) \subset \mathbb{C}_-$ and $\sigma(A + KC) \subset \mathbb{C}_-$.



Question: Is the closed loop system asymptotically stable?
This question can be rephrased as follows:

$$\frac{d}{dt}\begin{bmatrix} x \\ \hat{x} \end{bmatrix} = \underbrace{\begin{bmatrix} A & BF \\ -KC & A + BF + KC \end{bmatrix}}_{A_{Cl}} \begin{bmatrix} x \\ \hat{x} \end{bmatrix}$$

Does $\sigma(A_{Cl}) \subset \mathbb{C}_-$ hold?
We can do a similarity transformation. For

$$\begin{bmatrix} \mathrm{I} & 0 \\ -\mathrm{I} & \mathrm{I} \end{bmatrix} A_{Cl} \begin{bmatrix} \mathrm{I} & 0 \\ \mathrm{I} & \mathrm{I} \end{bmatrix} = \begin{bmatrix} A + BF & BF \\ 0 & A + KC \end{bmatrix}$$

it holds

$$\chi_{A_{Cl}} = \chi_{A+BF}\chi_{A+KC}, \qquad \sigma(A_{Cl}) = \sigma(A + BF) \cup \sigma(A + KC) \subset \mathbb{C}_-.$$

Thus, we can draw the conclusion:
The system Σ is stabilizable by dynamic output feedback $\hat{\Sigma}$ if and only if (A, B) is stabilizable and (A, C) is detectable.

MATLAB-Example: magnetically suspended ball 2

Let us consider the example from above and add a observer design, as discussed in the last section. Check out the code in OLAT.

4.5 Controllability and observability of polynomial models

In this section, we relate the terms introduced for the state-space formulation to the formulation as polynomial model. Consider the polynomial model $R\left(\frac{d}{dt}\right) w = 0$.

Definition 4.26. Let $w^{(1)}, w^{(2)}$ be admissible. Then $w^{(1)}$ is *controllable* to $w^{(2)}$, if there exist times $t_1 \leq t_2$ and an admissible signal w such that

$$w(t) = \begin{cases} w^{(1)}(t), & t \leq t_1, \\ w^{(2)}(t), & t \geq t_2 \end{cases}$$

holds. The model is *controllable*, if every $w^{(1)}$ is controllable to every $w^{(2)}$.

Let us have a look at the scalar case, i.e. $R\left(\frac{d}{dt}\right) = \sum_{j=0}^{k} r_j \frac{d^j}{dt^j},\ r_k \neq 0$.
If $k = 0$, then $k_0 w = 0 \Leftrightarrow w = 0$. The system is controllable. If $k > 0$, either $w \equiv 0$ or w has only isolated zeros (identity principle) $\rightsquigarrow$ the system is not controllable.

For the following theorem applies for example to $R(s) = [s\,\mathrm{I} - A, B]$.

Theorem 4.27. *Let $R \in \mathbb{R}[s]^{n\times(n+m)}$, $\operatorname{rank} R = n$. The following statements are equivalent:*

(i) R is controllable

(ii) R has Smith-Form $[I, 0]$

(iii) $\forall s \in \mathbb{C}$ it holds $\operatorname{rank} R(s) = n$ (Hautus)

(iv) $\exists T \in \mathbb{R}[s]^{m\times(n+m)}$, $\begin{bmatrix} R \\ T \end{bmatrix}$ is unimodular

(v) $\exists S \in \mathbb{R}[s]^{(n+m)\times n}$ with $RS = \mathrm{I}_n$

(vi) If $R = U_1 R_1,\ U_1 \in \mathbb{R}[s]^{n\times n}; R_1 \in \mathbb{R}[s]^{n\times(n+m)}$, then U_1 is unimodular.

Proof. (i)⇔(ii) without loss of generality $R = [D, 0]$. Equivalence follows from scalar case.
(ii)⇔(iii) is obvious
(ii)⇔(iv) $R = U[\mathrm{I}, 0]V = U[\mathrm{I}, 0]\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = UV_1 \Rightarrow \begin{bmatrix} R \\ V_2 \end{bmatrix} = \begin{bmatrix} UV_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} U & \\ & \mathrm{I} \end{bmatrix} V$ unimodular, $T = V_2$.
(iv)⇔(v) $\begin{bmatrix} R \\ T \end{bmatrix}^{-1} = [S, Q] \in \mathbb{R}[s]^{(n+m)\times(n+m)} \Rightarrow RS = \mathrm{I}_n$.
(v)⇔(vi) $R = U_1 R_1, RS = \mathrm{I} \Rightarrow U_1 R_1 S = \mathrm{I} \Rightarrow U_1$ unimodular.
(vi)⇔(ii) Let $R = U[D, 0]V = \underbrace{UD}_{U_1}\underbrace{[I, 0]V}_{R_1}$. With U_1 unimodular we get $D = \mathrm{I}$. □

Definition 4.28. $L(s)$ is called observable, if $L\left(\frac{d}{dt}\right)\ell = 0 \Rightarrow \ell = 0$.

Altogeher, we find the following main result:

Duality: R controllable $\Leftrightarrow R^T$ observable.

5 Transfer functions and realizations

We consider matrices of rational functions $H \in \mathbb{R}(s)^{p\times m}$ and *realizations* $H(s) = C(s\,\mathrm{I} - A)^{-1}B + D$. Let $H(s) = \frac{N(s)}{\delta(s)}$ with $N \in \mathbb{R}[s]^{p\times m}, \delta \in \mathbb{R}[s]$.

Remark 5.1. Note the difference in notation $\mathbb{R}(s)$ and $\mathbb{R}[s]$. The smallest field containing both $\mathbb{R}$ and s is denoted by $\mathbb{R}(s)$ and the smallest ring containing both $\mathbb{R}$ and s is denoted by $\mathbb{R}[s]$, In the special case of polynomials, this leads to

$$\mathbb{R}[x] = \Big\{f(x)\colon f \text{ is a polynomial with coefficients in } \mathbb{R}\Big\},$$
$$\mathbb{R}(x) = \left\{\frac{f(x)}{g(x)}\colon f, g \text{ are polynomials with coefficients in } \mathbb{R} \text{ and } g(x) \neq 0\right\}.$$

We introduce a generalization of the Smith form.

Theorem 5.2 (Smith-McMillen form)**.** *For all $H \in \mathbb{R}(s)^{p\times m}$ exists $U \in \mathbb{R}[s]^{p\times p}, V \in \mathbb{R}[s]^{m\times m}$ unimodular such that*

$$UHV = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, D = \operatorname{diag}(\gamma_1/\delta_1, \dots, \gamma_r/\delta_r), \quad \gamma_j, \delta_j \in \mathbb{R}[s]$$
$$\gamma_1 | \dots | \gamma_r, \quad \delta_r | \dots | \delta_1, \gcd(\gamma_j, \delta_j) = 1.$$

where $\gcd(a, b)$ *denotes the greatest common division of a and b.*

Proof. We use the Smith form of N to obtain

$$UHV = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, D = \operatorname{diag}(d_1/\delta, \dots, d_r/\delta).$$

Now, cancellation of terms in d_j/δ yields the desired result. □

Definition 5.3. The number $d(H) = \sum_{j=1}^{r} \deg \delta_j$ is called the *McMillen degree* of H.

Definition 5.4. A factorization...
... $H(s) = P(s)^{-1}Q(s)$ is called *left-coprime*, if $[P, Q]$ is controllable.
... $H(s) = Q(s)P(s)^{-1}$ is called *right-coprime*, if $\begin{bmatrix} P \\ Q \end{bmatrix}$ is observable.

Lemma 5.5. *The following statements hold true:*

1) $d(H) = \min\{\deg\det P\colon H = P^{-1}Q,\ P, Q \text{ polynomial}\}$

1') $d(H) = \min\{\deg\det P\colon H = QP^{-1},\ P, Q \text{ polynomial}\}$

2) $H = P^{-1}Q$ *left-coprime* $\Rightarrow d(H) = \deg\det P.$

2') $H = QP^{-1}$ *right-coprime* $\Rightarrow d(H) = \deg\det P.$

Proof. We prove only 1) and 2). Then 1') and 2') are analogous.
We claim: $\deg\det P_1 \le \deg\det P_2$, if $P_1^{-1}Q_1 = P_2^{-1}Q_2$ and $P_1^{-1}Q_1$ left coprime.

$$[P_1, Q_1] \text{ controllable } \Leftrightarrow \exists S, T\colon P_1S + Q_1T = \mathrm{I}\,.$$

Therefore, it holds
$$P_1^{-1} = S + P_1^{-1}Q_1T = S + P_2^{-1}Q_2T$$
and further
$$P_2 = (P_2S + Q_2T)P_1 \Rightarrow \det P_1 | \det P_2.$$
In particular, if also $P_2^{-1}Q_2$ left coprime, then

$$\deg\det P_1 = \deg\det P_2,$$

so $\deg\det P$ is minimal for left coprime factorizations. If

$$H = \underbrace{U^{-1}\operatorname{diag}(\delta_1,\dots,\delta_r,1,\dots,1)^{-1}}_{=:P^{-1}}\underbrace{\operatorname{diag}(\gamma_1,\dots,\gamma_r,0,\dots,0)V^{-1}}_{=:Q} = P^{-1}Q$$

and
$$[P, Q] = \underbrace{\operatorname{diag}(\delta_1,\dots,\delta_r,1,\dots,1)\operatorname{diag}(\gamma_1,\dots,\gamma_r,0,\dots,)}_{\text{controllable!}}\begin{bmatrix} U & \\ & V^{-1} \end{bmatrix},$$

then $d(H) = \deg\det P.$ □

5.1 Minimal realizations

Definition 5.6. $H \in \mathbb{R}(s)^{p\times q}$ with $H(s) = \frac{N(s)}{\delta(s)}$ is called *proper*, if $\deg\delta \ge \deg N$ and *strictly proper*, if $\deg\delta > \deg N$.
If $H(s) = C(s\,\mathrm{I} - A)^{-1}B + D$, then we call (A, B, C, D) a *realization of* H, sometimes noted as $H = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$. If $H = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ with $A \in \mathbb{R}^{n\times n}$, then n is called *order* of the realization. The realization is called *minimal*, if n is minimal.

Theorem 5.7. *$H \in \mathbb{R}(s)^{p\times q}$ possesses a realization if and only if it is proper.*

Proof. "$\Rightarrow$":
From $H(s) = C(s\,\mathrm{I} - A)^{-1}B + D \to D$ as $|s| \to \infty$, we obtain H is proper.
"$\Leftarrow$": Let H be proper and set $D = \lim_{s\to\infty} H(s)$. Then $H_0(s) = H(s) - D = \frac{N(s)}{\delta(s)}$ is strictly proper. If $\delta(s) = s^n + \delta_{n-1}s^{n-1} + \dots + \delta_1 s + \delta_0$ and $N(s) = N_{n-1}s^{n-1} + \dots + N_1 s + N_0$, then we can choose e.g.

$$A = \begin{bmatrix} 0 & \mathrm{I} & & \\ & & \ddots & \\ & & & \mathrm{I} \\ -\delta_0\,\mathrm{I} & \dots & \dots & -\delta_{n-1}\,\mathrm{I} \end{bmatrix},\ B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathrm{I} \end{bmatrix},\ C = [N_0, \dots, N_{n-1}]$$

or

$$A = \begin{bmatrix} 0 & & & -\delta_0\,\mathrm{I} \\ \mathrm{I} & & & \vdots \\ & \ddots & & \vdots \\ & & \mathrm{I} & -\delta_{n-1}\,\mathrm{I} \end{bmatrix}, \; B = \begin{bmatrix} N_0 \\ \vdots \\ N_{n-1} \end{bmatrix}, \; C = [0, \dots, 0, \mathrm{I}].$$

We show that $H_0(s) = C(s\,\mathrm{I} - A)^{-1}B$ for (i) using the following trick:

$$B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathrm{I} \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} s\,\mathrm{I} & -\,\mathrm{I} & & \\ & & \ddots & \\ & & & -\,\mathrm{I} \\ \delta_0\,\mathrm{I} & & & (s+\delta_{n-1})\,\mathrm{I} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{n-1} \end{bmatrix}$$

Then $sX_k - X_{n-k} = 0$ for $k = 0, \dots, n-2$ which yields $X_k = s^k X_0$ for $k = 0, \dots, n-1$. Further

$$s^n X_0 + \sum_{k=0}^{n-1} \delta_k s^k X_0 = \mathrm{I} \Rightarrow X_0 = \frac{1}{\delta(s)}\,\mathrm{I}$$

and hence

$$(s\,\mathrm{I} - A)^{-1}B = \frac{1}{\delta(s)} \begin{bmatrix} s^0\,\mathrm{I} \\ s^1\,\mathrm{I} \\ \vdots \\ s^{n-1}\,\mathrm{I} \end{bmatrix} \Rightarrow C(s\,\mathrm{I} - A)^{-1}B = H_0(s).$$

□

Next, we want to characterize the minimal realizations. In order to do this, we need the next lemma.

Lemma 5.8. *Let $V = [V_1, V_2, V_3, V_4] \in \mathbb{R}^{n\times n}$ nonsingular, such that image$V_1 = K(A,B) \cap \ker O(A,C)$, image$[V_1, V_2] =$ image$K(A,B)$ and image$[V_1, V_3] = \ker O(A,C)$. Then the transformation $(A,B,C) \to (V^{-1}AV, V^{-1}B, CV)$ leads to the form*

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ 0 & A_{22} & 0 & A_{24} \\ 0 & 0 & A_{33} & A_{34} \\ 0 & 0 & 0 & A_{44} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ 0 \\ 0 \end{bmatrix} u, y = C_2 x_2 + C_4 x_4.$$

Proof. Easy excercise. □

Theorem 5.9. *The realization $H(s) = C(s\,\mathrm{I} - A)^{-1}B + D$ is minimal if and only if (A,B) is controllable and (A,C) is observable. In this case $n = d(H)$. All minimal realizations are similar.*

Proof. "⇒": Assume $H(s) = C(s\,\mathrm{I} - A)^{-1}B + D$, but (A,B) not controllable or (A,C) not observable. Using the previous lemma we get

$$(A,B,C,D) \text{ is similar to } \left(\underbrace{\begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ 0 & A_{22} & 0 & A_{24} \\ 0 & 0 & A_{33} & A_{34} \\ 0 & 0 & 0 & A_{44}, \end{bmatrix}}_{\tilde{A} = V^{-1}AV}, \underbrace{\begin{bmatrix} B_1 \\ B_2 \\ 0 \\ 0 \end{bmatrix}}_{\tilde{B} = V^{-1}B}, \underbrace{\begin{bmatrix} 0, C_2, 0, C_4 \end{bmatrix}}_{\tilde{C} = CV}, D \right),$$

where (A_{22}, B_2) is controllable and (A_{22}, C_2) is observable.

$$H(s) = \tilde{C}(s\,\mathrm{I} - \tilde{A})^{-1}\tilde{B} + \tilde{D} = C_2(s\,\mathrm{I} - A_{22})^{-1}B_2 + D.$$

Hence (A,B,C,D) is not minimal.

"⇐": Let now (A,B) controllable, (A,C) observable and consider another realization $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ such that $C(s\,\mathrm{I} - A)^{-1}B = \tilde{C}(s\,\mathrm{I} - \tilde{A})^{-1}\tilde{B}$ with $A \in \mathbb{R}^{n\times n}$ and $\tilde{A} \in \mathbb{R}^{\tilde{n}\times\tilde{n}}$. We shall show that $\tilde{n} \geq n$.
Since $C(s\,\mathrm{I} - A)^{-1}B = \frac{1}{s}C(\mathrm{I} - \frac{1}{s}A)^{-1}B = \sum_{j=0}^{\infty} CA^j B\, s^{-j-1}$ for large |s|, we have

$$CA^i B = \tilde{C}\tilde{A}^i \tilde{B} \text{ for all } i \in \mathbb{N}.$$

For

$$O = O(A, C) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}, \tilde{O} = O_n(\tilde{A}, \tilde{C}) = \begin{bmatrix} \tilde{C} \\ \tilde{C}\tilde{A} \\ \vdots \\ \tilde{C}\tilde{A}^{n-1} \end{bmatrix}$$

and

$$K = K(A, B) = [B, AB, \ldots, A^{n-1}B], \ \tilde{K} = K_n(\tilde{A}, \tilde{B}) = [\tilde{B}, \tilde{A}\tilde{B}, \ldots, \tilde{A}^{n-1}\tilde{B}]$$

we have

$$OK = \left(CA^{jk-1}B\right)_{j,k=1}^{n} = \tilde{O}\tilde{K} \quad \text{ and } OAK = \tilde{O}\tilde{A}\tilde{K}, OB = \tilde{O}\tilde{B}, CK = \tilde{C}\tilde{K}.$$

Hence rank $\tilde{O}\tilde{K} =$ rank $OK = n$, thus rank $\tilde{O} \geq n$, rank $\tilde{K} \geq n$ and therefore $\tilde{n} \geq n$.
We are left to prove the similarity and $n = d(H)$. We begin with the similarity. Let $(\tilde{A}, \tilde{B})$ controllable, $(\tilde{A}, \tilde{C})$ observable, then $\tilde{O}, \tilde{K}$ have maximal row / column rank. There exists L and R such that

$$L\tilde{O} = \mathrm{I}_n, \tilde{K}R = \mathrm{I}_n, V := KR$$

Multiplying the equations yields

$$\mathrm{I}_n = L\tilde{O}\tilde{K}R = LOKR, \quad V^{-1} = LO.$$

Altogether we obtain the similarity, indeed

$$\begin{aligned} V^{-1}AV &= LOAKR = L\tilde{O}\tilde{A}\tilde{K}R = \tilde{A} \\ V^{-1}B &= LOB = L\tilde{O}\tilde{B} = \tilde{B} \\ CV &= CKR = \tilde{C}\tilde{K}R = \tilde{C}. \end{aligned}$$

Now we prove $n = d(H)$ given (A, B) is controllable and (A, C) is observable.
Set $G(s) = (s\,\mathrm{I} - A)^{-1}B$ which is a left coprime factorization. If $G = QP^{-1}$ is an arbitrary right coprime factorization, then $H = CG = CQP^{-1}$ with $d(H) \leq \deg\det P = d(G) = n$. We have equality, if CQP^{-1} is also right coprime, i.e. rank $\begin{bmatrix} CQ(s) \\ P(s) \end{bmatrix} = m$ for all $s \in \mathbb{C}$. Assume that for some $\lambda \in \mathbb{C}, v \neq 0$ it holds $\begin{bmatrix} CQ(s) \\ P(s) \end{bmatrix} v = 0$. Since $(\lambda\,\mathrm{I} - A)^{-1}B = Q(\lambda)P(\lambda)^{-1}$, we have

$$0 \overset{2!}{=} BP(\lambda)v \overset{1}{=} (\lambda\,\mathrm{I} - A)Q(\lambda)v.$$

Thus either $Q(\lambda)v = 0$, but then $\begin{bmatrix} Q(\lambda) \\ P(\lambda) \end{bmatrix} v = 0$ which contradict that QP^{-1} is right coprime. Or $Q(\lambda)v$ is an eigenvector of A, but then $CQ(\lambda)v = 0$ which contradicts the observability of (A, C). □