



**BERGISCHE
UNIVERSITÄT
WUPPERTAL**

Einführung in die Mathematische Systemtheorie

JOCHEN GLÜCK

Vorlesungsmanuskript
Wintersemester 2023/24

Vorläufig finale Version,
21. April 2025

Inhaltsverzeichnis

1	Differentialgleichungen in einer Veränderlichen	3
1.1	Motivation und Grundbegriffe	3
1.2	Lipschitzstetigkeit und der Banachsche Fixpunktsatz	8
1.3	Lösungen von Anfangswertproblemen: Existenz und Eindeutigkeit .	10
1.4	Lineare Differentialgleichungen und die Matrix-Exponentialfunktion	12
1.5	Inhomogenitäten und Faltung	19
2	Eingänge und Ausgänge	23
2.1	Systeme mit Ein- und Ausgängen	23
2.2	Steuerbarkeit	25
2.3	Beobachtbarkeit	31
3	Langzeitverhalten	35
3.1	Langzeitverhalten der Matrixexponentialfunktion	35
3.2	Kriterien für die Negativität der Spektralschranke	37
4	Rückkopplung und Stabilisierung	41
4.1	Statische Rückkopplung	41
4.2	Polverschiebungssatz	42
4.3	Stabilisierbare Systeme	47
4.4	Port-Hamiltonsche Systeme	49
4.5	Dynamische Rückkopplung	51
5	Beobachter und optimale Steuerung	55
5.1	Luenberger-Beobachter	55
5.2	Lineare Optimalsteuerung mit quadratischen Kosten	57
6	Lineare zeitinvariante Systeme	65
6.1	Die Laplace-Transformation	65
6.2	Zeitinvarianz und Impulsantwort	66
6.3	Transferfunktionen	71
6.4	Realisierungstheorie	75
6.5	Realisierungen positiver Systeme	80
6.6	Externe und interne Stabilität	81

Einleitung

Notation

Folgende Notation wird im Manuskript durchgehend verwendet:

Notation	Bedeutung
$\mathbb{N}$	$\{1, 2, 3, \dots\}$
$\mathbb{N}_0$	$\{0, 1, 2, 3, \dots\}$
$\mathbb{K}$	$\mathbb{R}$ or $\mathbb{C}$
m, n, p	Elemente von $\mathbb{N}$

Stand

Diese Version des Manuskript stammt vom 21. April 2025. Es ist vorläufige die finale Version des Manuskripts.

Kommentare zum Inhalt

- Ich bin Julian Hölz zu Dank verpflichtet, der mich in mehreren Vorlesungen vertreten und die zu diesen Vorlesungen gehörigen Inhalte des Manuskripts erstellt hat.
- Weite Teile des Inhalts von Kapitel 1 wurden aus meinem Vorlesungsmanuskript *Gewöhnliche Differentialgleichungen* aus dem Sommersemester 2020 an der Universität Passau übernommen.
- Gegen Ende der Vorlesung, insbesondere in Kapitel 6, hatte ich an einigen Stellen nicht genügend Zeit, alle Beweise, die ich in der Vorlesung gezeigt habe, für das Manuskript zu tippen. An diesen Stellen steht im Manuskript ein Hinweis darauf, dass der entsprechende Beweis in der Vorlesung gegeben wurde.

Kapitel 1

Differentialgleichungen in einer Veränderlichen

In diesem Kapitel geben wir eine knappe Einführung in die Theorie der gewöhnlichen Differentialgleichungen.¹ Wir behandeln die Dinge nicht in größtmöglicher Allgemeinheit, wollen aber ein paar wichtige Grundprinzipien herausstellen. In Abschnitt 1.3 zeigen wir eine einfache Variante des Existenz- und Eindeigkeitssatzes von Picard–Lindelöf für (im Allgemeinen) nichtlineare Differentialgleichungen. Im Rest des Kapitels konzentrieren wir uns auf lineare Differentialgleichungen, die ein wichtiger Spezialfall sind. In allen verbleibenden Kapiteln diskutieren wir **lineare Systemtheorie**. Die lineare Theorie eignet sich gut für eine Einführung in die Systemtheorie, weil sich viele wichtige Konzepte damit gut einführen und illustrieren lassen und weil sie zugleich einfacher zugänglich als die nichtlineare Theorie ist.

1.1 Motivation und Grundbegriffe

In quantitativen Wissenschaften betrachtet man sehr häufig Größen, die sich mit der Zeit ändern – z.B. Ortsangaben in der Physik, Konzentrationen in der Chemie, Populationsgrößen in der Biologie, Aktienkurse in der Finanzmathematik, usw. Wenn wir solch eine Größe mit dem Symbol x bezeichnen und die Zeit mit t , dann ist x also eine Funktion, die von der Variablen t abhängt. In vielen Modellen ist es unmöglich, die Funktion x ad hoc zu bestimmen. Hier einige sehr einfache Beispiele:

Beispiele 1.1.1. (a) Sei b die Anzahl Bakterien in einer Petrischalen-Kultur. Welche Form hat die von der Zeit abhängige Funktion b ?

(b) Es bezeichne $x(t) \in \mathbb{R}^3$ die Position eines Satelliten zum Zeitpunkt t (in diesem Fall ist $x(t)$ also nicht eine einzelne Größe, sondern ein Vektor, der aus

¹Hinweis: Manche Teile dieses Kapitels – insbesondere in den Abschnitten 1.1, 1.2 und 1.3 – wurden aus meinem Manuskript *Gewöhnliche Differentialgleichungen* (SoSe 2020, Universität Passau) übernommen und an diese Vorlesung angepasst.

drei Größen besteht). Wenn Sie den Ort $x(0)$ sowie die Geschwindigkeit zum Zeitpunkt 0 kennen, woher wissen Sie dann, wo sich der Satellit zum Zeitpunkt t befinden wird?

- (c) Es bezeichne k die Konzentration des Isotops C^{14} in der gesamten Kohlenstoffmenge einer archäologischen Probe. Wie hängt k von der Zeit t ab?

Während es in vielen – erst recht in komplizierteren – Situationen nicht möglich ist, ad hoc zu erkennen, von welcher Gestalt die gesuchte Funktion sein muss, ist es andererseits häufig möglich einen Zusammenhang zwischen der Größe und ihrer zeitlichen Änderung, d.h. ihrer Ableitung, anzugeben. Wir veranschaulichen das anhand der vorangehenden Beispiele 1.1.1:

Beispiele 1.1.2. (a) Bakterien vermehren sich durch Zellteilung, und solange sie genügend Ressourcen und Platz zur Verfügung haben und die Rahmenbedingungen sich nicht ändern, tun sie dies mit einer festen Rate. Deshalb ist es sinnvoll anzunehmen, dass die Zunahme der Bakterienanzahl, d.h. die Ableitung $\dot{b}(t)$, proportional zur Anzahl der bereits vorhandenen Bakterien ist.

Dies bedeutet, dass es eine reelle Zahl $c > 0$ gibt mit der Eigenschaft

$$\dot{b}(t) = cb(t).$$

Diese Gleichung ist plausibel, solange die Anzahl der Bakterien nicht zu groß ist – denn wenn die Petrischale sich füllt, bleiben nicht mehr genügend Ressourcen um die Reproduktionsrate aufrecht zu erhalten.

- (b) In der klassischen Mechanik gelten die sogenannten *Newtonschen Bewegungsgleichungen*: Sie besagen, dass die Beschleunigung a , die ein Objekt erfährt, multipliziert mit seiner Masse m gleich der Kraft F ist, die auf das Objekt wirkt, d.h. es gilt $ma = F$.

Weil $x(t)$ die Position des Satelliten zum Zeitpunkt t ist, ist seine Geschwindigkeit zum Zeitpunkt t gleich der Ableitung $\dot{x}(t)$ und seine Beschleunigung a gleich der zweiten Ableitung $\ddot{x}(t)$. Es gilt also $m\ddot{x}(t) = F$.

Die Gravitationskraft, die auf einen Satelliten wirkt hängt aber von einer aktuellen position ab – sie nimmt zum Beispiel ab, wenn der Satellit sich weiter von der Erde entfernt. Das heißt, F ist in Wirklichkeit eine Funktion, die von der aktuellen Position des Satelliten abhängt. Somit erhalten wir insgesamt die Gleichung

$$m\ddot{x}(t) = F(x(t))$$

für die Position des Satelliten. (In der Physik ist eine genaue Formel für $F(x(t))$ bekannt – Sie können diese unter dem Stichwort *Newtonsches Gravitationsgesetz* nachlesen.)

- (c) Aus der Theorie des radioaktiven Zerfalls weiß man, dass ein C14-Atom (ebenso wie jedes andere radioaktive Atom) in einem vorgegebenen Zeitraum mit einer festen Wahrscheinlichkeit zerfällt. Also ist die Rate mit der die Konzentration $k(t)$ abnimmt – d.h. die negative Ableitung $-\dot{k}(t)$ – direkt proportional zur momentanen Konzentration $k(t)$. Es gibt also eine Zahl $c > 0$ so, dass die Gleichung

$$-\dot{k}(t) = ck(t)$$

gilt.

An den Beispielen 1.1.2 erkennt man folgendes: Die Beschreibung einer zeitabhängigen Größe x passiert in vielen Modellen nicht, indem man direkt eine Formel für x angibt. Stattdessen findet man häufig eine Gleichung für den Zusammenhang zwischen x und ihrer Ableitung $\dot{x}$ (oder auch ihren höheren Ableitungen wie z.B. $\ddot{x}$). Eine solche Gleichung nennt man eine **Differentialgleichung**. In den vorangehenden Beispielen hängt die gesuchte Funktion nur von einer reellen Variablen ab (in den Beispielen von der Zeit t) – in solch einer Situation spricht man von einer **gewöhnlichen Differentialgleichung**. Die höchste vorkommenden Ableitung der gesuchten Funktion ist die **Ordnung** der Differentialgleichung.

Lassen Sie uns nun kurz ansprechen, inwiefern man die **Eindeutigkeit** von Lösungen einer Differentialgleichungen anschaulich erwarten kann. Betrachten wir wieder die drei konkreten Situationen aus den Beispielen 1.1.1 und 1.1.2:

Beispiele 1.1.3. (a) Die Anzahl an Bakterien $b(t)$ zum Zeitpunkt t haben wir durch die Differentialgleichung

$$\dot{b}(t) = cb(t)$$

für eine Konstante $c > 0$ beschrieben. Können wir erwarten, dass es nur eine Lösung dieser Differentialgleichung gibt?

Stellen Sie sich vor, im Labor werden mehrere Bakterienkulturen in verschiedenen Petrischalen gezüchtet; nehmen wir an, diese Petrischalen sollen alle identisch sein, mit der Ausnahme dass sich die Anzahl der Bakterien beim Ansetzen der Kulturen zwischen den einzelnen Schalen wesentlich unterscheidet. Dann wäre es nicht sinnvoll anzunehmen, dass die Anzahl der Bakterien in jeder Schale zu jedem Zeitpunkt gleich ist – sie sind ja noch nicht einmal zum Zeitpunkt $t = 0$ gleich!

- (b) Sei wieder $x(t) \in \mathbb{R}^3$ die Position unseres Satelliten zum Zeitpunkt t . Aus den Newtonschen Bewegungsgleichungen hatten wir die Differentialgleichung

$$m\ddot{x}(t) = F(x(t))$$

erhalten, wobei $F(x(t))$ die Kraft bezeichnet, die auf den Satelliten wirkt, wenn er sich an der Stelle $x(t)$ befindet. Können wir erwarten, dass diese Differentialgleichung nur eine Lösung hat?

Das wäre mehr als verwegen, denn wo der Satellit sich zum Zeitpunkt t befindet, wird ja nicht nur vom Kraftfeld der Erdanziehung abhängen, sondern auch davon, wo der Satellit sich befunden hat, als er von einem Raumfahrzeug ausgesetzt wurde – also von seiner Position $x(0)$ zum Zeitpunkt 0.

Selbst wenn aber die Position des Satelliten zum Zeitpunkt 0 bekannt ist, können wir nicht erwarten, dass dadurch seine Bahn bereits bestimmt ist: Stellen Sie sich zwei verschiedene Satelliten vor, die an der selben Position starten, aber zum Startzeitpunkt einen deutlich verschiedenen Geschwindigkeitsvektor besitzen. Dann werden die beiden Positionen dieser Satelliten sofort auseinander laufen. Also müssen wir nicht nur die Position x zum Zeitpunkt 0, sondern auch die Geschwindigkeit $\dot{x}(0)$ zum Zeitpunkt 0 kennen, damit wir erwarten können, dass die Bahn des Satelliten (d.h. die Lösung x unserer Differentialgleichung) durch unsere Kenntnisse eindeutig bestimmt ist.

- (c) Mit der Konzentration $k(t)$ des Kohlenstoff-Isotops C14 in unserer archäologischen Probe verhält es sich ähnlich wie mit der Bakterienkultur. Wir hatten diese durch die Differentialgleichung

$$-\dot{k}(t) = ck(t) \tag{1.1.1}$$

beschrieben (mit einer festen Konstante $c > 0$). Anschaulich ist es aber nicht sinnvoll zu erwarten, dass diese Differentialgleichung nur eine Lösung hat – denn wenn wir verschiedene Proben mit verschiedenen Startkonzentrationen betrachten, so wird in jeder Probe der zeitliche Verlauf der C14-Konzentration die Differentialgleichung (1.1.1) lösen. Aber trotzdem wird der zeitliche Verlauf nicht in allen Proben gleich sein (d.h. es gibt mehrere Funktionen k , die (1.1.1) lösen) – den die C14-Konzentration kann sich ja vielleicht schon zum Zeitpunkt 0 unterscheiden (je nachdem, welche Proben wir vorliegen haben).

Dass also die Differentialgleichung (1.1.1) nur eine Lösung besitzt, können wir vernünftiger Weise nur dann erwarten, wenn wir denn Wert von k zum Zeitpunkt 0 fixieren.

In den oben stehenden Beispielen erkennen Sie, dass eine gewöhnliche Differentialgleichungen üblicherweise keine eindeutige Lösung hat. Damit man überhaupt eine Chance auf Eindeutigkeit hat, muss man zusätzlich den Wert der gesuchten Funktion zu einem Zeitpunkt t_0 (den man in konkreten Beispielen oft als 0 wählt) vorgeben. Diesen Zeitpunkt t_0 interpretiert man oft als Anfangszeitpunkt – wenn man für eine Differentialgleichung den Wert der gesuchten Funktion zum Zeitpunkt t_0 vorgibt, dann erhält man ein **Anfangswertproblem**, dass aus der Differentialgleichung und der Anfangsbedingung besteht.

Die folgende Proposition zeigt, wie man ein Anfangswertproblem in eine Integralgleichung umschreiben kann. Weil es später sehr nützlich sein wird, betrachten wir nicht nur Differentialgleichungen im $\mathbb{R}^n$, sondern auch im $\mathbb{C}^n$. Sei also $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

Proposition 1.1.4. *Sei $f : \mathbb{K}^n \rightarrow \mathbb{K}^n$ eine stetige Abbildung; sei zudem $t_0 \in \mathbb{R}$ und $x_0 \in \mathbb{K}^n$. Sei $J \subseteq \mathbb{R}$ ein Intervall mit mehr als einem Punkt, das t_0 enthält. Für jede Funktion $x : J \rightarrow \mathbb{K}^n$ sind die folgenden Aussagen äquivalent:*

(i) *Die Funktion x ist differenzierbar und erfüllt*

$$\begin{cases} \dot{x}(t) = f(x(t)) & \text{für alle } t \in J, \\ x(t_0) = x_0. \end{cases} \quad (1.1.2)$$

(ii) *Die Funktion x ist stetig und es gilt*

$$x(t) = x_0 + \int_{t_0}^t f(x(s)) \, ds \quad (1.1.3)$$

für alle $t \in J$.

Falls die äquivalenten Aussagen (i) and (ii) erfüllt sind, dann ist x sogar stetig differenzierbar.

Man beachte, dass das Integral in (1.1.3) wohldefiniert ist (als Riemann-Integral), denn aus der Stetigkeit von x und f folgt, dass der Integrand stetig ist.

Eine Funktion $x : J \rightarrow \mathbb{K}^d$, die die Aussage (i) in Proposition 1.1.4 erfüllt, nennt man eine **Lösung** des Anfangswertproblems 1.1.2.

Beweis von Proposition 1.1.4. “(i) $\Rightarrow$ (ii)” Es gelte (i). Weil x differenzierbar ist, ist x insbesondere stetig. Da f nach Voraussetzung ebenfalls stetig ist, folgt aus der Differentialgleichung, die x löst, dass $\dot{x}$ ebenfalls stetig ist. Also ist x sogar stetig differenzierbar; somit können wir den Hauptsatz der Differential- und Integralrechnung anwenden. Mit dessen Hilfe erhalten für jedes $t \in J$

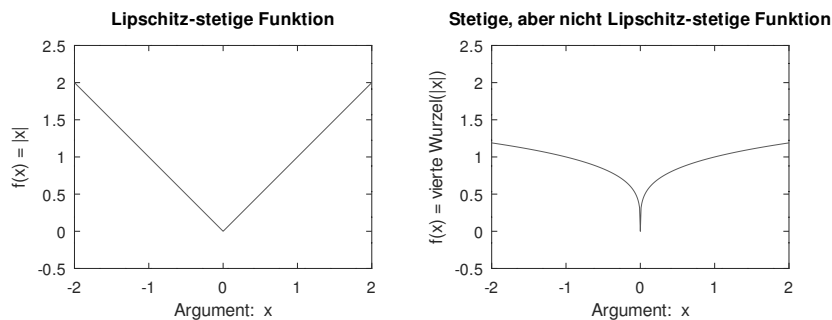
$$\int_{t_0}^t f(x(s)) \, ds = \int_{t_0}^t \dot{x}(s) \, ds = x(t) - x(t_0) = x(t) - x_0;$$

dies zeigt (ii).

“(ii) $\Rightarrow$ (i)” Es gelte nun (ii). Dann ist $x(t_0) = x_0 + 0 = x_0$. Aus dem Hauptsatz der Differential- und Integralrechnung folgt, dass x stetig differenzierbar ist und dass $\dot{x}(t) = f(x(t))$ für alle $t \in J$ gilt. Dies zeigt (i). $\square$

In den nächsten beiden Abschnitten werden wir Proposition 1.1.4 verwenden um ein Kriterium für Existenz und Eindeutigkeit der Lösungen von Anfangswertproblemen herzuleiten.

Abbildung 1.2.1: Graphische Darstellung der Lipschitz-stetigen Funktion $[-2, 2] \ni x \mapsto |x| \in \mathbb{R}$ und der Funktion $[-2, 2] \ni x \mapsto |x|^{1/4}$, die zwar stetig, aber nicht Lipschitz-stetig ist.



1.2 Lipschitzstetigkeit und der Banachsche Fixpunktsatz

Definition 1.2.1 (Lipschitz-Stetigkeit). Seien (M, d_M) und (N, d_N) metrische Räume. Eine Abbildung $f : M \rightarrow N$ heißt *Lipschitz-stetig*, wenn es eine reelle Zahl $L \geq 0$ gibt derart, dass

$$d_N(f(x), f(y)) \leq L d_M(x, y) \quad \text{für alle } x, y \in M \quad (1.2.1)$$

gilt. In diesem Fall heißt die kleinste reelle Zahl $L \geq 0$, für die (1.2.1) gilt, die *Lipschitz-Konstante* von f .²

Anschaulich bedeutet Lipschitz-Stetigkeit, dass eine Funktion „sich nirgends schneller als linear ändert“. Ein einfaches Beispiel einer Lipschitz-stetigen Funktion ist

$$[-2, 2] \ni x \mapsto |x| \in \mathbb{R};$$

dass diese Funktion Lipschitz-stetig ist, kann man mit Hilfe der unteren Dreiecksungleichung beweisen. Ein Beispiel einer Funktion, die stetig, aber nicht Lipschitz-stetig ist, ist

$$[-2, 2] \ni x \mapsto |x|^{1/4} \in \mathbb{R};$$

der Grund, weshalb diese Funktion nicht Lipschitz-stetig ist, ist ihr Verhalten im Punkt $x = 0$: je näher man diesem Punkt kommt, desto schneller ändert sich die Funktion – man kann, anschaulich gesprochen, ihre Änderung also nicht durch eine lineare Funktion abschätzen. Die beiden soeben diskutierten Funktionen sind in der Abbildung 1.2.1 graphisch dargestellt.

²Warum gibt es überhaupt eine kleinste Zahl $L \geq 0$, für die (1.2.1) gilt?

Aufgaben 1.2.2. (a) Seien (M_k, d_k) für $k = 1, 2, 3$ metrische Räume und seien $f : M_1 \rightarrow M_2$ and $g : M_2 \rightarrow M_3$ Lipschitz-stetig mit Lipschitz-Konstanten L_f und L_g . Zeigen Sie: Die Hintereinanderausführung $g \circ f : M_1 \rightarrow M_3$ ist ebenfalls Lipschitz-stetig mit Lipschitz-Konstante höchstens $L_g L_f$.

(b) Sei (M, d) ein metrischer Raum und sei $f : M \rightarrow M$ Lipschitz-stetig mit Lipschitz-Konstante $L \geq 0$. Zeigen Sie: Für jedes $n \in \mathbb{N}_0$ ist f^n (die n -fache Hintereinanderausführung von f mit sich selbst) Lipschitz-stetig mit Lipschitz-Konstante höchstens L^n .

Die folgende Variante des Banachschen Fixpunktsatzes eignet sich besonders gut um in Abschnitt 1.3 den Existenz- und Eindeutigkeitssatz von Picard–Lindelöf zu beweisen. Eine andere (und vielleicht üblichere) Variante, die voraussetzt, dass die Lipschitz-Konstante von f selbst < 1 ist, kann man direkt daraus folgern (siehe Übungsblatt 1).

Satz 1.2.3 (Banachscher Fixpunktsatz³). *Sei (M, d) ein nicht-leerer vollständiger metrischer Raum und sei $f : M \rightarrow M$ eine Lipschitz-stetige Abbildung. Für jedes $n \in \mathbb{N}_0$ bezeichne L_n die Lipschitzkonstante von f^n und es sei $\sum_{k=0}^{\infty} L_k < \infty$.*

Dann besitzt f genau einen Fixpunkt $\hat{x} \in M$. Außerdem konvergiert für jedes $x \in M$ die Folge der Iterierten $(f^k(x))_{k \in \mathbb{N}}$ gegen $\hat{x}$.

Beweis. Eindeutigkeit: Seien $\hat{x}_1$ und $\hat{x}_2$ Fixpunkte von f . Aus $\sum_{k=0}^{\infty} L_k < \infty$ folgt $L_k \rightarrow 0$ für $k \rightarrow \infty$, also können wir $k \in \mathbb{N}$ so groß wählen, dass $L_k < 1$ gilt. Dann ist

$$d(\hat{x}_1, \hat{x}_2) = d(f^k(\hat{x}_1), f^k(\hat{x}_2)) \leq L_k d(\hat{x}_1, \hat{x}_2).$$

Wegen $L_k < 1$ folgt hieraus $d(\hat{x}_1, \hat{x}_2) = 0$, also $\hat{x}_1 = \hat{x}_2$.

Konvergenz von $(f^k(x))_{k \in \mathbb{N}}$ und Existenz eines Fixpunktes: Sei $x \in M$ beliebig. Wir zeigen, dass $(f^k(x))_{k \in \mathbb{N}}$ eine Cauchy-Folge ist; sei dazu $\varepsilon > 0$. Wegen der Summierbarkeit der L_k gibt es ein $k_0 \in \mathbb{N}_0$ mit $d(x, f(x)) \sum_{k=k_0}^{\infty} L_k \leq \varepsilon$. Für jedes $k_1 \geq k_0$ gilt somit

$$d(f^{k_0}(x), f^{k_1}(x)) \leq \sum_{k=k_0}^{k_1-1} d(f^k(x), f^{k+1}(x)) \leq \sum_{k=k_0}^{k_1-1} L_k d(x, f(x)) \leq \varepsilon.$$

Also ist $(f^k(x))_{k \in \mathbb{N}}$ tatsächlich eine Cauchy-Folge und konvergiert somit wegen der Vollständigkeit von (M, d) gegen ein $\hat{x} \in M$. Dieses $\hat{x}$ ist wegen

$$f(\hat{x}) = f\left(\lim_{k \rightarrow \infty} f^k(x)\right) = \lim_{k \rightarrow \infty} f(f^k(x)) = \lim_{k \rightarrow \infty} f^{k+1}(x) = \hat{x}$$

ein Fixpunkt von f . □

³Benannt nach Stefan Banach (1892 – 1945), polnischer Mathematiker; Mitbegründer der Funktionalanalysis und einer der bedeutendsten Mathematiker in der ersten Hälfte des 20. Jahrhunderts.

1.3 Lösungen von Anfangswertproblemen: Existenz und Eindeutigkeit

Satz 1.3.1 (Globaler Existenz- und Eindeutigkeitsatz von Picard–Lindelöf⁴). *Sei $f : \mathbb{K}^n \rightarrow \mathbb{K}^n$ eine Lipschitz-stetige Funktion. Zudem seien $t_0 \in \mathbb{R}$ und $x_0 \in \mathbb{K}^n$. Dann gibt es genau eine differenzierbare Funktion $x : \mathbb{R} \rightarrow \mathbb{K}^n$, die das Anfangswertproblem*

$$\begin{cases} \dot{x}(t) = f(x(t)) & \text{für alle } t \in \mathbb{R}, \\ x(t_0) = x_0 \end{cases} \quad (1.3.1)$$

löst.

Beweis. Wir verstehen $\mathbb{K}^n$ im ganzen Beweis mit der Euklidischen Norm $\|\cdot\|$. Sei $J \subseteq \mathbb{R}$ ein kompaktes Intervall mit mehr als einem Punkt, das t_0 enthält. Wir zeigen, dass es genau eine differenzierbare Funktion $x : J \rightarrow \mathbb{K}^n$ gibt, die

$$\begin{cases} \dot{x}(t) = f(x(t)) & \text{für alle } t \in J, \\ x(t_0) = x_0 \end{cases} \quad (1.3.2)$$

erfüllt. Dies genügt um die Behauptung zu folgern.⁵

Es bezeichne $C(J; \mathbb{K}^n)$ den Raum der stetigen Funktionen von J nach $\mathbb{K}^n$; wir statten ihn mit der Supremumsnorm $\|\cdot\|_\infty$ aus, die durch $\|x\|_\infty := \sup_{t \in J} \|x(t)\|$ für alle $x \in C(J; \mathbb{K}^n)$ gegeben ist. Es ist ein Standardresultat aus der Analysis, dass $C(J; \mathbb{K}^n)$ ein Banachraum und somit insbesondere ein vollständiger metrischer Raum ist.⁶ Wir betrachten die Abbildung $T : C(J; \mathbb{K}^n) \rightarrow C(J; \mathbb{K}^n)$, folgendermaßen definiert ist: für jedes $x \in C(J; \mathbb{K}^n)$ sei $T(x) \in C(J; \mathbb{K}^n)$ gegeben durch

$$(T(x))(t) = x_0 + \int_{t_0}^t f(x(s)) \, ds \quad \text{für alle } t \in J.$$

Laut Proposition 1.1.4 sind die Lösungen des Anfangswertproblems (1.3.2) genau die Fixpunkte der Funktion T .

Wir werden nun den Banachschen Fixpunktsatz 1.2.3 verwenden um zu zeigen, dass T genau einen Fixpunkt hat. Dazu müssen wir zeigen, dass die Iterierten T^k Lipschitz-stetig sind und dass ihre Lipschitz-Konstanten summierbar sind. Zu diesem Zweck zeigen wir für alle $x, y \in C(J; \mathbb{K}^n)$, für jedes $k \in \mathbb{N}_0$ und jedes $t \in J$ die Ungleichung

$$\left\| (T^k(x))(t) - (T^k(y))(t) \right\| \leq \frac{|t - t_0|^k L^k}{k!} \|x - y\|_\infty, \quad (1.3.3)$$

⁴Benannt nach Ernst Leonard Lindelöf (1870 – 1946), finnischer Mathematiker, und Charles Émile Picard (1856 – 1941), französischer Mathematiker.

⁵Wieso?

⁶Wie hängen hier nochmal Metrik und Norm zusammen?

wobei $L \geq 0$ die Lipschitz-Konstante von f bezeichnet. Wir zeigen dies zunächst für alle $t \geq t_0$. Für $k = 0$ ist die Aussage offensichtlich; also sei die Aussage nun für ein festes $k \in \mathbb{N}_0$ und alle $t_0 \leq t \in J$ bereits bewiesen. Dann gilt

$$\begin{aligned} \left\| (T^{k+1}(x))(t) - (T^{k+1}(y))(t) \right\| &= \left\| \int_{t_0}^t f\left((T^k(x))(s)\right) - f\left((T^k(y))(s)\right) ds \right\| \\ &\leq \int_{t_0}^t L \left\| (T^k(x))(s) - (T^k(y))(s) \right\| ds \\ &\leq \int_{t_0}^t L \frac{s^k L^k}{k!} ds \|x - y\|_\infty \\ &= \frac{(t - t_0)^{k+1} L^{k+1}}{(k+1)!} \|x - y\|_\infty. \end{aligned}$$

Für $t \leq t_0$ zeigt man (1.3.3) auf dieselbe Weise (man muss nur etwas auf die Beträge und das Vorzeichen achten).

Wenn wir $\ell \geq 0$ als die Länge des Intervalls J wählen, dann liefert (1.3.3) die Abschätzung

$$\left\| T^k(x) - T^k(y) \right\|_\infty \leq \frac{\ell^k L^k}{k!} \|x - y\|_\infty$$

für alle $k \in \mathbb{N}_0$ und alle $x, y \in C(J; \mathbb{K}^n)$ – d.h., jedes T^n ist Lipschitz-stetig und seine Lipschitz-Konstante ist höchstens $\frac{\ell^n L^n}{n!}$. Diese Zahlen sind summierbar über $n \in \mathbb{N}_0$ (mit Summe $\exp(\ell L)$), also ist die Voraussetzung des Banachschen Fixpunktsatzes 1.2.3 erfüllt. Somit besitzt T genau einen Fixpunkt in $C(J; \mathbb{K}^n)$, was zu zeigen war. $\square$

Bemerkung 1.3.2 (Picard–Lindelöf-Iterierte). Der Beweis von Satz 1.3.1 zeigt nicht nur, dass das Anfangswertproblem (1.3.1) genau eine Lösung besitzt, sondern er gibt auch eine Möglichkeit an, um die Lösung mittels eines Iterationsverfahren zu bestimmen:

Wenn man nämlich ein nicht-triviales kompaktes Intervall $J \subseteq \mathbb{R}$ betrachtet, das t_0 enthält, und mit einer beliebigen Funktion $y \in C(J; \mathbb{R}^d)$ startet, dann konvergiert die Funktionenfolge $(T^k x)_{k \in \mathbb{N}}$ in $\|\cdot\|_\infty$ -Norm gegen die Lösung von (1.3.1) auf dem Intervall J . Das liegt daran, dass die Iterierten im Banachschen Fixpunktsatz 1.2.3 für jeden Startwert gegen das Fixpunkt konvergieren.

Die Funktionen $T^k x$ nennt man die *Picard–Lindelöf-Iterierten* zu x .

Bemerkung 1.3.3 (Verallgemeinerungen). Satz 1.3.1 hat recht spezielle Voraussetzungen, weshalb man ihn bei weitem nicht auf alle interessanten Anfangswertprobleme anwenden kann. Zum Beispiel gibt es folgende Einschränkungen:

- Wir haben nur Differentialgleichungen erster Ordnung betrachtet. Das ist aber keine große Einschränkung, weil man eine Differentialgleichung höherer Ordnung in eine Differentialgleichung erster Ordnung umschreiben kann. Auf Übungsblatt 1 werden Sie anhand eines Beispiels sehen, wie das geht.

- Wir haben angenommen, dass die rechte Seite der Differentialgleichung $f(x(t))$ lautet – die Zeit t geht auf der rechten Seite also nur indirekt ein, weil $x(t)$ von t abhängt. Differentialgleichungen mit dieser Eigenschaft nennt man **autonom**.

Eine Differentialgleichung von der Form $\dot{x}(t) = f(t, x(t))$ nennt man nicht-autonom. Für solche Differentialgleichungen kann man ein ähnliches Resultat wie in Satz 1.3.1 beweisen, wenn man die Voraussetzungen an f geeignet anpasst.

Eine nicht-autonome Differentialgleichung kann man zum Beispiel verwenden um eine Bakterienpopulation zu modellieren, deren Wachstumsrate von der Zeit abhängt (zum Beispiel, wenn sich die Temperatur über die Zeit ändert).

- Wir haben angenommen, dass f auf ganz $\mathbb{R}^d$ definiert ist und das f außerdem Lipschitz-stetig ist. In Anwendungen hat man aber häufig eine rechte Seite, die nur auf einer offenen Teilmenge $\Omega \subseteq \mathbb{K}^n$ definiert ist und die zudem nicht Lipschitz-stetig, sondern nur **lokal Lipschitz-stetig**⁷ sind. Für solche rechten Seite kann man ebenfalls einen Existenz- und Eindeutigkeitssatz beweisen; man spricht häufig vom **lokalen** Existenz- und Eindeutigkeitssatz von Picard-Lindelöf, weil es in diesem Fall vorkommen kann, dass die Lösung nur für Zeiten t in der Nähe des Startwertes t_0 existiert. Für Details verweisen wir auf Bücher über gewöhnliche Differentialgleichungen, zum Beispiel [Heu09, Abschnitte 11, 12 und 60] oder [PW10, Kapitel 2].

1.4 Lineare Differentialgleichungen und die Matrix-Exponentialfunktion

Im Folgenden sei $\mathbb{K}^n$ stets mit der euklidischen Norm ausgestattet. Eine Matrix $A \in \mathbb{K}^{n \times m}$ identifizieren wir stets mit der zugehörigen linearen Abbildung

$$\begin{aligned}\mathbb{K}^m &\rightarrow \mathbb{K}^n, \\ x &\mapsto Ax.\end{aligned}$$

Proposition 1.4.1 (Die Matrixnorm als Lipschitzkonstante linearer Abbildungen). *Sei $A \in \mathbb{K}^{n \times m}$. Die Abbildung A ist Lipschitz-stetig und ihre Lipschitzkonstante ist gleich der Zahl*

$$\|A\| := \max \{ \|Ax\| \mid x \in \mathbb{K}^m \text{ and } \|x\| \leq 1 \}.$$

Außerdem ist die Abbildung $\mathbb{K}^{n \times m} \rightarrow [0, \infty)$, $A \mapsto \|A\|$ eine Norm auf $\mathbb{K}^{n \times m}$.

Man nennt $\|A\|$ die von der euklidischen Norm induzierte **Matrixnorm** oder **Operatornorm** von A .

⁷D.h. jeder Punkt $x \in \Omega$ besitzt eine Umgebung U in Ω derart, dass die Einschränkung $f|_U$ Lipschitz-stetig ist.

Beweis von Proposition 1.4.1. Es bezeichne $L \geq 0$ die Lipschitz-Konstante von A ; wir müssen $L = \|A\|$ zeigen.

“ $\leq$ ” Für alle $x \in \mathbb{K}^m$ folgt aus der Definition von $\|A\|$ die Ungleichung $\|Ax\| \leq \|A\| \|x\|$. Für alle $x, y \in \mathbb{K}^m$ gilt somit $\|Ax - Ay\| \leq \|A\| \|x - y\|$, also ist tatsächlich $L \geq \|A\|$.

“ $\geq$ ” Für jedes $x \in \mathbb{K}^m$ mit Norm $\|x\| \leq 1$ gilt

$$\|Ax\| = \|Ax - A0\| \leq L \|x - 0\| \leq L.$$

Indem wir das Supremum über alle x mit $\|x\| \leq 1$ nehmen, folgt $\|A\| \leq L$.

Dass $A \mapsto \|A\|$ tatsächlich eine Norm ist, ist ein Standardresultat aus der Analysis – man kann leicht nachrechnen, dass die Axiome einer Norm erfüllt sind. $\square$

Proposition 1.4.2 (Submultiplikativität der Matrixnorm). *Seien $A \in \mathbb{K}^{m \times \ell}$ und $B \in \mathbb{K}^{n \times m}$. Dann gilt $\|BA\| \leq \|B\| \|A\|$.*

Beweis. Das ist ein Spezialfall von Aufgabe 1.2.2(a). $\square$

Wenn $A \in \mathbb{K}^{n \times n}$, $x_0 \in \mathbb{K}^n$ und $t_0 \in \mathbb{R}$ ist, dann nennt man

$$\begin{cases} \dot{x}(t) = Ax(t) & \text{für alle } t \in \mathbb{R}, \\ x(t_0) = x_0 \end{cases} \quad (1.4.1)$$

ein **lineares, homogenes Anfangswertproblem**. Weil $A : \mathbb{K}^n \rightarrow \mathbb{K}^n$ laut Proposition 1.4.1 Lipschitz-stetig ist, hat das Anfangswertproblem 1.4.1 laut des globalen Existenz- und Eindeutigkeitssatzes von Picard–Lindelöf 1.3.1 genau eine Lösung $x : \mathbb{R} \rightarrow \mathbb{K}^d$.

Diese Lösung wollen wir nun berechnen. Dazu brauchen wir die folgende Definition.

Definition 1.4.3 (Matrix-e-Funktion). Sei $A \in \mathbb{K}^{n \times n}$. Dann nennt man

$$e^A := \exp(A) := \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

die Matrixexponentialfunktion von A .

Beachten Sie, dass die Reihe in der vorangehenden Definition immer konvergiert. Die Details hierzu besprechen wir in den Übungen.

Satz 1.4.4 (Lösung linearer, homogener Anfangswertprobleme). *Seien $A \in \mathbb{K}^{n \times n}$, $x_0 \in \mathbb{K}^n$ und $t_0 \in \mathbb{R}$. Dann ist die Funktion*

$$\begin{aligned} \mathbb{R} &\rightarrow \mathbb{K}^d, \\ t &\mapsto e^{(t-t_0)A} x_0 \end{aligned}$$

stetig differenzierbar und erfüllt das Anfangswertproblem (1.4.1).

Beweis. Sei $x : \mathbb{R} \rightarrow \mathbb{K}^n$ diejenige differenzierbare Funktion, die (1.4.1) löst. Laut Proposition 1.1.4 ist x sogar stetig differenzierbar. Wir müssen also nur $x(t) = e^{(t-t_0)A}x_0$ für alle $t \in \mathbb{R}$ zeigen.

Sei dazu $J \subseteq \mathbb{R}$ ein nicht-triviales Intervall, das t_0 enthält. Wir verwenden die Picard–Lindelöf-Iterierten aus Bemerkung 1.3.2, angewendet auf die konstante Funktion $x_0 \mathbb{1}_J \in C(J; \mathbb{K}^n)$. Sei also $T : C(J; \mathbb{K}^n) \rightarrow C(J; \mathbb{K}^n)$ definiert wie im Beweis von Satz 1.3.1. Dann konvergiert $(T^k x_0 \mathbb{1}_J)$ bezüglich $\|\cdot\|_\infty$ gegen $x|_J$. Induktiv rechnet man leicht nach, dass

$$(T^k x_0 \mathbb{1}_J)(t) = \sum_{j=0}^k \frac{(t-t_0)^j}{j!} A^j x_0$$

für alle $t \in J$ und alle $t \in J$ gilt. Also ist für jedes $t \in J$

$$x(t) = x|_J(t) = \lim_{k \rightarrow \infty} (T^k x_0 \mathbb{1}_J)(t) = \sum_{j=0}^{\infty} \frac{(t-t_0)^j}{j!} A^j x_0 = e^{(t-t_0)A} x_0.$$

Weil J ein beliebiges nicht-triviales kompaktes Intervall war, dass t_0 enthält, gilt somit $x(t) = e^{(t-t_0)A}x_0$ für alle $t \in \mathbb{R}$. $\square$

Korollar 1.4.5 (Eigenschaften der Matrix-e-Funktion). *Sei $A \in \mathbb{K}^{n \times n}$.*

- (a) *Die Funktion $\mathbb{R} \rightarrow \mathbb{K}^{n \times n}$, $t \mapsto e^{tA}$ ist stetig differenzierbar und es gilt $\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}A$ für alle $t \in \mathbb{R}$.*
- (b) *Es gilt $e^{(t+s)A} = e^{tA}e^{sA}$ für alle $s, t \in \mathbb{R}$.*
- (c) *Es gilt $e^{0 \cdot A} = \text{id}$, wobei $\text{id} \in \mathbb{K}^{n \times n}$ die Einheitsmatrix bezeichnet.*
- (d) *Für jedes $t \in \mathbb{R}$ ist e^{tA} invertierbar und erfüllt $(e^{tA})^{-1} = e^{-tA}$.*
- (e) *Für jedes $\lambda \in \mathbb{K}$ und jedes $t \in \mathbb{R}$ gilt $e^{t(\lambda \text{id} + A)} = e^{t\lambda}e^{tA}$.*

Beweis. (a) Lassen Sie uns den k -ten kanonischen Einheitsvektor $e_k \in \mathbb{K}^n$ betrachten für ein $k \in \{1, \dots, n\}$. Dann ist $x : \mathbb{R} \rightarrow \mathbb{K}^n$, $t \mapsto e^{tA}e_k$ laut Satz 1.4.4 stetig differenzierbar und erfüllt

$$\begin{cases} \dot{x}(t) = Ax(t) & \text{für alle } t \in \mathbb{R}, \\ x(0) = e_k \end{cases}.$$

Somit gilt $\frac{d}{dt}e^{tA}e_k = Ae^{tA}e_k$. Weil k beliebig war, ist $\mathbb{R} \rightarrow \mathbb{K}^{n \times n}$ stetig differenzierbar und für die Ableitung erhalten wir

$$\frac{d}{dt}e^{tA} = \left(\frac{d}{dt}e^{tA}e_1 \quad \dots \quad \frac{d}{dt}e^{tA}e_n \right) = \left(Ae^{tA}e_1 \quad \dots \quad Ae^{tA}e_n \right) = Ae^{tA} \text{id} = Ae^{tA}$$

für alle $t \in \mathbb{R}$. Dass A und e^{tA} kommutieren folgt direkt aus der Definition der Matrix-e-Funktion, weil A mit jeder seiner Potenzen kommutiert.

(b) Wir betrachten wieder den k -ten kanonischen Einheitsvektor $e_k \in \mathbb{K}^n$ für ein $k \in \{1, \dots, n\}$ und wir fixieren ein $s \in \mathbb{R}$.

Laut Satz 1.4.4 gilt $\frac{d}{dt}e^{tA}e_k = Ae^{tA}e_k$ für alle $t \in \mathbb{R}$ und somit folgt aus der Kettenregel $\frac{d}{dt}e^{(t+s)A}e_k = Ae^{(t+s)A}e_k$ für alle $t \in \mathbb{R}$, d.h. $x : \mathbb{R} \rightarrow \mathbb{K}^d$, $t \mapsto e^{(t+s)A}e_k$ löst das Anfangswertproblem

$$\begin{cases} \dot{x}(t) = Ax(t) & \text{für alle } t \in \mathbb{R}, \\ x(0) = e^{sA}e_k. \end{cases}$$

Zugleich besagt Satz 1.4.4 aber, dass die einzige Lösung dieses Anfangswertproblems durch $x(t) = e^{tA}e^{sA}e_k$ für alle $t \in \mathbb{R}$ gegeben ist. Also gilt für jedes $t \in \mathbb{R}$ die Gleichung $e^{(t+s)A}e_k = e^{tA}e^{sA}e_k$, das heißt die k -te Spalte der Matrizen $e^{(t+s)A}$ und $e^{tA}e^{sA}$ stimmt überein. Weil $k \in \{1, \dots, n\}$ beliebig war, folgt die Gleichheit der Matrizen.

(c) Das folgt direkt aus der Definition der Matrix-e-Funktion.

(d) Das folgt aus (c) und (b) für $s = -t$.

(e) Die Produktregel⁸ und (a) liefern

$$\frac{d}{dt}(e^{t\lambda}e^{tA}) = \lambda e^{t\lambda}e^{tA} + e^{t\lambda}Ae^{tA} = (\lambda \text{id} + A)e^{t\lambda}e^{tA} = \frac{d}{dt}e^{t(\lambda \text{id} + A)}$$

für alle $t \in \mathbb{R}$. Außerdem sind $e^{t\lambda}e^{tA}$ und $e^{t(\lambda \text{id} + A)}$ laut (c) für $t = 0$ beide gleich id . Also folgt die Behauptung aus dem Hauptsatz der Differential- und Integralrechnung. $\square$

Beispiele 1.4.6 (Einfache Beispiele für die Matrix-e-Funktion). (a) Es gilt

$$\exp\left(t \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

für alle $t \in \mathbb{R}$.

(b) Allgemeiner gilt für jedes $\lambda \in \mathbb{C}$

$$\exp\left(t \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}\right) = e^{t\lambda} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{t\lambda} & te^{t\lambda} \\ 0 & e^{t\lambda} \end{pmatrix}$$

für alle $t \in \mathbb{R}$.

(c) Es gilt

$$\exp\left(t \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$$

für alle $t \in \mathbb{R}$.

⁸Warum gilt sie in diesem Fall?

Beweis. (a) Das folgt sofort aus der Definition der Matrix-e-Funktion, wenn man verwendet, dass

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = 0$$

ist.

(b) Das ist eine direkte Konsequenz aus Teil (a) und Korollar 1.4.5(e).

(c) Das kann man leicht nachrechnen, indem man beobachtet, dass alle ungeraden Potenzen von

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

gleich der Matrix selbst sind und all gerade Potenzen gleich der Einheitsmatrix. $\square$

Wir wollen nun noch allgemeine Verfahren besprechen um die Exponentialfunktion einer Matrix zu berechnen. Wir beginnen mit dem Fall diagonalisierbarer Matrizen:

Satz 1.4.7 (Matrix-e-Funktion für diagonalisierbare Matrizen). *Sei $A \in \mathbb{K}^{n \times n}$ diagonalisierbar, d.h. es gebe eine invertierbare Matrix $T \in \mathbb{K}^{n \times n}$ und eine Diagonalmatrix $D \in \mathbb{K}^{n \times n}$ mit Diagonaleinträgen $\lambda_1, \dots, \lambda_n$ derart, dass $A = TDT^{-1}$ ist. Dann gilt*

$$e^A = Te^DT^{-1} = T \begin{pmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{pmatrix} T^{-1}.$$

Beweis von Satz 1.4.7. Die Definition von e^A und e^D als Exponentialreihe liefert

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \sum_{k=0}^{\infty} \frac{(TDT^{-1})^k}{k!} = \sum_{k=0}^{\infty} \frac{TD^kT^{-1}}{k!} = T \sum_{k=0}^{\infty} \frac{D^k}{k!} T^{-1} = Te^DT^{-1};$$

für die vorletzte Gleichheit haben wir verwendet, dass die Matrixmultiplikation stetig ist, damit wir die Faktoren T und T^{-1} aus der unendlichen Reihe ziehen dürfen.

Dass e^D die behauptete Form hat, kann man leicht nachrechnen, da man Diagonalmatrizen multiplizieren kann, indem man ihre jeweiligen Diagonaleinträge multipliziert. $\square$

Beachten Sie, dass man mit Satz 1.4.7 auch e^{tA} für jedes $t \in \mathbb{R}$ ausrechnen kann, wenn A diagonalisierbar ist; denn aus $A = TDT^{-1}$ folgt $tA = T(tD)T^{-1}$ und tD ist ebenfalls eine Diagonalmatrix. Denken Sie auch daran, dass es Matrizen $A \in \mathbb{R}^{n \times n}$ gibt, die über $\mathbb{R}$ nicht diagonalisierbar sind, über $\mathbb{C}$ (d.h., wenn man A als Element von $\mathbb{C}^{n \times n}$ auffasst) hingegen schon. Hier ist ein Beispiel ein solchen Falls:

Beispiel 1.4.8 (e-Funktion einer Matrix mit den Eigenwerten). Betrachten wir die Matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2 \times 2}.$$

Sie hat keine reellen Eigenwerte, aber über $\mathbb{C}$ hat sie die beiden Eigenwerte $-i$ und i und ist somit diagonalisierbar.⁹ Eigenvektoren zu den Eigenwerten iu und i sind zum Beispiel $\begin{pmatrix} 1 \\ i \end{pmatrix}$ und $\begin{pmatrix} 1 \\ -i \end{pmatrix}$. Wir können also

$$D := \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \quad \text{und} \quad T := \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$$

definieren. Somit gilt für alle $t \in \mathbb{R}$

$$\begin{aligned} e^{tA} &= T e^{tD} T^{-1} = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} e^{-it} & 0 \\ 0 & e^{it} \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} e^{-it} & -ie^{-it} \\ e^{it} & ie^{it} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} e^{-it} + e^{it} & i(-e^{-it} + e^{it}) \\ i(e^{-it} - e^{it}) & e^{-it} + e^{it} \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}. \end{aligned}$$

Alternativ kann man diese Formel auch mit einer ähnlichen Rechnung wie in Beispiel 1.4.6(c) zeigen, indem man verwendet, dass $A^4 = \text{id}$ gilt.

Genauso wie man die e-Funktion einer Diagonalmatrix sehr leicht ausrechnen kann, funktioniert dies auch für Block-Diagonalmatrizen. Eine Matrix $A \in \mathbb{K}^{n \times n}$ heißt **Block-Diagonalmatrix**, wenn man sie in der Form

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix} \tag{1.4.2}$$

schreiben kann, wobei $A_1 \in \mathbb{K}^{n_1 \times n_1}, \dots, A_m \in \mathbb{K}^{n_m \times n_m}$ quadratische Matrizen sind.^{10,11}

Wenn man die Exponentialfunktion einer Block-Diagonalmatrix ausrechnen möchte, dann kann man dies einfach blockweise tun:

Proposition 1.4.9 (Matrix-e-Funktion von Blockmatrizen). *Für eine Block-Diagonalmatrix A wie in (1.4.2) gilt*

$$e^A = \begin{pmatrix} e^{A_1} & & 0 \\ & \ddots & \\ 0 & & e^{A_m} \end{pmatrix},$$

⁹Warum?

¹⁰Hierbei gilt natürlich $n_1 + \dots + n_m = n$.

¹¹Rein formal betrachtet kann man jede Matrix als Block-Diagonalmatrix auffassen, die aus nur einem Block besteht (mit $n_1 = n$). Das ist aber natürlich nicht besonders hilfreich.

wobei die Ausdrücke $e^{A_1}, \dots, e^{A_m}$ ebenfalls im Sinne der Matrix-Exponentialfunktion zu verstehen sind.

Beweis. Man kann leicht nachrechnen, dass

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}^k = \begin{pmatrix} (A_1)^k & & 0 \\ & \ddots & \\ 0 & & (A_m)^k \end{pmatrix}$$

für alle $k \in \mathbb{N}_0$ gilt. Somit folgt die Behauptung aus der Definition der Matrix-Exponentialfunktion. $\square$

Auch für nicht-diagonalisierbare Matrizen kann man die Matrix-e-Funktion bestimmen, wenn man die Jordan-Normalform der Matrix kennt. Wegen Proposition 1.4.9 genügt es dazu, wenn wir in der Lage sind, die Matrix-e-Funktion eines Jordan-Blocks zu bestimmen.

Dabei ist jede Zahl λ_k ($k = 1, \dots, m$) ein Eigenwert von A , und es können auch mehrere der Zahlen $\lambda_1, \dots, \lambda_m$ gleich sein (d.h. in der Jordan-Normalform von A können mehrere Jordan-Blöcke mit demselben Eigenwert vorkommen).

Satz 1.4.10. Sei $J \in \mathbb{K}^{n \times n}$ ein Jordanblock, das heißt es sei

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}$$

für ein $\lambda \in \mathbb{K}$. Dann gilt für jedes $t \in \mathbb{R}$

$$e^{tJ} = e^{t\lambda} \begin{pmatrix} \frac{t^0}{0!} & \frac{t^1}{1!} & \frac{t^2}{2!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ 0 & \frac{t^0}{0!} & \frac{t^1}{1!} & \cdots & \vdots \\ \vdots & & \ddots & \ddots & \frac{t^2}{2!} \\ \vdots & & & \frac{t^0}{0!} & \frac{t^1}{1!} \\ 0 & \cdots & \cdots & 0 & \frac{t^0}{0!} \end{pmatrix}.$$

Beweis. Sei $t \in \mathbb{R}$ fest. Laut Korollar 1.4.5(e) gilt

$$e^{tJ} = e^{t\lambda \text{id} + t(J - \lambda \text{id})} = e^{t\lambda} e^{t(J - \lambda \text{id})}.$$

Deshalb können wir ab sofort ohne Einschränkung $\lambda = 0$ annehmen.

Somit gilt $J^k = 0$ für alle $k \geq n$ und im Fall $k \in \{0, \dots, n-1\}$ besteht die Matrix J^k nur aus Nullen, mit Ausnahme der k -ten Superdiagonalen, die nur aus Einsen besteht. Mit Hilfe der Exponentialreihe

$$e^{tJ} = \sum_{k=0}^{\infty} \frac{(tJ)^k}{k!} = \sum_{k=0}^{n-1} \frac{t^k J^k}{k!}$$

erhalten wir deshalb die behauptete Formel. $\square$

Übrigens sind die Beispiele 1.4.6(a) und (b) Spezialfälle von Satz 1.4.10 für den Fall $n = 2$.

1.5 Inhomogenitäten und Faltung

Das Ziel dieses und des nächsten Abschnitts ist es, Anfangswertprobleme von der Form

$$\begin{cases} \dot{x}(t) = Ax(t) + g(t), \\ x(0) = x_0 \end{cases}$$

zu besprechen. Die Differentialgleichung in diesem Anfangswertproblem nennt man eine **inhomogene linear Differentialgleichung**. Wir weichen hier in zweierlei Hinsicht von der Situation ab, die wir zuvor betrachtet hatten: (i) Da die rechte Seite $g(t)$ enthält, kann sie nun explizit von t abhängen. (ii) Wir wollen auch Funktionen g zulassen, die nicht stetig sind. Das wird später sinnvoll sein, wenn wir solche eine **Inhomogenität** stückweise aus mehreren Funktionen zusammensetzen wollen.

In diesem Abschnitt sehen wir uns aber zunächst den Fall stetiger Inhomogenitäten an. In diesem Fall lässt sich das Anfangswertproblem folgendermaßen lösen:

Proposition 1.5.1 (Differentialgleichung mit stetiger Inhomogenität). *Sei $A \in \mathbb{K}^{n \times n}$, $x_0 \in \mathbb{K}^n$, und $g : \mathbb{R} \rightarrow \mathbb{K}^n$ stetig. Dann ist die Funktion $x : \mathbb{R} \rightarrow \mathbb{K}^n$, die durch*

$$x(t) := e^{tA}x_0 + \int_0^t e^{(t-s)A}g(s) \, ds = e^{tA}x_0 + \int_0^t e^{sA}g(t-s) \, ds$$

für alle $t \in \mathbb{R}$ gegeben ist, stetig differenzierbar und löst das Anfangswertproblem

$$\begin{cases} \dot{x}(t) = Ax(t) + g(t) & \text{für alle } t \in \mathbb{R}. \\ x(0) = x_0. \end{cases}$$

Beweis. Das kann man direkt durch Nachrechnen zeigen, $\square$

Die Lösungsformel in der vorangehenden Proposition wird manchmal als **Variation der Konstanten-Formel** bezeichnet. Wir bezeichnen Sie im folgenden als **Faltungsformel**, da sie die Lösung durch die Faltung der Matrix-e-Funktion mit der Inhomogenität ausdrückt.

Bemerkung 1.5.2. Eine Intuition dafür, warum die Lösungsformel dieses Faltungsgestalt hat, kann man am einfachsten bekommen, indem man das zeitdiskrete Analogon einer Differentialgleichung betrachtet. Wir besprechen das in der Vorlesung im Detail.

Bevor wir fortfahren, zeigen wir eine nette Anwendung von Proposition 1.5.1: Die Gleichung $e^{A+B} = e^A e^B$ ist für Matrizen A und B im Allgemeinen falsch.¹² Man kann aber trotzdem eine Formel für e^{A+B} angeben, die sogenannte **Dyson–Phillips-Reihe**.¹³

Satz 1.5.3 (Dyson–Phillips-Reihe). *Seien $A, B \in \mathbb{K}^{n \times n}$. Für $t \in \mathbb{R}$ und $k \in \mathbb{N}_0$ definieren wir $V_k(t) \in \mathbb{K}^{n \times n}$ rekursiv durch*

$$\begin{aligned} V_0(t) &:= e^{tA}, \\ V_{k+1}(t) &:= \int_0^t e^{sA} B V_k(t-s) \, ds. \end{aligned}$$

Dann konvergiert für jedes $t \in \mathbb{R}$ die Reihe $\sum_{k=0}^{\infty} V_k(t)$ und ist gleich $e^{t(A+B)}$.

Beweis. Fixiere $x_0 \in \mathbb{K}^n$. Dann ist $x : \mathbb{R} \rightarrow \mathbb{K}^n$, $t \mapsto e^{t(A+B)} x_0$ die Lösung des Anfangswertproblems

$$\begin{cases} \dot{x}(t) = Ax(t) + Bx(t), \\ x(0) = x_0. \end{cases}$$

Indem wir Proposition 1.5.1 $g(t) := Bx(t)$ für alle t wählen, sehen wir, dass

$$x(t) = e^{tA} x_0 + \int_0^t e^{sA} Bx(t-s) \, ds$$

für alle $t \in \mathbb{R}$ gilt.

Sei jetzt $J \subseteq \mathbb{R}$ ein nicht-triviales kompaktes Intervall; dass 0 enthält. Dann ist die Einschränkung von x auf J also ein Fixpunkt der Abbildung $T : C(J; \mathbb{K}^n) \rightarrow C(J; \mathbb{K}^n)$, die durch

$$(Ty)(t) = e^{tA} x_0 + \int_0^t e^{sA} By(t-s) \, ds$$

für alle $y \in C(J; \mathbb{K}^n)$ und alle $t \in J$ gegeben ist.

Nun können wir wie im Beweis des Satzes 1.3.1 von Picard–Lindelöf zeigen, dass T Lipschitz-stetig ist und die Lipschitz-Konstanten der Iterierten von T summierbar sind. Also konvergiert für jedes $y \in C(J; \mathbb{K}^n)$ die Folge $(T^k y)$ gegen $x|_J$. Indem man $T^k y$ für $y(t) := e^{tA} x_0$ für alle $k \in \mathbb{N}_0$ explizit berechnet erhält man die Behauptung. $\square$

¹²Es sei denn, A und B kommutieren.

¹³Diese Reihe ist besonders in unendlicher Dimension von Relevanz. Obwohl wir uns in dieser Vorlesung nur mit dem endliche-dimensionalen Fall beschäftigen, zeigen wir sie hier, da sie eine schöne Anwendung von Proposition 1.5.1 ist und außerdem nochmals eine Anwendung des Banachschen Fixpunktsatz liefert.

Zuletzt betrachten wir noch messbare anstelle von stetigen Inhomogenitäten. Für ein Intervall $J \subseteq \mathbb{R}$ bezeichnen wir mit $L^1_{\text{loc}}(\mathbb{R}; \mathbb{K}^n)$ die Menge aller messbaren Funktionen von J nach $\mathbb{K}^n$,¹⁴ derart, dass auf jeder kompakten Teilmenge K von J jede Komponentenfunktion in $L^1(K; \mathbb{K})$ liegt. Wir benötigen das folgende Resultat:

Proposition 1.5.4 (Fast überall-Variante des Hauptsatzes). *Sei $J \subseteq \mathbb{R}$ ein nicht-triviales Intervall und $t_0 \in J$ und sei $g \in L^1_{\text{loc}}(J; \mathbb{K}^n)$. Sei $h : J \rightarrow \mathbb{K}^n$,*

$$h(t) := \int_{t_0}^t g(s) \, ds$$

für alle $t \in J$. Dann ist h stetig und fast überall differenzierbar und es gilt $\dot{h}(t) = g(t)$ für fast alle $t \in J$.

Wir verzichten an dieser Stelle auf den Beweis.

Ähnlich wie in Proposition 1.5.1 kann man nun zeigen:

Proposition 1.5.5 (Differentialgleichung mit messbarer Inhomogenität). *Sei $A \in \mathbb{K}^{n \times n}$ und $x_0 \in \mathbb{K}^n$, Sei $J \subseteq \mathbb{R}$ ein nicht-triviales Intervall, das 0 enthält, und sei $g \in L^1_{\text{loc}}(J; \mathbb{K}^n)$. Dann ist die Funktion $x : J \rightarrow \mathbb{K}^n$, die durch*

$$x(t) := e^{tA}x_0 + \int_0^t e^{(t-s)A}g(s) \, ds = e^{tA}x_0 + \int_0^t e^{sA}g(t-s) \, ds$$

für alle $t \in J$ gegeben ist, absolut stetig und fast überall differenzierbar und es gilt

$$\begin{cases} \dot{x}(t) = Ax(t) + g(t) & \text{für fast alle } t \in J, \\ x(0) = x_0. \end{cases}$$

Die Lösung x in der oben stehenden Proposition bezeichnen wir als **Caratheodory-Lösung** der Differentialgleichung (statt als “klassische” Lösung, da sie nur absolut stetig und fast überall differenzierbar sein muss).

¹⁴Wobei wir Funktionen identifizieren, die fast überall gleich sind.

Kapitel 2

Eingänge und Ausgänge

2.1 Systeme mit Ein- und Ausgängen

Grundbegriffe

Von nun an werden wir lineare Differentialgleichungen mit Ein- und Ausgängen studieren. Es sei dazu $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$ und $D \in \mathbb{K}^{p \times m}$. Die zugehörigen linearen Abbildungen bilden also folgendermaßen ab:

$$\begin{array}{ccccccc} \mathbb{K}^m & \xrightarrow{B} & \mathbb{K}^n & \xrightarrow{A} & \mathbb{K}^n & \xrightarrow{C} & \mathbb{K}^p \\ & & & & & & \\ \mathbb{K}^m & \xrightarrow{\quad D \quad} & \mathbb{K}^p & & & & \end{array}$$

Wir wollen die Gleichungen

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

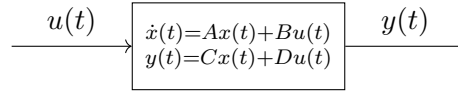
studieren. Man bezeichnet (Σ) als ein **System**. Dabei ist u eine Funktion von einem reellen Intervall nach $\mathbb{K}^m$ und Bu somit eine Funktion nach $\mathbb{K}^n$. Wir nennen u die **Eingangsfunktion** oder **Inputfunktion** von (Σ) . Den Vektor $x(t)$ nennt man den **Zustand** oder den **Zustandsvektor** des Systems (Σ) zum Zeitpunkt t . Die erste Zeile ist also eine inhomogene lineare Differentialgleichung auf $\mathbb{K}^n$ mit Inhomogenität Bu . Die zweite Zeile berechnet aus der Lösung x der Differentialgleichung und der Eingangsfunktion u zu jedem Zeitpunkt t einen Vektor $y(t) \in \mathbb{K}^p$. Die Funktion y , die von einem reellen Intervall nach $\mathbb{K}^p$ abbildet, nennen wir die **Ausgangsfunktion** oder **Outputfunktion** von (Σ) .

Man interpretiert (Σ) üblicherweise folgendermaßen: Das Eingangssignal kann von uns gewählt werden – damit **steuern** oder **kontrollieren** wir das System (Σ) , weshalb man auch von **Steuerungs-** oder **Kontrolltheorie** spricht. Aus der Differentialgleichung bestimmt sich dann die Lösung x (beachten Sie aber, dass man dazu noch einen Anfangswert für x festlegen muss!), und aus der zweiten Zeile ergibt sich schließlich der Wert von y zu jedem Zeitpunkt.

Wir verwenden die folgenden Begriffe für die Matrizen und die Räume $\mathbb{K}^m, \mathbb{K}^n, \mathbb{K}^p$:

- Der Raum $\mathbb{K}^m$ heißt **Eingangsraum** oder **Inputraum**.
- Der Raum $\mathbb{K}^n$ heißt **Zustandsraum** (englisch **state space**).
- Der Raum $\mathbb{K}^p$ heißt **Ausgangsraum** oder **Outputraum**.
- Die Matrix B heißt **Eingangsmatrix** oder **Inputmatrix**.
- Die Matrix A heißt **Systemmatrix**.
- Die Matrix C heißt **Ausgangsmatrix** oder **Outputmatrix**.
- Die Matrix D heißt **Durchgriffsmatrix** oder **Feedthroughmatrix**.

Das System (Σ) wird häufig mit Hilfe des folgenden Diagramms veranschaulicht:

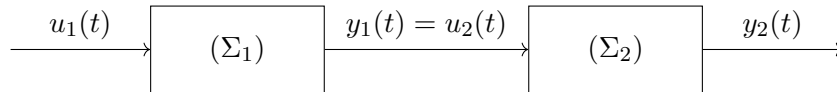


Oder, wenn man das System im Diagramm mit (Σ) abkürzen will:



Verknüpfung mehrere Systeme

Man kann, was für Modellierungszwecke äußerst nützlich ist, auch mehrere Systeme miteinander verknüpfen. Hat man zum Beispiel zwei Systeme (Σ_1) und (Σ_2) mit den entsprechenden Matrizen $A_k \in \mathbb{K}^{n_k \times n_k}$, $B \in \mathbb{K}^{n_k \times m_k}$, $C \in \mathbb{K}^{p_k \times n_k}$ und $D \in \mathbb{K}^{p_k \times m_k}$ (für $k \in \{1, 2\}$) und gilt $m_2 = p_1$, so kann man die beiden System wie in folgendem Diagramm hintereinander schalten:



Die Verknüpfung der beiden System entsteht hier also, indem man den Eingang $u_2(t)$ von (Σ_2) gleich dem Ausgang $y_1(t)$ setzt – und dazu war die Annahme $m_2 = p_1$ nötig. Das Gesamtsystem, dass wir so erhalten haben, hat also den Eingang u_1 , den Ausgang y_2 , und den Zustandsvektor $(x_1(t), x_2(t))$. Als Gleichung des verknüpften Systems erhält man

$$\begin{cases} \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ B_2 C_1 & A_2 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 D_1 \end{pmatrix} u_1(t), \\ y_2(t) = \begin{pmatrix} D_2 C_1 & C_2 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + D_2 D_1 u_1(t) \end{cases}$$

2.2 Steuerbarkeit

In diesem Abschnitt sehen wir uns an, wie man den Zustand eines Systems mit Hilfe der Eingangsfunktion beeinflussen kann. Der Ausgang ist dafür nicht relevant, das heißt wir betrachten nur eine Systemmatrix $A \in \mathbb{K}^{n \times n}$ und eine Eingangsmatrix $B \in \mathbb{K}^{n \times m}$ sowie einen Startvektor (auf den waren wir im vorangehenden Abschnitt noch gar nicht richtig eingegangen) $x_0 \in \mathbb{K}^n$ und betrachten das Anfangswertproblem

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0. \quad (2.2.1)$$

Wir werden im Folgenden in erster Linie an der Entwicklung in positiver Zeitrichtung – d.h. für $t \geq 0$ interessiert sein. Die Eingangsfunktion u nehmen wir als L^1 -Funktion an, die auf einem Intervall mit linkem Randpunkt 0 definiert ist. Für die (Caratheodory-)Lösung des Anfangswertproblems kennen wir aus Proposition 1.5.5, der wir nun folgende Bezeichnung geben:

Definition 2.2.1 (Lösungsoperator für Systeme mit Eingang). Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Sei $x_0 \in \mathbb{K}^n$ und $u \in L^1(J; \mathbb{K}^m)$ für ein Intervall J , das 0 und einen weiteren Punkt rechts von 0 enthält. Dann definieren wir

$$\Phi_u^t(x_0) := e^{tA}x_0 + \int_0^t e^{(t-s)A}Bu(s) \, ds \in \mathbb{K}^n$$

für jedes $0 \leq t \in J$.¹

Das heißt, $t \mapsto \Phi_u^t(x_0)$ löst das Anfangswertproblem (2.2.1) für $0 \leq t \in J$. Man beachte, dass der Wert $\Phi_u^t(x_0)$ nicht davon abhängt, wie sich u rechts von t verhält oder links von 0 verhält. Insbesondere hängt $\Phi_u^t(x_0)$ auch nicht davon ab, ob wie weit rechts von t und wie weit links von 0 die Funktion u noch definiert ist – d.h., um über $\Phi_u^t(x_0)$ zu sprechen können wir uns ab sofort vorstellen, dass u nur auf $[0, t]$ definiert ist, selbst wenn der Definitionsbereich von u in Wirklichkeit größer ist.

Die folgende Eigenschaft des Lösungsoperators Φ spiegelt wieder, was passiert, wenn man die Differentialgleichung zu einem Zeitpunkt s “neu startet”:

Proposition 2.2.2 (Kozykeleigenschaft). Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Seien $s, t \geq 0$ und es sei $u \in L^1([0, s+t]; \mathbb{K}^m)$. Dann gilt

$$\Phi_u^{t+s} = \Phi_{u(\cdot+s)}^t \circ \Phi_u^s.$$

Beweis. Durch Nachrechnen. □

Beachten Sie, dass die Formel in Proposition 2.2.2 für die Eingangsfunktion $u = 0$ sich zu $\Phi_0^{t+s} = \Phi_0^t \circ \Phi_0^s$ vereinfacht – das ist gerade die Funktionalgleichung $e^{(t+s)A} = e^{tA}e^{sA}$ der Matrixexponentialfunktion.

Man kann die Kozykeleigenschaft stattdessen auch mit Hilfe der Konkatenation von Eingangsfunktionen beschreiben:

¹Beachten Sie, dass wir A und B in der Notation für Φ unterdrücken.

Definition 2.2.3 (Konkatenation von Funktionen). Sei $s, t \geq 0$ und $u \in L^1([0, s]; \mathbb{K}^n)$ sowie $v \in L^1([0, t]; \mathbb{K}^n)$. Dann definiert man $v \diamond u \in L^1([0, s+t]; \mathbb{K}^n)$ durch

$$(v \diamond u)(r) := \begin{cases} u(r) & \text{für } r \in [0, s), \\ v(r-s) & \text{für } r \in [s, s+t]. \end{cases}$$

Korollar 2.2.4 (Kozykeleigenschaft via Konkatenation). Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Seien $s, t \geq 0$ und $u \in L^1([0, s]; \mathbb{K}^m)$ sowie $v \in L^1([0, t]; \mathbb{K}^m)$. Dann gilt

$$\Phi_{v \diamond u}^{t+s} = \Phi_v^t \circ \Phi_u^s.$$

Nun können wir wie angekündigt diskutieren, wie wir ein System durch Wahl der Eingangsfunktion beeinflussen können.

Definition 2.2.5 (Erreichbarkeit und Kontrollierbarkeit). Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Sei $t > 0$.

- (a) Seien $x_0, x_1 \in \mathbb{K}^n$. Wir sagen, x_1 ist von x_0 aus **in der Zeit t erreichbar**, wenn es ein $u \in L^1([0, t]; \mathbb{K}^m)$ gibt mit $\Phi_u^t(x_0) = x_1$. Diese Eigenschaft notieren wir mit $x_0 \overset{t}{\rightsquigarrow} x_1$.
- (b) Das Paar (A, B) heißt **steuerbar in der Zeit t** oder **kontrollierbar in der Zeit t** , falls für alle $x_0, x_1 \in \mathbb{K}^n$ gilt: $x_0 \overset{t}{\rightsquigarrow} x_1$.

Da wir uns nur lineare Systeme ansehen, gilt folgende nützliche Eigenschaft:

Proposition 2.2.6 (Der Anfangswert ist nicht relevant für Steuerbarkeit). Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Sei $t > 0$. Dann sind äquivalent:

- (i) Das Paar (A, B) ist kontrollierbar in Zeit t .
- (ii) Für jedes $x_1 \in \mathbb{K}^n$ gilt: $0 \overset{t}{\rightsquigarrow} x_1$.
- (iii) Die lineare Abbildung

$$\begin{aligned} L^1([0, t]; \mathbb{K}^m) &\rightarrow \mathbb{K}^n, \\ u &\mapsto \Phi_u^t(0) \end{aligned}$$

is surjektiv.

Beweis. “(i) $\Rightarrow$ (ii)” Diese Implikation ist klar.

“(ii) $\Rightarrow$ (i)” Seien $x_0, x_1 \in \mathbb{K}^n$. Wegen (ii) gilt $0 \overset{t}{\rightsquigarrow} x_1 - e^{tA}x_0$. Also gibt es ein $u \in L^1([0, t]; \mathbb{K}^m)$ mit $\Phi_u^t(0) = x_1 - e^{tA}x_0$. Diese Gleichung ist laut Definition von Φ äquivalent zu $\Phi_u^t(x_0) = x_1$.

“(ii) $\Leftrightarrow$ (iii)” Diese Äquivalenz ist klar. □

Bevor wir ein Kriterium für Kontrollierbarkeit beweisen, dass sich in konkreten Fällen leicht nachrechnen lässt, ist es sinnvoll durch einige einfache Beispiele eine besseren Intuition zu bekommen:

Beispiele 2.2.7 (Einfache Beispiele für (nicht-)Steuerbarkeit). Sei $t > 0$.

- (a) Das Paar $(A, 0)$ ist nicht kontrollierbar in der Zeit t (egal, wie wir A wählen).
- (b) Sei $m = n$. Das Paar (A, id) ist immer kontrollierbar in der Zeit t . Man kann sich nämlich überlegen, dass die Matrix $M := \int_0^t e^{(t-s)A} ds = \int_0^t e^{sA} ds$ für $t > 0$ immer invertierbar ist.² Für gegebenes $x_1 \in \mathbb{K}^n$ erfüllt somit die konstante Eingangsfunktion $u(t) := M^{-1}x_1$ die Bedingung $\Phi_u^t(0) = x_1$.
- (c) Für die Wahl

$$A := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \text{id} \quad \text{und} \quad B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

ist das Paar (A, B) nicht kontrollierbar in Zeit t , weil die Wahl der Eingangsfunktion u keinen Einfluss auf die Entwicklung der zweiten Komponente $x_2(t)$ des Zustandsvektors $x(t)$ hat.

Satz 2.2.8 (Bild der Erreichbarkeitsmatrix). *Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Sei $t > 0$. Dann ist das Bild der Matrix*

$$R(A, B) := [B, AB, A^2B, \dots, A^{n-1}B] \in \mathbb{K}^{n \times nm}$$

gleich dem Bild der linearen Abbildung

$$L^1([0, t]; \mathbb{K}^m) \rightarrow \mathbb{K}^n, \\ u \mapsto \Phi_u^t(0).$$

Insbesondere hängt letzteres nicht von t ab.

Die Matrix $R(A, B)$ in oben stehendem Satz bezeichnet man als **Erreichbarkeitsmatrix**³ des Systems.

Beweis von Satz 2.2.8. Für die Dauer dieses Beweises bezeichnen das Bild der Erreichbarkeitsmatrix mit R und das Bild von $u \mapsto \Phi_u^t(0)$ mit W . Beides sind Untervektorräume von $\mathbb{K}^n$ und somit abgeschlossen in $\mathbb{K}^n$.

“ $R \supseteq W$ ” Wir jedes Polynom $q \in \mathbb{K}[X]$ und jedes $u_0 \in \mathbb{K}^m$ gilt wegen des Satzes von Cayley–Hamilton $q(A)Bu_0 \in R$ und somit $e^{tA}Bu_0 = \lim_{k \rightarrow \infty} \sum_{j=0}^k \frac{A^j}{j!} Bu_0 \in R$. Für jedes stetige $u : [0, 1] \rightarrow \mathbb{K}^m$ ist der Vektor $\Phi_0^t(u)$ gleich dem Riemannintegral $\int_0^t e^{tA}Bu(t-s)ds$, also ein Limes von Linearkombinationen von Vektoren

²Warum ist das so?

³Oder auch als **Steuerbarkeitsmatrix** oder **Kontrollierbarkeitsmatrix**.

$e^{t_1 A} u_1, \dots, e^{t_\ell A} B u_\ell$, und somit in R . Weil $C([0, 1]; \mathbb{K}^n)$ dicht ist in $L^1([0, 1]; \mathbb{K}^n)$ und die Abbildung $u \mapsto \Phi_u^t(0)$ stetig ist, folgt $W \subseteq R$.

“ $R \subseteq W$ ” Offensichtlich liegt das Bild von B in W . Als nächstes beobachten wir, dass für jedes $s \in [0, t)$ auch das Bild von $e^{sA} B$ in W liegt. Es folgt nämlich für jedes solche s und für jedes $u_0 \in \mathbb{K}^m$ wegen der Stetigkeit der Exponentialfunktion

$$\begin{aligned} e^{sA} B u_0 &= \lim_{\delta \downarrow 0} \frac{1}{\delta} \int_s^{s+\delta} e^{rA} B u_0 \, dr \\ &= \lim_{\delta \downarrow 0} \frac{1}{\delta} \int_0^t e^{rA} B u_0 \mathbb{1}_{[s, s+\delta]}(r) \, dr = \lim_{\delta \downarrow 0} \frac{1}{\delta} \Phi_{u_\delta}^t(0) \end{aligned}$$

für $u_\delta := u_0 \mathbb{1}_{[t-s-\delta, t-s]}$.

Damit liegt für jedes $u_0 \in \mathbb{K}^m$ aber auch $\left(\frac{d}{ds}\right)^k e^{sA} B u_0$ in W für jedes $s \in [0, \infty)$ und jedes $k \in \mathbb{N}_0$. Für $s = 0$ erhalten wir insbesondere $A^k B u_0 \in \mathbb{K}$.

Weil R die Summer der Bilder von $A^0 B, \dots, A^{n-1} B$ ist, folgt wie behauptet $R \subseteq W$. $\square$

Aus Satz 2.2.8 und Proposition 2.2.6 erhält man sofort:

Korollar 2.2.9 (Kalman'sches Rangkriterium für Steuerbarkeit). *Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Dann sind äquivalent:*

- (i) *Es gibt eine Zeit $t > 0$ derart, das Paar (A, B) in der Zeit t steuerbar ist.*
- (ii) *Für jedes Zeit $t > 0$ ist das Paar (A, B) in der Zeit t steuerbar.*
- (iii) *Die Erreichbarkeitsmatrix $R(A, B) = [B, AB, A^2 B, \dots, A^{n-1} B] \in \mathbb{K}^{n \times nm}$ hat Rang vollen Rang, das heißt Rang n .*

Motiviert durch das vorangehende Korollar werden wir in Zukunft nur noch sagen, dass (A, B) **steuerbar** oder **kontrollierbar** ist, um auszudrücken, dass es für ein (äquivalent: alle) t steuerbar ist.

Lassen Sie uns die Erreichbarkeitsmatrix in den Beispielen 2.2.7 ausrechnen.

Beispiele 2.2.10 (Einfache Beispiele für (nicht-)Steuerbarkeit). (a) Es gilt $R(A, 0) = 0 \in \mathbb{K}^{n \times nm}$. Also hat $R(A, 0)$ den Rang 0.

(b) Sei $m = n$. Es gilt $R(A, \text{id}) = [\text{id}, A, \dots, A^{n-1}]$, und diese Matrix hat Rang n , weil bereits id am Anfang dieser Matrix Rang n hat.

(c) Für

$$A := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \text{id} \quad \text{und} \quad B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

gilt

$$R(A, B) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

diese Matrix hat Rang 1, also nicht vollen Rang.

Lassen Sie uns noch weiteres Beispiel betrachten:

Beispiel 2.2.11 (Ein einfaches Beispiel mit skalarem Eingang). Sei $n = 2$, $m = 1$ und

$$A := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \text{id} \quad \text{und} \quad B := \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Dann ist

$$R(A, B) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Diese Matrix hat Rang 2, also ist (A, B) kontrollierbar.

Ein weiteres Kriterium für Steuerbarkeit ist:

Satz 2.2.12 (Hautus-Test für Steuerbarkeit). *Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Dann sind äquivalent:*

- (i) *Das Paar (A, B) ist steuerbar.*
- (ii) *Für alle $\lambda \in \mathbb{C}$ ist der Rang der Matrix $[B, \lambda \text{id} - A] \in \mathbb{C}^{n \times (m+n)}$ voll, also gleich n .*
- (iii) *Für alle Eigenwerte $\lambda \in \mathbb{C}$ von A ist der Rang der Matrix $[B, \lambda \text{id} - A] \in \mathbb{C}^{n \times (m+n)}$ voll, also gleich n .*

Beweis. Wir können ohne Einschränkung $\mathbb{K} = \mathbb{C}$ annehmen.⁴

“(i) $\Rightarrow$ (ii)” Sei $\lambda \in \mathbb{C}$ und habe $(B \ \lambda \text{id} - A)$ nicht vollen Rang, das heißt, das Bild dieser Matrix ist ein echter Untervektorraum von $\mathbb{K}^n$. Dann gibt es einen Zeilenvektor $0 \neq z \in \mathbb{C}^{1 \times n}$ mit $z(B \ \lambda \text{id} - A) = 0$. Also ist $zB = 0$ und $\lambda z = zA$. Für jedes $k \in \mathbb{N}_0$ folgt somit

$$zA^k B = \lambda^k zB = 0,$$

also verschwindet z auf dem Bild der Steuerbarkeitsmatrix, das heißt, diese hat nicht vollen Rang. Wegen des Kalmanschen Rangkriteriums ist (A, B) deshalb nicht steuerbar.

“(ii) $\Rightarrow$ (i)” Wir bezeichnen mit $R \subseteq \mathbb{C}^n$ das Bild der Erreichbarkeitsmatrix $R(A, B)$ und nehmen an, dass es ein echter Unterraum von $\mathbb{K}^n$ ist. Wegen des Satzes

⁴Warum?

von Cayley–Hamilton ist R invariant unter A , also induziert A eine lineare Abbildung $A_/\text{ auf dem Quotientenraum } \mathbb{C}^n/R$. Da der Quotientenraum nicht 0 ist, besitzt $A_/\text{ einen Eigenwert } \lambda \in \mathbb{C}$. Hierzu existiert ein Linkseigenvektor, das heißt ein lineares Funktional } 0 \neq \varphi : \mathbb{C}^n/R \text{ mit der Eigenschaft } \varphi A_/\text{ = } \lambda A_/\text{.}

Wir betrachten nun die Hintereinanderausführung $\psi : \mathbb{C}^n \xrightarrow{Q} \mathbb{C}^n/R \xrightarrow{\varphi} \mathbb{C}$, wobei die erste Abbildung Q die Quotientenabbildung bezeichnet.⁵ Dann ist ψ nicht Null⁶ und es gilt $\psi(\lambda \text{id} - A) = \varphi(\lambda \text{id} - A_/)Q = 0$, also ist ψ ein Linkseigenvektor von A zum Eigenwert λ . Außerdem gilt $\psi B = \varphi QB = 0$, denn das Bild von B liegt in R und somit verschwindet Q auf dem Bild von B .

Insgesamt verschwindet ψ also auf dem Bild von $[B, \lambda \text{id} - A]$, das heißt das Bild dieser Matrix ist nicht der ganze $\mathbb{C}^n$.

“(ii) $\Leftrightarrow$ (iii)” Diese Äquivalenz folgt daraus, dass für λ aus der Resolventenmenge von A bereits die Matrix $\lambda \text{id} - A$ Rang n hat. $\square$

Korollar 2.2.13 (Steuerbarkeit ist invariant unter skalarer Verschiebung). *Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Sei $\mu \in \mathbb{K}$. Es ist (A, B) genau dann steuerbar, wenn $(A - \mu \text{id}, B)$ steuerbar ist.*

Für ein nicht-steuerbares Paar (A, B) kann man das System in einen steuerbaren Teil und einen Teil ohne Eingang zerlegen. Um das sauber aufzuschreiben, verwenden wir den folgenden Begriff:

Definition 2.2.14. Seien $A, \tilde{A} \in \mathbb{K}^{n \times n}$ und $B, \tilde{B} \in \mathbb{K}^{n \times m}$. Wir sagen, dass System mit Systemmatrix A und Eingangsmatrix B ist **ähnlich** zum System mit Systemmatrix $\tilde{A}$ und Eingangsmatrix $\tilde{B}$, wenn es eine invertierbare Matrix $T \in \mathbb{K}^{n \times n}$ gibt derart, dass $A = T\tilde{A}T^{-1}$ und $B = T\tilde{B}$ gilt.

Für zwei ähnliche Systeme sieht man leicht, dass das eine genau dann steuerbar ist, wenn das andere steuerbar ist.

Proposition 2.2.15 (Zerlegung nicht-steuerbarer Systeme). *Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Eingangsmatrix $B \in \mathbb{K}^{n \times m}$. Wenn (A, B) nicht steuerbar ist und $B \neq 0$ gilt, dann ist das System ähnlich zu einem System mit System- und Eingangsmatrix*

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} \quad \text{und} \quad \tilde{B} = \begin{pmatrix} \tilde{B}_1 \\ 0 \end{pmatrix}$$

mit folgenden Eigenschaften:

- (a) *Es ist $\tilde{A}_{11} \in \mathbb{K}^{n_1 \times n_1}$ und $\tilde{A}_{22} \in \mathbb{K}^{n_2 \times n_2}$ wobei die Dimensionen $n_1, n_2 \geq 1$ sich zu n addieren.*

Entsprechend gilt $\tilde{A}_{12} \in \mathbb{K}^{n_1 \times n_2}$ und $\tilde{B}_1 \in \mathbb{K}^{n_1 \times m}$.

⁵Weil ψ ein lineares Funktional auf $\mathbb{K}^n$ ist, können wir ψ als einen Zeilenvektor in $\mathbb{C}^{1 \times n}$ schreiben. Das Funktional ψ spielt hier im Prinzip dieselbe Rolle wie der Zeilenvektor z im vorangehenden Teil des Beweises.

⁶Wieso nicht?

(b) Das Paar $(\tilde{A}_{11}, \tilde{B}_1)$ ist steuerbar.

Beweis. Bezeichne $R \subseteq \mathbb{K}^n$ das Bild der Erreichbarkeitsmatrix $R(A, B)$. Weil (A, B) nicht steuerbar ist, ist $R \neq \mathbb{K}^n$. Also können wir einen von $\{0\}$ verschiedenen Untervektorraum $V \subseteq \mathbb{K}^n$ finden, der $R \oplus V = \mathbb{K}^n$ erfüllt.

Setze $n_1 := \dim R$ und $n_2 := \dim V$. Es ist $n_1 + n_2 = n$. Wegen $V \neq \{0\}$ gilt $n_2 \neq 0$ und weil $B \neq 0$ ist, gilt $R \neq \{0\}$ und somit $n_1 \neq 0$. Wähle eine lineare Bijektion $T : \mathbb{K}^n \rightarrow \mathbb{K}^n$, die R auf die ersten n_1 Koordinaten und V auf die verbleibenden n_2 Koordinaten schickt. Setze $\tilde{A} := TAT^{-1}$ und $\tilde{B} := TB$. Wir unterteilen diese beiden Matrizen in Blockmatrizen

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{pmatrix} \quad \text{und} \quad \tilde{B} = \begin{pmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{pmatrix}$$

entsprechend der Aufteilung von $\mathbb{K}^n$ in die ersten n_1 die die verbleibenden n_2 Koordinaten.

Weil das Bild von B in R liegt, bildet TB in die ersten n_1 Koordinaten von $\mathbb{K}^n$ ab, das heißt, für jedes $u \in \mathbb{K}^m$ sind die letzten n_2 Koordinaten von TBu gleich 0. Also ist $\tilde{B}_2 = 0$. Außerdem gilt für jeden Vektor $\tilde{x} \in \mathbb{K}^n$, der nur in den ersten n_1 Koordinaten lebt: $T^{-1}\tilde{x} \in R$ und somit $AT^{-1}\tilde{x} \in R$, da R wegen des Satz von Cayley–Hamilton invariant unter A ist; also lebt auch $\tilde{A}\tilde{x} = TAT^{-1}\tilde{x}$ nur in den ersten n_1 Koordinaten. Damit ist $\tilde{A}_{21} = 0$ gezeigt.

Man rechnet nun leicht noch, dass

$$\tilde{A}^k \tilde{B} = \begin{pmatrix} \tilde{A}_{11}^k \tilde{B}_1 \\ 0 \end{pmatrix}$$

für alle $k \in \mathbb{N}_0$ gilt, also lebt das Bild von $R(\tilde{A}, \tilde{B})$ nur in den ersten n_1 Koordinaten und ist – wieder aufgrund des Satzes von Cayley–Hamiltons – gleich dem Bild von

$$\begin{pmatrix} R(\tilde{A}_{11}, \tilde{B}_1) \\ 0 \end{pmatrix}.$$

Somit hat $R(\tilde{A}, \tilde{B}) = TR(A, B)T^{-1}$ denselben Rang wie $R(A, B)$, also n_1 . Das zeigt, dass das Bild von $R(\tilde{A}_{11}, \tilde{B}_1)$ gleich $\mathbb{K}^{n_1}$ ist, das heißt, $(\tilde{A}_{11}, \tilde{B}_1)$ ist in der Tat steuerbar. $\square$

Bemerkung 2.2.16 (Zerlegung vs. Quotientenräume). Diese Bemerkung wird in der Vorlesung diskutiert.

2.3 Beobachtbarkeit

In diesem Abschnitt betrachten wir Systeme mit Ausgängen, aber ohne Eingänge. Das heißt für $A \in \mathbb{K}^{n \times n}$ und $C \in \mathbb{K}^{p \times n}$ sowie einen Startvektor $x_0 \in \mathbb{K}^n$ betrachten wir

$$\begin{aligned} \dot{x}(t) &= Ax(t), & x(0) &= x_0, \\ y(t) &= Cx(t). \end{aligned} \tag{2.3.1}$$

Für gegebenes x_0 können wir x und y konkret berechnen: Es gilt für alle t

$$\begin{aligned}x(t) &= e^{tA}x_0, \\y(t) &= Ce^{tA}x_0.\end{aligned}$$

Wir fragen uns nun, wieviele Informationen man über x_0 erlangen kann, wenn man nur $x(t)$ beobachten kann.

Definition 2.3.1 (Beobachtbarkeit). Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Ausgangsmatrix $C \in \mathbb{K}^{p \times n}$. Sei $t > 0$. Das Paar (A, C) heißt **beobachtbar in der Zeit t** , wenn für alle $x_0, x_1 \in \mathbb{K}^n$ gilt: Falls $Ce^{sA}x_0 = Ce^{sA}x_1$ für alle $s \in [0, t]$ gilt, dann ist $x_0 = x_1$.

Proposition 2.3.2 (Beobachtbarkeit durch Vergleich mit $\mathbf{0}$). Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Ausgangsmatrix $C \in \mathbb{K}^{p \times n}$. Sei $t > 0$. Dann sind äquivalent:

- (i) Das Paar (A, C) ist beobachtbar in der Zeit t .
- (ii) Für alle $x_0 \in \mathbb{K}^n$ gilt: Falls $Ce^{sA}x_0 = 0$ für alle $s \in [0, t]$, dann ist $x_0 = 0$.
- (iii) Die Abbildung

$$\begin{aligned}\mathbb{K}^n &\rightarrow C([0, t]; \mathbb{K}^p), \\x_0 &\mapsto (s \mapsto Ce^{sA}x_0)\end{aligned}$$

ist injektiv.

Beweis. “(i) $\Leftrightarrow$ (ii)” Das folgt daraus, dass die Abbildung $x_0 \mapsto Ce^{tA}x_0$ für jedes feste t linear ist.

“(i) $\Leftrightarrow$ (iii)” Das ist einfach die Definition des Begriffs injektiv. $\square$

Satz 2.3.3 (Beobachtbarkeit ist dual zu Steuerbarkeit). Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$ und $C \in \mathbb{K}^{p \times n}$. Sei $t > 0$.

- (a) Es ist (A, B) steuerbar in der Zeit t genau dann, wenn (A^T, B^T) beobachtbar ist.
- (b) Es ist (A, C) beobachtbar in der Zeit t genau dann, wenn (A^T, C^T) steuerbar in der Zeit t ist.

Beweis. (a) Steuerbarkeit von (A, B) in der Zeit t ist laut Proposition 2.2.6(iii) äquivalent zur Surjektivität der Abbildung

$$\begin{aligned}L^1([0, t]; \mathbb{K}^m) &\rightarrow \mathbb{K}^n, \\u &\mapsto \Phi_u^t(0)\end{aligned}$$

aus dieser Proposition. Durch Zeitumkehr im Argument der Funktionen in $L^1([0, t]; \mathbb{K}^n)$ kann man leicht sehen, dass dieses Bild gleich dem Bild der Abbildung

$$\begin{aligned}\Gamma_t : L^1([0, t]; \mathbb{K}^m) &\rightarrow \mathbb{K}^n, \\ u &\mapsto \int_0^t e^{sA} B u(s) \, ds\end{aligned}$$

ist.

Zugleich ist (A^T, B^T) laut Proposition 2.3.2(iii) genau dann beobachtbar, wenn die Abbildung

$$\begin{aligned}\psi_t : \mathbb{K}^n &\rightarrow L^\infty([0, t]; \mathbb{K}^m), \\ x_0 &\mapsto (s \mapsto B^T e^{sA^T} x_0)\end{aligned}$$

injektiv ist.⁷ Wir müssen also zeigen, dass Γ_t genau dann surjektiv ist, wenn Ψ_t injektiv ist.

Aus der Funktionalanalysis weiß man, dass $L^\infty([0, t]; \mathbb{K}^m)$ dem Dualraum von $L^1([0, t]; \mathbb{K}^m)$ entspricht. Wir zeigen nun, dass $\Psi_t = (\Gamma_t)'$ gilt, das heißt, dass Ψ_t gleich dem dualen Operator von Γ_t ist. Seien dazu $u \in L^1([0, t]; \mathbb{K}^m)$ und $x_0 \in \mathbb{K}^n$. Dann rechnet man leicht nach, dass

$$\langle \Psi_t u, x_0 \rangle = \langle u, \Gamma_t x_0 \rangle$$

gilt. Also ist tatsächlich $\Psi_t = (\Gamma_t)'$.

Somit ist Ψ_t genau dann injektiv, wenn Γ_t dichtes Bild hat. Weil Γ_t in einen endlich-dimensionalen Raum abbildet, ist das wiederum äquivalent dazu, dass Γ_t surjektiv ist.

(b) Das folgt direkt aus (a) durch Vertauschen der Matrizen mit ihren Transponierten. $\square$

Somit hängt also Steuerbarkeit ebenfalls nicht von der Wahl der Zeit t ab. Wir nennen das Paar (A, C) **steuerbar**, wenn es für ein (äquivalent: alle) $t > 0$ steuerbar ist.

Korollar 2.3.4 (Kalman-Rangbedingung und Hautus-Test für Beobachtbarkeit). *Wir betrachten ein System mit Systemmatrix $A \in \mathbb{K}^{n \times n}$ und Ausgangsmatrix $C \in \mathbb{K}^{p \times n}$. Dann sind äquivalent:*

- (i) *Das Paar (A, C) ist beobachtbar.*
- (ii) *Die **Beobachtbarkeitsmatrix***

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \in \mathbb{K}^{pn \times n}$$

⁷Hier haben wir den Funktionenraum $C([0, t]; \mathbb{K}^m)$ aus der Proposition mit dem Funktionenraum $L^\infty([0, t]; \mathbb{K}^p)$ ersetzt. Warum bereitet das hier keine Probleme?

ist injektiv, hat also Rang n .

(iii) *Für jedes $\lambda \in \mathbb{C}$ ist die Matrix $\begin{bmatrix} C \\ \lambda \operatorname{id} - A \end{bmatrix} \in \mathbb{K}^{(p+n) \times n}$ injektiv, hat also Rang n .*

(iv) *Für jeden Eigenwert $\lambda \in \mathbb{C}$ von A ist die Matrix $\begin{bmatrix} C \\ \lambda \operatorname{id} - A \end{bmatrix} \in \mathbb{K}^{(p+n) \times n}$ injektiv, hat also Rang n .*

Beweis. Das folgt aus Satz 2.3.3 und aus den entsprechenden Charakterisierungen von Steuerbarkeit. $\square$

Kapitel 3

Langzeitverhalten

3.1 Langzeitverhalten der Matrixexponentialfunktion

Satz 3.1.1 (Stabilität der Matrixexponentialfunktion). *Sei $A \in \mathbb{K}^{n \times n}$. Dann sind äquivalent:*

- (i) *Es gilt $e^{tA} \rightarrow 0$ für $t \rightarrow \infty$.*
- (ii) *Die Matrixexponentialfunktion von A ist **exponentiell stabil**, das heißt es gibt Zahlen $M \geq 1$ and $\omega < 0$ derart, dass $\|e^{tA}\| \leq Me^{t\omega}$ für alle $t \in [0, \infty)$ gilt.*
- (iii) *Alle Eigenwerte von A (über $\mathbb{C}$) haben Realteil < 0 .*
- (iv) *Es gibt es positive semidefinite selbst-adjungierte Matrix $Q \in \mathbb{K}^{n \times n}$, für die $AQ + QA^* = -\text{id}$ gilt.*
- (v) *Für jede positiv definite selbst-adjungierte Matrix $P \in \mathbb{K}^{n \times n}$ gibt es eine positiv semidefinite selbst-adjungierte Matrix $Q \in \mathbb{K}^{n \times n}$ derart, dass $AQ + QA^* = -P$ gilt.*

In den Aussagen (iv) und (v) ist die Matrix Q automatisch sogar positiv definit.¹

Beweis. “(i) $\Rightarrow$ (iii)” Wenn $\lambda \in \mathbb{C}$ ein von Eigenwert mit $\text{Re } \lambda \geq 0$ und Eigenvektor $z \in \mathbb{C}^n$ ist, dann gilt für alle $t \in [0, \infty)$

$$\|e^{tA}z\| = \|e^{t\lambda}z\| = e^{t\text{Re } \lambda} \|z\|,$$

und dieser Ausdruck konvergiert für $t \rightarrow \infty$ nicht gegen 0: Damit finden wir auch ein $x \in \mathbb{K}^n$ mit $\|e^{tA}x\| \not\rightarrow 0$ für $t \rightarrow \infty$.²

“(iii) $\Rightarrow$ (i)” Das folgt aus der Darstellung der Matrixexponentialfunktion mit Hilfe der Jordanschen Normalform, siehe Satz 1.4.10.

¹Man kann auch zeigen, dass Q eindeutig bestimmt ist, dass Q in (iv) und (v) jeweils eindeutig (durch P) bestimmt ist. Haben Sie eine Idee, wie das zum Beispiel geht?

²Warum?

“(ii) $\Rightarrow$ (i)” Klar.

“(i) $\Rightarrow$ (ii)”³ Wegen (i) gibt es ein $t_0 > 0$ mit $\delta := \|e^{t_0 A}\| < 1$. Wir setzen $C := \sup_{t \in [0, t_0]} \|e^{tA}\|$. Wegen der Stetigkeit der Matrix-e-Funktion ist $C < \infty$.

Sei jetzt $t \in [0, \infty)$. Dann finden wir eine Zahl $k \in \mathbb{N}$ und es $s \in [0, t_0)$ mit $t = kt_0 + s$. Also ist

$$\|e^{tA}\| \leq C\delta^k = \frac{C}{\delta} \exp((k+1)t_0 \frac{\log \delta}{t_0}) \leq \frac{C}{\delta} \exp(t \frac{\log \delta}{t_0}).$$

Das zeigt (ii) mit $M = \frac{C}{\delta}$ und $\omega = \frac{\log \delta}{t_0}$.

“(ii) $\Rightarrow$ (v)” Für gegebenes P setzen wir

$$Q := \int_0^\infty e^{tA} P e^{tA^*} dt \in \mathbb{K}^{n \times n};$$

wegen (ii) existiert das Integral. Die Matrix Q ist selbst-adjungiert und positiv semidefinit.

Um zu zeigen, dass Q die gewünschte Gleichung löst, setzen wir $\varphi(t) := e^{tA} P e^{tA^*}$ für alle $t \in [0, \infty)$. Dann ist φ stetig differenzierbar und aus der Produktregel folgt $\dot{\varphi}(t) = A\varphi(t) + \varphi(t)A^*$ für alle $t \in [0, \infty)$. Also ist wegen des Hauptsatzes der Differential- und Integralrechnung

$$\varphi(t) - \varphi(0) = \int_0^t A e^{sA} P e^{sA^*} ds + \int_0^t e^{sA} P e^{sA^*} A^* ds \rightarrow AQ + QA^*$$

für $t \rightarrow \infty$. Zugleich gilt aber wegen (ii) $\varphi(t) \rightarrow 0$ für $t \rightarrow \infty$, also ist $-P = \varphi(0) = AQ + QA^*$.

An dieser Stelle beobachten wir noch, dass es positiv semidefinites Q , dass $AQ + QA^* = -P$ für ein positiv definites P erfüllt, selbst positiv definit sein muss. Wäre nämlich 0 ein Eigenwert von Q mit Eigenvektor $x \in \mathbb{K}^n$, dann wäre

$$0 \neq -x^* P x = x^* (AQ + QA^*) x = 0.$$

“(v) $\Rightarrow$ (iv)” Klar.

“(iv) $\Rightarrow$ (iii)” Sei $\lambda \in \mathbb{C}$ ein Eigenwert von A . Dann ist λ auch ein Eigenwert von A^* ; sei $z \in \mathbb{C}^n$ ein Eigenvektor von A^* zu diesem Eigenwert. Dann gilt

$$-\|z\|^2 = -z^* z = z^* (AQ + QA^*) z = \bar{\lambda} z^* Q z + \lambda z^* Q z = 2 \operatorname{Re} \lambda z^* Q z.$$

Wegen $z^* Q z \geq 0$ folgt $\operatorname{Re} \lambda < 0$. □

Bemerkung 3.1.2 (Die Resolventen von implementierten Halbgruppen). Diese Bemerkung wird in der Vorlesung diskutiert und ist vor allem für diejenigen Personen interessant, die die Grundlagen C_0 -Halbgruppentheorie kennen.

³Aus der Darstellung der Matrix-e-Funktion mit Hilfe der Jordanschen Normalform könnte man stattdessen auch direkt “(iii) $\Rightarrow$ (i)” zeigen. Wir geben stattdessen ein anderes Argument an, dass auch im unendlich Dimensionalen nützlich ist.

Definition 3.1.3 (Spektralschranke). Sei $A \in \mathbb{K}^{n \times n}$. Die Zahl

$$s(A) := \max\{\operatorname{Re} \lambda \mid \lambda \in \mathbb{C} \text{ ist ein Eigenwert von } A\}$$

nennt man die **Spektralschranke von A** .

Bemerkung 3.1.4 (Verhalten von M und ω). (a) Die Äquivalenz von (i) und (iii) in Satz 3.1.1 sagt also gerade, dass die Matrix-e-Funktion von A genau dann exponentiell stabil ist, wenn $s(A) < 0$ gilt. Aus der Darstellung der e-Funktion mit Hilfe der Jordanschen Normalform kann man erkennen, dass man für ω jede Zahl nehmen kann, die $> s(A)$ ist.

(b) Allerdings kann man die beiden Zahlen im Allgemeinen nicht gleich wählen und es kann passieren, dass M explodiert, wenn man mit ω nahe an $s(A)$ herangeht: Betrachte die Matrix

$$A := \begin{pmatrix} -\varepsilon & 1 \\ 0 & -\varepsilon \end{pmatrix}$$

für ein $\varepsilon > 0$. Dann ist $s(A) = -\varepsilon$ und es gilt

$$e^{tA} = e^{-t\varepsilon} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

für alle $t \in [0, \infty)$. Wegen des Eintrags oben rechts sehen sie, dass es kein $M \geq 1$ gibt derart, dass $\|e^{tA}\| \leq Me^{-t\varepsilon}$ für alle $t \in [0, \infty)$ wäre.

Satz 3.1.5 (Konvergenz der Matrixexponentialfunktion für $t \rightarrow \infty$). Sei $A \in \mathbb{K}^{n \times n}$. Dann sind äquivalent:

- (i) Es konvergiert e^{tA} für $t \rightarrow \infty$.
- (ii) Es gilt $s(A) \leq 0$ und $\sigma(A) \cap i\mathbb{R} \subseteq \{0\}$ und falls 0 ein Eigenwert von A ist, dann ist er **semisimpel**, das heißt seine algebraische und geometrische Vielfachheit stimmen überein.

Beweis. Das folgt direkt aus der Existenz der Jordanschen Normalform und der Formel für die Matrix-e-Funktion von Jordanblöcken in Satz 1.4.10. $\square$

3.2 Kriterien für die Negativität der Spektralschranke

Proposition 3.2.1 (Stabilität in für 2×2 -Matrizen). Sei $A \in \mathbb{K}^{2 \times 2}$. Es sei A selbst-adjungiert oder es sei $\mathbb{K} = \mathbb{R}$.

- (a) Es gilt $s(A) < 0$ genau dann, wenn $\operatorname{tr} A < 0$ und $\det A > 0$ ist.⁴

⁴Mit $\operatorname{tr} A$ bezeichnen wir die Spur von A .

(b) Es gilt $s(A) \leq 0$ genau dann, wenn $\operatorname{tr} A \leq 0$ und $\det A \geq 0$ ist.

Beweis. Seien $\lambda_1, \lambda_2 \in \mathbb{C}$ die beiden Eigenwerte von A

Wir betrachten zuerst den Fall, in dem $\lambda_1, \lambda_2 \in \mathbb{R}$ gilt (das deckt insbesondere den Fall ab, in dem A selbstadjungiert ist).

(a) “ $\Rightarrow$ ” Wegen $s(A) < 0$ gilt $\lambda_1, \lambda_2 < 0$ und somit folgt $\operatorname{tr} A = \lambda_1 + \lambda_2 < 0$ und $\det A = \lambda_1 \lambda_2 > 0$.

“ $\Leftarrow$ ” Sei $s(A) \geq 0$. Dann liegt einer der beiden Eigenwerte – sagen wir λ_1 – in $[0, \infty)$. Für den anderen Eigenwert gibt es zwei Möglichkeiten: Entweder λ_2 liegt in $[0, \infty)$ und somit ist $\operatorname{tr} A \geq 0$. Oder λ_2 liegt in $(-\infty, 0)$ und somit ist $\det A \leq 0$.

(b) “ $\Rightarrow$ ” Wenn $s(A) \leq 0$ ist, gilt $\lambda_1, \lambda_2 \leq 0$ und somit folgt $\operatorname{tr} A = \lambda_1 + \lambda_2 \leq 0$ und $\det A = \lambda_1 \lambda_2 \geq 0$.

“ $\Leftarrow$ ” Sei $s(A) > 0$. Dann liegt einer der beiden Eigenwerte – sagen wir λ_1 – in $(0, \infty)$. Für den anderen Eigenwert gibt es zwei Möglichkeiten: Entweder λ_2 liegt in $[0, \infty)$ und somit gilt $\operatorname{tr} A > 0$. Oder λ_2 liegt in $(-\infty, 0)$ und somit gilt $\det A < 0$.

Nun betrachten wir noch den Fall, dass einer der Eigenwerte – sagen wir λ_1 – nicht reell ist. Somit ist A nicht selbst-adjungiert und laut Voraussetzung ist A somit reell. Dann ist aber auch λ_2 nicht reell und es gilt $\lambda_2 = \overline{\lambda_1}$. Nun beweisen wir wieder beide Äquivalenzen:

(a) “ $\Rightarrow$ ” Wegen $s(A) < 0$ gilt $\operatorname{tr} A = \lambda_1 + \lambda_2 = 2 \operatorname{Re} \lambda_1 < 0$ und $\det A = \lambda_1 \lambda_2 = |\lambda_1|^2 > 0$.

“ $\Leftarrow$ ” Sei $s(A) \geq 0$. Weil λ_1 und λ_2 denselben Realteil haben, ist dieser ≥ 0 und somit folgt $\operatorname{tr} A \geq 0$. Insbesondere gilt $\operatorname{tr} A \geq 0$ oder $\det A \leq 0$.

(b) “ $\Rightarrow$ ” Wenn $s(A) \leq 0$ ist, gilt $\operatorname{tr} A = \lambda_1 + \lambda_2 = 2 \operatorname{Re} \lambda_1 \leq 0$ und $\det A = \lambda_1 \lambda_2 = |\lambda_1|^2 \geq 0$.

“ $\Leftarrow$ ” Sei $s(A) > 0$. Weil λ_1 und λ_2 denselben Realteil haben, ist dieser > 0 und somit folgt $\operatorname{tr} A > 0$. Insbesondere gilt $\operatorname{tr} A > 0$ oder $\det A < 0$.⁵ $\square$

Wir geben nun noch ein schönes Kriterium für $s(A) < 0$ an unter der Bedingung, dass die Einträge der Matrix-e-Funktion ≥ 0 sind für alle Zeiten $t \geq 0$. Dazu charakterisieren wir diese Eigenschaft zuerst in folgendem Satz. Für eine Matrix M schreiben wir $M \geq 0$, wenn alle Einträge von M aus $[0, \infty)$ sind.

Satz 3.2.2 (Matrixexponentialfunktion mit Einträge ≥ 0). Sei $A \in \mathbb{K}^{n \times n}$. Es sind äquivalent:

(i) Für jedes $t \in [0, \infty)$ gilt $e^{tA} \geq 0$.

(ii) Alle Einträge von A sind reell and alle nicht-diagonal Einträge von A sind ≥ 0 .

Beweis. “(i) $\Rightarrow$ (ii)” Wegen $A = \lim_{t \downarrow 0} \frac{1}{t}(e^{tA} - \operatorname{id})$ ist A reell. Für Indizes $j \neq k$ und Zeiten $t > 0$ gilt zudem

$$0 \leq \frac{1}{t}(e^{tA})_{jk} = \frac{1}{t}(e^{tA} - \operatorname{id})_{jk} \rightarrow A_{jk}$$

⁵Wobei $\det A < 0$ in diesem Fall nie auftreten kann, denn es ist $\det A = |\lambda_1|^2 \geq 0$.

für $t \downarrow 0$, also $A_{jk} \geq 0$.

“(ii) $\Rightarrow$ (i)” Wegen (ii) gibt es eine Zahl $c \in \mathbb{R}$ mit $A + c \operatorname{id} \geq 0$. Aus der Definition der Matrixexponentialfunktion folgt somit $e^{t(A+c\operatorname{id})} \geq 0$ für alle $t \in [0, \infty)$ und somit $e^{tA} = e^{-tc} e^{t(A+c\operatorname{id})} \geq 0$ für all diese t . $\square$

Satz 3.2.3 (Stabilität von Matrix-e-Funktionen mit nicht-negativen Einträgen). *Sei $A \in \mathbb{R}^{n \times n}$ und alle nicht-diagonal-Einträge von A seien ≥ 0 . Dann sind äquivalent:*

- (i) $s(A) < 0$
- (ii) A ist invertierbar und $-A^{-1} \geq 0$.

Für den Beweis verwenden wir das folgende Lemma:

Lemma 3.2.4. *Sei $A \in \mathbb{K}^{n \times n}$ und sei $\lambda \in \mathbb{C}$ mit $\operatorname{Re} \lambda > s(A)$. Dann gilt*

$$(\lambda \operatorname{id} - A)^{-1} = \int_0^\infty e^{-t\lambda} e^{tA} dt,$$

wobei das Integral absolut konvergiert.

Beweis. Wegen $s(A - \lambda) = s(A) - \operatorname{Re} \lambda < 0$ geht $\|e^{-t\lambda} e^{tA}\|$ für $t \rightarrow \infty$ exponentiell gegen 0 (Bemerkung 3.1.4(a)). Deshalb konvergiert das Integral absolut.

Wir setzen nun $\varphi : [0, \infty) \rightarrow \mathbb{K}^{n \times n}$, $\varphi(t) := e^{-t\lambda} e^{tA} = e^{t(A-\lambda)}$ für $t \geq 0$. Dann ist φ stetig differenzierbar mit Ableitung $\dot{\varphi}(t) = (A - \lambda)\varphi(t)$. Aus dem Hauptsatz der Differential- und Integralrechnung folgt somit für alle $t \geq 0$

$$\varphi(t) - \varphi(0) = \int_0^t (A - \lambda)\varphi(s) ds.$$

Wir haben bereits gesehen, dass $\varphi(t)$ für $t \rightarrow \infty$ gegen 0 strebt, also ist

$$-\operatorname{id} = -\varphi(0) = (A - \lambda) \int_0^t \varphi(s) ds.$$

Durch Multiplikation mit $(A - \lambda)^{-1}$ folgt die Behauptung. $\square$

Beweis von Satz 3.2.3. Vorbemerkung: Laut Satz 3.2.2 gilt $e^{tA} \geq 0$ für alle $t \in [0, \infty)$. Aus Lemma 3.2.4 folgt somit für alle $\lambda \in (s(A), \infty)$

$$(\lambda - A)^{-1} = \int_0^\infty e^{-t\lambda} e^{tA} dt \geq 0.$$

“(i) $\Rightarrow$ (ii)” Wegen $s(A) < 0$ folgt aus der Vorbemerkung mit $\lambda := 0$ die Ungleichung $-A^{-1} \geq 0$.

“(ii) $\Rightarrow$ (i)” Wir zeigen, dass für keine Zahl $\eta < s(A)$ in der Resolventenmenge von A die Ungleichung $(\eta - A)^{-1} \geq 0$. Widerspruchshalber nehmen wir an, dass es ein solches η gibt.

Sei nun $\nu \in \mathbb{C}$ ein Eigenwert von A mit $\operatorname{Re} \nu = s(A)$ und mit Eigenvektor $z \in \mathbb{C}^n$. Mit dem Betrag von Vektoren bezeichnen wir im Folgenden den eintragsweisen Betrag.

Sei $\lambda \in (s(A), \infty)$ und betrachten diejenige Zahl $\mu \in \mathbb{C}$ mit Realteil λ und dem gleichen Imaginärteil wie ν . Wegen Lemma 3.2.4 gilt⁶

$$\begin{aligned} \frac{1}{\lambda - s(A)} |v| &= |(\mu - A)^{-1} v| = \left| \int_0^\infty e^{-t\mu} e^{tA} v \, dt \right| \\ &\leq \int_0^\infty e^{-t\lambda} e^{tA} |v| \, dt = (\lambda - A)^{-1} |v|. \end{aligned}$$

Also folgt $\|(\lambda - A)^{-1} v\| \rightarrow \infty$ für $\lambda \downarrow s(A)$.

Wegen der Annahme $(\eta - A)^{-1} \geq 0$ und wegen $\eta < s(A)$ folgt aber zugleich aus der Resolventengleichung für alle $\lambda > s(A)$

$$0 \leq (\lambda - \eta)(\lambda - A)^{-1}(\eta - A)^{-1} = (\eta - A)^{-1} - (\lambda - A)^{-1},$$

und somit

$$0 \leq (\lambda - A)^{-1} |v| \leq (\eta - A)^{-1} |v|.$$

Das zeigt, dass $\|(\lambda - A)^{-1} v\|$ für $\lambda \downarrow 0$ beschränkt ist. Widerspruch. $\square$

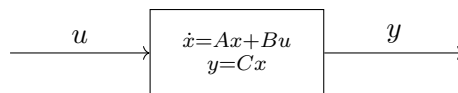
⁶Es ist sehr verlockend, in der folgenden Abschätzung am Ende der ersten Zeile die Eigenwertgleichung $e^{tA} v = e^{t\nu} v$ zu verwenden. Das wollen wir hier aber nicht tun – wir verwenden stattdessen die Abschätzung $|e^{tA} v| \leq e^{tA} |v|$, die für alle Matrizen ≥ 0 und für alle Vektoren gilt, denn das ermöglicht uns eine untere Abschätzung für die Resolvente $(\lambda - A)^{-1}$.

Kapitel 4

Rückkopplung und Stabilisierung

4.1 Statische Rückkopplung

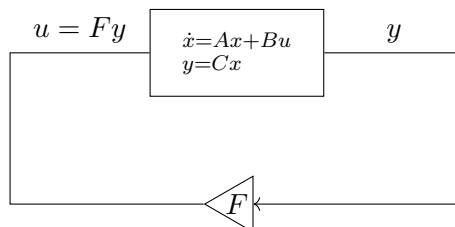
Betrachten Sie folgendes Eingangs-Ausgangs-System ohne Durchgriff:



Wenn A Eigenwerte besitzt, die Realteil > 0 haben, dann kann es – je nach Anfangswert – passieren, dass $x(t)$ exponentiell wächst, selbst wenn man den Eingang u auf 0 setzt. Um das zu verhindern, kann man versuchen, zu jedem Zeitpunkt t das Ausgangssignal $y(t)$ zu nehmen, es mit einer geeigneten Matrix $F \in \mathbb{K}^{m \times p}$ zu multiplizieren, und $Cy(t)$ zum selben Zeitpunkt als Eingang $u(t)$ ins System zurück zu speisen. Man konstruiert also eine **Rückkopplungsschleife** (englisch **feedback loop**). Die Matrix F beschreibt also die Rückkopplung oder, englisch ausgedrückt, das Feedback.

Man spricht von **statischer** Rückkopplung, da $u(t)$ lediglich von y zum Zeitpunkt t abhängt + und nicht zum Beispiel von den Werten, die y in der Vergangenheit – also zu Zeitpunkten für $t -$ hatte.

Das rückgekoppelte System sieht graphisch folgendermaßen aus:



Man sieht direkt, dass der Zustand x im rückgekoppelten System der Differentialgleichung

$$\dot{x} = (A + BFC)x$$

genügt. Wenn man also F so wählen kann, dass alle Eigenwerte von $A+BFC$ Realteil < 0 haben, ist dann konvergiert $x(t)$, unabhängig vom Startwert, für $t \rightarrow \infty$ gegen 0 – man hat das System stabilisiert.

Deshalb liegt es nahe die Frage zu stellen, für welche A , B und C man ein F finden kann, derart, dass $A+BFC$ genau die Eigenwerte hat, die man gerne haben möchte.

Für den Fall $C = \text{id}$ lässt sich die Frage sehr konkret beantworten. Das ist der Inhalt des nächsten Abschnitts.

4.2 Polverschiebungssatz

In diesem Abschnitt argumentieren wir sehr spektraltheoretisch, deshalb konzentrieren wir uns auf den Fall $\mathbb{K} = \mathbb{C}$. Man kann aber auch eine Version des nachfolgenden Satzes 4.2.1 über $\mathbb{R}$ zeigen. Der folgende Satz wird auf Englisch als **pole shifting theorem** oder **pole placement theorem** bezeichnet.

Satz 4.2.1 (Polverschiebungssatz). *Sei $A \in \mathbb{C}^{d \times d}$ und $B \in \mathbb{C}^{n \times n}$. Dann sind äquivalent:*

- (i) *Das Paar (A, B) ist steuerbar.*
- (ii) *Für jedes Polynom $h \in \mathbb{C}[X]$ vom Grad n und mit führendem Koeffizienten 1 gibt es eine Matrix $F \in \mathbb{C}^{m \times n}$ derart, dass $A + BF$ das charakteristische Polynom h hat.*
- (iii) *Es gibt eine Matrix $F \in \mathbb{C}^{m \times n}$ derart, dass $\sigma(A + BF)$ und $\sigma(A)$ sich nicht schneiden.*

Beweis von Satz 4.2.1, Teil 1. “(ii) $\Rightarrow$ (iii)” Wähle ein Polynom $h \in \mathbb{C}[X]$ vom Grad d und mit führendem Koeffizienten 1, dass keine gemeinsame Nullstelle mit dem charakteristischen Polynom von A hat. Laut (ii) gibt es ein $F \in \mathbb{C}^{m \times n}$ derart, dass $A + BF$ charakteristisches Polynom h hat. Somit haben $A + BF$ und A keinen gemeinsamen Eigenwert.

“(iii) $\Rightarrow$ (i)” Wenn (A, B) nicht steuerbar ist, dann gibt es laut Hautus-Kriterium einen Eigenwert $\lambda \in \mathbb{C}$ von A derart, dass $[A - \lambda | B]$ Rang $< n$ hat. Somit finden wir einen Vektor $0 \neq z \in \mathbb{C}^n$ derart, dass $z^T(A - \lambda) = 0$ und $z^T B = 0$ gilt. Für jedes $F \in \mathbb{C}^{m \times n}$ gilt somit

$$z^T(\lambda - (A + BF)) = 0,$$

das heißt λ ist auch ein Eigenwert von $A + BF$. □

Die interessanteste Implikation in Satz 4.2.1 ist die von (i) nach (ii). Den Rest von Abschnitt 4.2 verbringen wir damit, diese Implikation zu beweisen.

Der Beweis, den wir angeben, ist nicht auf größtmögliche Effizienz oder Stringenz getrimmt, sondern soll einige Einblicke und Einsichten in interessante und nützliche Konzepte vermitteln. Wir verfolgen folgende Strategie:

- Wir betrachten zuerst den “single input Fall” (das heißt $m = 1$) und zeigen, dass die Implikation stimmt, falls A diagonalisierbar mit paarweise verschiedenen Eigenwerten ist. (das wird Teil 2 des Beweises sein).
- Wir betrachten nochmals den single input Fall und zeigen die Implikation für den Fall, dass das Zielpolynom h nur einfache Nullstellen hat, von denen keine ein Eigenwert von A ist. Dann kombinieren wir das mit dem vorangehenden Punkt um die Behauptung ohne Einschränkungen im single input Fall zu erhalten (das alles wird Teil 3 des Beweis sein).
- Wir reduzieren den Fall für allgemeines m auf den Fall $m = 1$ (das wird Teil 4 des Beweises sein).

Proposition 4.2.2 (Determinante von $\mathbf{1} + \mathbf{BF}$). *Seien $B \in \mathbb{K}^{n \times m}$ und $F \in \mathbb{K}^{m \times n}$. Dann gilt $\det(\text{id} + BF) = \det(\text{id} + FB)$.*

Beweis. Das besprechen wir in den Übungen. $\square$

Lemma 4.2.3 (Determinante von Rang-1-Störungen von Diagonalmatrizen). *Sei $A \in \mathbb{C}^{n \times n}$ eine Diagonalmatrix und seien $f, b \in \mathbb{C}^n$. Dann gilt*

$$\det(A + bf^T) = \det A + \sum_{k=1}^n b_k f_k \prod_{j \neq k} A_{jj}.$$

Beweis. Weil beide Seiten stetig bezüglich der Einträge von A sind, genügt es die Formel für invertierbare A zu zeigen; sei also $A_{jj} \neq 0$ für alle j . Aus $A + bf^T = A(\text{id} + A^{-1}bf^T)$ folgt mit Proposition 4.2.2

$$\begin{aligned} 0 \neq \det(A + bf^T) &= \det(A) \det(\text{id} + A^{-1}bf^T) \\ &= \det(A) \det(1 + f^T A^{-1}b) = \det(A)(1 + f^T A^{-1}b), \end{aligned}$$

Wegen der Diagonalstruktur von A gilt zudem $f^T A^{-1}b = \sum_{k=1}^n \frac{b_k f_k}{A_{kk}}$, woraus wegen $\det A = \prod_{j=1}^n A_{jj}$ die Behauptung folgt. $\square$

Beispiel 4.2.4 (Steuerbarkeit für diagonale Systemmatrizen mit einem Eingang). Seien $A \in \mathbb{K}^{n \times n}$ eine Diagonalmatrix mit paarweise verschiedenen Diagonaleinträgen und sei $b \in \mathbb{K}$.

Es ist (A, b) genau dann steuerbar, wenn alle Einträge von b ungleich 0 sind.

Beweis. “ $\Rightarrow$ ” Wenn der k -te Eintrag von b gleich 0 ist, dann ist die k -te Zeile der Erreichbarkeitsmatrix $R(A, b)$ gleich 0.

“ $\Leftarrow$ ” Es ist

$$R(A, b) = \begin{pmatrix} b_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & b_n \end{pmatrix} \begin{pmatrix} A_{11}^0 & A_{11}^1 & \dots & A_{11}^{n-1} \\ \vdots & & & \vdots \\ A_{nn}^0 & A_{nn}^1 & \dots & A_{nn}^{n-1} \end{pmatrix}.$$

Der linke Faktor ist invertierbar, weil alle b_k ungleich 0 sind und der rechte Faktor ist invertierbar, weil es sich um eine Vandermonde-Matrix zu n paarweise verschiedenen Stützstellen handelt. $\square$

Beweis von Satz 4.2.1, Teil 2. Wir zeigen die Implikation “(i) $\Rightarrow$ (ii)” im Fall $m = 1$, falls alle n Eigenwerte von A paarweise verschieden sind. Wegen $m = 1$ können wir B als einen Spaltenvektor $b \in \mathbb{C}^n$ schreiben. Das gesuchte F schreiben wir als f^T für einen Spaltenvektor $f \in \mathbb{C}^n$.

In diesem Fall ist A diagonalisierbar und nach einer Ähnlichkeitstransformation dürfen wir annehmen, dass A selbst diagonal ist und alle Diagonaleinträge paarweise verschieden sind. Laut Beispiel 4.2.4 gilt dann wegen der Steuerbarkeit von (A, b) , dass $b_k \neq 0$ ist für alle Indizes k .

Bezeichne nun $p_A(\mu) := \det(\mu - A)$ das charakteristische Polynom von A und für jedes $k \in \{1, \dots, n\}$ bezeichne p_k das Polynom vom Grad $n - 1$, dass durch $p_k(\mu) = \prod_{j \neq k} (\mu - A_{jj})$ definiert ist. Für jedes $f \in \mathbb{C}^n$ ist laut Lemma 4.2.3 ist das charakteristische Polynom von $A + bf^T$ gleich

$$\det(\mu - A - bf^T) = p_A(\mu) - \sum_{k=1}^n b_k f_k p_k(\mu).$$

Die Polynome $p_1, \dots, p_n$ bildet eine Basis des Raumes aller komplexen Polynome vom Grad $\leq n - 1$. Ist nämlich g ein Polynom vom Grad $\leq n - 1$, dann gilt

$$g(\mu) = \sum_{k=1}^n \frac{g(A_{kk})}{p_k(A_{kk})} p_k(\mu)$$

für alle $\mu \in \mathbb{C}$; das überprüft man zuerst an den Stellen $\mu = A_{jj}$ und anschließend folgt es für alle $\mu \in \mathbb{C}$ aufgrund des Identitätssatzes für Polynome.¹

Weil $p_A - h$ ein Polynom vom Grad $\leq n - 1$ ist, finden wir also Zahlen $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ mit $p_A(\mu) - h(\mu) = \sum_{k=1}^n \alpha_k p_k(\mu)$. Nun nutzen wir noch, dass $b_k \neq 0$ für alle k gilt – deshalb gibt es einen Vektor $f \in \mathbb{C}^d$ mit $b_k f_k = \alpha_k$ für alle k . Also ist für dieses f

$$\det(\mu - A - bf^T) = p_A(\mu) - \sum_{k=1}^n \alpha_k p_k(\mu) = h(\mu),$$

was im betrachteten Fall die Behauptung zeigt. $\square$

Satz 4.2.5 (Sherman–Morrison-Formel für Inverse von Rang-1-Störungen). *Sei $A \in \mathbb{C}^{n \times n}$ invertierbar und seien $b, f \in \mathbb{C}^n$. Es ist $A + bf^T$ genau dann invertierbar, wenn $1 + f^T A^{-1} b \neq 0$ ist, und in diesem Fall gilt*

$$(A + bf^T)^{-1} = A^{-1} - \frac{A^{-1} b f^T A^{-1}}{1 + f^T A^{-1} b}.$$

¹Dieses Argument beweist übrigens den Interpolationssatz von Lagrange für paarweise verschiedene Stützstellen.

Beweis. Falls $1 + f^T A b \neq 0$ ist, kann man unmittelbar nachrechnen, dass die gegebene Formel eine Matrix definiert, die invers zu $A^{-1} + b f^T$ ist.

Sei nun umgekehrt $A + b f^T$ invertierbar. Aus Proposition 4.2.2 erhalten wir wie im Beweis von Lemma 4.2.3

$$\det(A + b f^T) = \det(A)(1 + f^T A^{-1} b).$$

also $1 + f^T A^{-1} b \neq 0$. □

Lemma 4.2.6 (Aufspann der Matrixpotenzen). *Sei $A \in \mathbb{C}^{n \times n}$ und seien $\mu_1, \dots, \mu_n \in \mathbb{C} \setminus \sigma(A)$ paarweise verschiedene Zahlen. Dann stimmen die folgenden Untervektorräume von $\mathbb{C}^{n \times n}$ überein:*

- (a) $V_a := \text{span}\{\text{id}, A, \dots, A^{n-1}\}.$
- (b) $V_b := \text{span}\{A^k \mid k \in \mathbb{N}_0\}.$
- (c) $V_c := \text{span}\{(\mu - A)^{-1} \mid \mu \in \mathbb{C} \setminus \sigma(A)\}.$
- (d) $V_d := \text{span}\{(\mu_1 - A)^{-1}, \dots, (\mu_n - A)^{-1}\}.$

Beweis. Wir werden mehrmals verwenden, dass Untervektorräume im endlich-dimensionalen immer abgeschlossen sind.

“ $V_a = V_b$ ” Folgt aus Caley–Hamilton.

“ $V_c \subseteq V_b$ ” Für $\mu > r(A)$ folgt aus der Neumannreihe, dass $(\mu - A)^{-1} \in V_b$ gilt. Für alle anderen $\mu \in \mathbb{C} \setminus \sigma(A)$ folgt es nun aus dem Identitätssatz für holomorphe Funktionen.²

“ $V_b \subseteq V_c$ ” Es gilt $\text{id} = \lim_{\mu \rightarrow \infty} \mu(\mu - A)^{-1} \in V_c$ und somit auch

$$A = \lim_{\mu \rightarrow \infty} \mu A(\mu - A)^{-1} = \lim_{\mu \rightarrow \infty} (\mu^2(\mu - A)^{-1} - \mu \text{id}) \in V_c.$$

Aus der Resolventengleichung folgt, dass V_c abgeschlossen unter Matrixmultiplikation ist, also folgt $A^k \in V_c$ für alle $k \in \mathbb{N}_0$.

“ $V_d \subseteq V_c$ ” Klar.

“ $V_c \subseteq V_d$ ” Sei $\varphi : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}$ ein lineares Funktional, dass auf V_d verschwindet. Wir müssen zeigen, dass φ auch auf V_c verschwindet.

Wegen der Cramerschen Regeln zur Berechnung von inversen Matrizen ist die Abbildung $G : \mu \mapsto \det(\mu - A)(\mu - A)^{-1}$ in jedem Eintrag ein Polynom vom Grad $\leq n - 1$ (und die Singularitäten in den Eigenwerte von A sind hebbar). Somit ist $\varphi \circ G : \mathbb{C} \rightarrow \mathbb{C}$ ein Polynom vom Grad $\leq n - 1$. Dieses Polynom verschwindet an den Stellen $\mu_1, \dots, \mu_n$ und ist deshalb wegen des Identitätssatzes für Polynome konstant 0. Somit ist $\det(\mu - A)\varphi((\mu - A)^{-1}) = 0$ für alle $\mu \in \mathbb{C} \setminus \sigma(A)$, also verschwindet φ auf V_c . □

²Alternativ kann man für diese Inklusion auch mit dem Lemma von Bézout im Ring $\mathbb{C}[X]$ und dem Satz von Cayley–Hamilton argumentieren.

Als unmittelbare Konsequenz aus diesem Lemma erhalten wir:

Korollar 4.2.7 (Steuerbarkeit via Resolventen). *Sei $A \in \mathbb{C}^{n \times n}$ und seien $\mu_1, \dots, \mu_n \in \mathbb{C} \setminus \sigma(A)$ paarweise verschiedene Zahlen. Sei $B \in \mathbb{C}^{n \times m}$. Dann sind äquivalent:*

- (i) *Das Paar (A, B) ist steuerbar.*
- (ii) *Die Matrix $[(\mu_1 - A)^{-1}B \ \dots \ (\mu_n - A)^{-1}B] \in \mathbb{C}^{n \times nm}$ hat vollen Rang, also Rang n .*

Proposition 4.2.8 (Steuerbarkeit bei Störung durch BF). *Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$ und $F \in \mathbb{K}^{m \times n}$. Es ist (A, B) genau dann steuerbar, wenn $(A + BF, B)$ steuerbar ist.*

Beweis. Das folgt leicht aus dem Hautus-Kriterium für Steuerbarkeit. $\square$

Beweis von Satz 4.2.1, Teil 3. Wir zeigen die Implikation “(i) $\Rightarrow$ (ii)” im Fall $m = 1$ für allgemeine h . Wir schreiben wieder $B = b \in \mathbb{C}^n$.

Seien $\lambda_1, \dots, \lambda_n$ die Eigenwerte von A (gezählt mit algebraischer Vielfachheit). Wähle zunächst ein Polynom $\tilde{h}$ vom Grad n mit führendem Koeffizienten 1, das auf keinem λ_k verschwindet und das paarweise verschiedene Nullstellen $\mu_1, \dots, \mu_n$ besitzt. Für jedes μ_k ist dann $\mu_k - A$ invertierbar, also folgt für jedes $f \in \mathbb{C}^n$ aus Satz 4.2.5, dass $\mu_k - A - bf^T$ genau dann singulär ist, wenn

$$f^T(\mu_k - A)^{-1}b = 1$$

ist. Wegen der Steuerbarkeit von (A, b) wissen wir aus Korollar 4.2.7, dass die $n \times n$ -Matrix $M := [(\mu_1 - A)^{-1}b \ \dots \ (\mu_n - A)^{-1}b] \in \mathbb{C}^{n \times nm}$ invertierbar ist, also gibt es (genau) ein $f \in \mathbb{C}^n$ mit $f^T M = (1, \dots, 1)$. Für diese Wahl von f ist $\mu_k - A - bf^T$ für jedes k singulär, also besitzt $A + bf^T$ die Eigenwerte $\mu_1, \dots, \mu_n$. Weil die μ_k alle paarweise verschieden sind, ist das charakteristische Polynom von $\tilde{A} := A + bf^T$ gleich $\tilde{h}$.

Laut Proposition 4.2.8 ist auch $(\tilde{A}, b)$ steuerbar. Da $\tilde{A}$ die paarweise verschiedenen Eigenwerte $\mu_1, \dots, \mu_n$ hat, gibt es laut dem bereits geführten Teil 2 des Beweises ein $\tilde{f} \in \mathbb{C}^n$ derart, dass

$$A + b(f + \tilde{f})^T = \tilde{A} + b\tilde{f}^T$$

das charakteristische Polynome h besitzt. $\square$

Zum Schluss beweisen wir nun noch den Fall für allgemeines $m \in \mathbb{N}$, indem wir ihn auf den Fall $m = 1$ zurückführen.

Beweis von Satz 4.2.1, Teil 3. Wir zeigen die Implikation “(i) $\Rightarrow$ (ii)” für allgemeine m , und zwar per Induktion über m .

Für $m = 1$ wissen wir bereits, dass die Implikation stimmt. Sei sie nun für ein festes $m \in \mathbb{N}$ gezeigt.

Sei $[B, b] \in \mathbb{C}^{n \times (m+1)}$ mit $B \in \mathbb{C}^{n \times m}$ und sei $(A, [B, b])$ steuerbar. Wir unterscheiden drei Fälle:

1. *Fall:* Wenn sogar (A, B) steuerbar ist, finden wir laut Induktionsvoraussetzung ein $\tilde{F} \in \mathbb{C}^{m \times n}$ derart, dass $A + B\tilde{F}$ das charakteristische Polynom h besitzt. Indem wir $\tilde{F}$ unten um eine Nullzeile ergänzen erhalten wir eine Matrix $F \in \mathbb{C}^{(m+1) \times n}$ mit $A + [B, b]F = A + B\tilde{F}$ und sind fertig.

2. *Fall:* Wenn $B = 0$ gilt, ist (A, b) steuerbar. Dann finden wir wie bereits gezeigt ein $f \in \mathbb{C}^n$ derart, dass $A + bf^T$ das charakteristische Polynom h besitzt. Indem wir über f^T insgesamt m Nullzeilen hinzufügen, sind wir fertig.

3. *Fall:* Es ist (A, B) nicht steuerbar, aber $B \neq 0$. Laut Proposition 2.2.15 können wir dann nach einer Ähnlichkeitstransformation annehmen, dass

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \quad \text{und} \quad B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$$

mit Dimensionen $A_{11} \in \mathbb{C}^{n_1 \times n_1}$, $A_{12} \in \mathbb{C}^{n_1 \times n_2}$, $A_{22} \in \mathbb{C}^{n_2 \times n_2}$ sowie $B_1 \in \mathbb{C}^{n_1 \times m}$. Ebenso können wir b zerlegen in zwei übereinander stehende Vektoren $b_1 \in \mathbb{C}^{n_1}$ und $b_2 \in \mathbb{C}^{n_2}$. Hierbei ist (A_{11}, B_1) steuerbar.

Bei Betrachten der Erreichbarkeitsmatrizen $R(A, [B, b])$ und $R(A_{22}, b_2)$ sieht man, dass (A_{22}, b_2) steuerbar ist (wegen der Steuerbarkeit von $(A, [B, b])$).

Wir faktorisieren nun das Polynom h in Polynome h_1 und h_2 vom Grad n_1 und n_2 und mit je führendem Koeffizienten 1. Laut Induktionshypothese³ gibt es eine Matrix $F_1 \in \mathbb{C}^{m \times n_1}$ derart, dass $A_{11} + B_1 F_1$ das charakteristische Polynom h_1 hat. Und laut des bereits vorab behandelten Falls $m = 1$ gibt es ein $f_2 \in \mathbb{C}^{n_2}$ derart, dass $A_{22} + b_2 f_2^T$ das charakteristische Polynom h_2 hat. Damit besitzt

$$A + \begin{bmatrix} B_1 & b_1 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} F_1 & 0 \\ 0 & f_2^T \end{bmatrix} = \begin{bmatrix} A_{11} + B_1 F_1 & A_{12} + b_1 f_2^T \\ 0 & A_{22} + b_2 f_2^T \end{bmatrix}$$

das charakteristische Polynom $h_1 h_2 = h$. □

4.3 Stabilisierbare Systeme

Satz 4.3.1 (Asymptotische 0-Steuerbarkeit und Stabilisierbarkeit). *Sei $A \in \mathbb{K}^{n \times n}$ und $B \in \mathbb{K}^{n \times m}$. Dann sind äquivalent:*

- (i) *Das Paar (A, B) ist **asymptotisch 0-steuerbar**, das heißt für alle $x_0 \in \mathbb{K}^n$ gibt es ein $u \in L_{\text{loc}}^1([0, \infty); \mathbb{K}^m)$ derart, dass die Lösung $x : [0, \infty) \rightarrow \mathbb{K}^n$ von*

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0$$

die Bedingung $\lim_{t \rightarrow \infty} x(t) = 0$ erfüllt.

³Wie genau lautet eigentlich die Induktionshypothese? Können Sie sie vollständig – inklusive aller Quantoren – aufschreiben?

- (ii) Das Paar (A, B) ist **stabilisierbar**, das heißt es gibt eine Matrix $F \in \mathbb{K}^{m \times n}$ derart, dass $s(A + BF) < 0$.

Falls A und B von der Form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} B_1 \\ 0 \end{pmatrix}$$

sind und (A_{11}, B_1) steuerbar ist,⁴ dann sind die oben stehenden Aussagen (i) und (ii) außerdem äquivalent zu:

- (iii) Die Spektralschranke von A_{22} erfüllt $s(A_{22}) < 0$.

Beweis. Aufgrund der Proposition 2.2.15 können wir durchgehend annehmen, dass A und B die oben genannte Blockstruktur haben.

“(i) $\Rightarrow$ (iii)” Falls A_{22} einen Eigenwert λ mit Realteil ≥ 0 und Eigenvektor x_2 hat, dann die die zweite Komponenten der Lösung x von

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = \begin{pmatrix} 0 \\ x_2 \end{pmatrix}$$

zum Zeitpunkt t gleich $e^{t\lambda}x_2$, und zwar unabhängig von der Eingangsfunktion. Insbesondere konvergiert die Lösung für kein u gegen 0 für $t \rightarrow \infty$ und somit ist (A, B) nicht asymptotisch 0-steuerbar.

“(iii) $\Rightarrow$ (ii)” Wegen der Steuerbarkeit von (A_{11}, B_1) gibt es wegen des Polverschiebungssatzes 4.2.1 eine Matrix F_1 von passender Dimension derart, dass $s(A_{11} + B_1F_1) < 0$ ist. Somit haben alle Eigenwerte von

$$A + B \begin{pmatrix} F & 0 \end{pmatrix} = \begin{pmatrix} A_{11} + B_1F_1 & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

Realteil < 0 .

“(ii) $\Rightarrow$ (i)” Wähle F wie in (ii) und sei $x_0 \in \mathbb{K}^n$. Sei $x(t) := e^{t(A+BF)}x_0$ und $u(t) := Fx(t)$ für alle $t \in [0, \infty)$. Dann gilt $x(t) \rightarrow 0$ für $t \rightarrow \infty$, es ist $x(0) = x_0$ und

$$\dot{x}(t) = Ax(t) + BFx(t) = Ax(t) + Bu(t)$$

für alle $t \in [0, \infty)$. Also haben wir ein $u \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^m)$ (sogar in $L^1([0, \infty); \mathbb{K}^m)$) gefunden, derart, dass die Lösung des Systems mit Startwert x_0 für $t \rightarrow \infty$ gegen 0 strebt. □

Bemerkung 4.3.2 (Entdeckbarkeit). Ähnlich wie Beobachtbarkeit dual zu Steuerbarkeit ist, gibt es auch zur Stabilisierbarkeit einen dualen Begriff, nämlich **Entdeckbarkeit** (englisch **detectability**). Darauf werden wir im Abschnitt 4.5 zurückkommen.

⁴Wobei wir hier zulassen wollen, dass manche der involvierten Matrizen die Höhe oder Breite 0 haben, damit wir auch die Fälle abdecken, in denen $B = 0$ ist oder in denen (A, B) selbst steuerbar ist, siehe Proposition 2.2.15. Hierbei fassen wir, falls A_{11} und B_1 leere Matrizen ist, die Aussage “ (A_{11}, B_1) ist steuerbar” als wahr auf.

4.4 Port-Hamiltonsche Systeme

Wir erinnern daran, dass wir $\mathbb{K}^n$ in diesem Kurs immer mit der Euklidischen Norm ausstatten und dass wir den Matrixraum $\mathbb{K}^{n \times n}$ mit der induzierten Operatornorm ausstatten.

Definition 4.4.1 (Dissipativität und Kontraktivität). Seien $A, T \in \mathbb{K}^{n \times n}$.

- (a) Die Matrix T heißt **kontraktiv**, wenn $\|T\| \leq 1$ gilt.⁵
- (b) Die Matrix A heißt **dissipativ**, falls $\operatorname{Re}(x^*Ax) \leq 0$ für alle $x \in \mathbb{K}^n$ gilt.⁶

Proposition 4.4.2 (Kontraktivität der e-Funktion). Für jedes $A \in \mathbb{K}^{n \times n}$ sind äquivalent:

- (a) Für jedes $t \in [0, \infty)$ ist e^{tA} kontraktiv.
- (b) Die Matrix A ist dissipativ.

Beweis. Wir bemerken zunächst, dass für jedes $x \in \mathbb{K}^n$ die Abbildung $[0, \infty) \ni t \mapsto \|e^{tA}x\|^2 \in [0, \infty)$ differenzierbar ist mit Ableitung

$$\frac{d}{dt} \|e^{tA}x\|^2 = 2 \operatorname{Re} \left((e^{tA}x)^* A e^{tA}x \right)$$

für alle $t \in [0, \infty)$.

“(a) $\Rightarrow$ (b)” Sei $x \in \mathbb{K}^n$. Wegen der Kontraktivität von e^{tA} ist die Abbildung $[0, \infty) \ni t \mapsto \|e^{tA}x\|^2$ monoton fallend. Wegen der Vorbemerkung des Beweis folgt $\operatorname{Re} x^*Ax \leq 0$.

“(b) $\Rightarrow$ (a)” Sei $\varepsilon > 0$. Es genügt zu zeigen, dass $\|e^{t(A-\varepsilon)}\| \leq 1$ für alle $t \in [0, \infty)$ gilt.⁷

Wir dazu das Gegenteil an und verwenden die Abkürzung $\tilde{A} := A - \varepsilon$. Sei also $x \in \mathbb{K}^n$ ein Element der abgeschlossenen Einheitskugel und ein $t > 0$ derart, dass $\|e^{t\tilde{A}}x\|^2 > 1$ ist. Setze $t_0 := \inf \{t \in [0, \infty) \mid \|e^{t\tilde{A}}x\|^2 > 1\}$. Wegen der Stetigkeit der Matrix-e-Funktion ist dann $\|e^{t_0\tilde{A}}x\|^2 = 1$ und in jeder rechtseitigen Umgebung von t_0 gibt es ein t mit $\|e^{t\tilde{A}}x\|^2 > 1$. Somit ist

$$0 \leq \frac{d}{dt} \|e^{t\tilde{A}}x\|^2 \Big|_{t=t_0} = 2 \operatorname{Re} \left((e^{t_0\tilde{A}}x)^* \tilde{A} e^{t_0\tilde{A}}x \right) \leq -2\varepsilon \|e^{t_0\tilde{A}}x\|^2 = -2\varepsilon,$$

wobei die letzte Ungleichung aus der Dissipativität von A folgt. Widerspruch. $\square$

⁵Die Literatur ist hier etwas uneinheitlich: Man Autorinnen und Autoren bezeichnen die deutlich stärkere Eigenschaft $\|T\| < 1$ als **kontraktiv**.

⁶In anderen Worten, wenn der numerische Wertebereich von A in der abgeschlossenen linken Halbebene liegt.

⁷Warum?

Definition 4.4.3 (Port-Hamiltonsche Systeme). Ein System der Form

$$\begin{cases} \dot{x} &= AHx + Bu, \\ y &= B^*Hx \end{cases}$$

heißt endlich-dimensionales, lineares **port-Hamiltonsches System**, falls $B \in \mathbb{K}^{n \times m}$ ist, $A \in \mathbb{K}^{n \times n}$ dissipativ ist und $H \in \mathbb{K}^{n \times n}$ selbst-adjungiert und positiv definit ist.

Für $x \in \mathbb{K}^n$ heißt die Zahl $\frac{1}{2}x^*Hx \in [0, \infty)$ die **Energie** des Systems im Zustand x .

Bemerkung 4.4.4 (Zerlegung in selbst-adjungierten und schiefadjungierten Teil). Jede Matrix $A \in \mathbb{K}^{n \times n}$ kann man in der Form $A = J - R$ schreiben, wobei R selbst-adjungiert und J schief-adjungiert ist.⁸ Man kann leicht nachprüfen, dass A genau dann dissipativ ist, wenn R positiv semidefinit ist.

Proposition 4.4.5 (Energieänderung von port-Hamiltonschen Systemen). Für eine⁹ Lösung x eines port-Hamiltonschen Systems gilt mit der Notation von Definition 4.4.3 und mit der Zerlegung $A = J - R$ aus Bemerkung 4.4.4

$$\frac{d}{dt} \frac{1}{2} x^*(t) H x(t) = -(Hx(t))^* R H x(t) + \operatorname{Re} \left(u(t)^* y(t) \right) \leq \operatorname{Re} \left(u(t)^* y(t) \right)$$

für jedes $t \in [0, \infty)$. Insbesondere ist die Energie monoton abnehmend, wenn der Eingang u konstant 0 ist.

Beweis. Es ist

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} x(t)^* H x(t) &= \operatorname{Re} \left(\dot{x}(t)^* H x(t) \right) \\ &= \operatorname{Re} \left((Hx(t))^* A^* H x(t) \right) + \operatorname{Re} \left(u(t)^* B^* H x(t) \right) \\ &= -(Hx(t))^* R H x(t) + \operatorname{Re} \left(u(t)^* y(t) \right) \end{aligned}$$

für alle $t \in [0, \infty)$. □

Theorem 4.4.6 (Stabilisierbarkeit und Steuerbarkeit port-Hamiltonscher Systeme). Sei $A \in \mathbb{K}^{n \times n}$ dissipativ und $H \in \mathbb{K}^{n \times m}$ selbst-adjungiert und positiv semidefinit. Sei $B \in \mathbb{K}^{n \times m}$. Dann sind äquivalent:

- (i) Das Paar (AH, B) ist stabilisierbar.
- (ii) Es gibt eine Matrix $F \in \mathbb{K}^{m \times m}$ derart, dass $s(AH + BFB^*H) < 0$ gilt.
- (iii) Für jede Zahl $k > 0$ gilt $s(AH - kBB^*H) < 0$.

⁸Wie geht das?

⁹Weshalb steht hier „eine“ Lösung und nicht „die“ Lösung?

Beweis. “(iii) $\Rightarrow$ (ii)” Wähle $F = -k \text{id}$.

“(ii) $\Rightarrow$ (i)” Dies Implikation folgt direkt aus der Definition von Stabilisierbarkeit – betrachte einfach die Matrix $\tilde{F} := FB^*H$, dann ist $s(AH + B\tilde{F}) < 0$.

“(i) $\Rightarrow$ (iii)” Sei $k \in (0, \infty)$ und sei $\lambda \in \mathbb{C}$ ein Eigenwert von $AH - kBB^*H$. Dann gibt es ein $x \in \mathbb{C}^n$ mit $x^*(A - kBB^*)H = \lambda x^*$. Wir setzen $z := H^{-1}x$. Es ist $z \neq 0$ und es gilt

$$x^*(A - kBB^*)x = \lambda x^*H^{-1}x.$$

Somit folgt

$$\operatorname{Re} \lambda = \frac{1}{x^*H^{-1}x} \left(\operatorname{Re} (x^*Ax) - k \|B^*x\|^2 \right) \leq 0,$$

wobei die Ungleichung aus der Dissipativität von A folgt. Wir nehmen nun widerspruchshalber an, dass die Ungleichung eine Gleichung ist. Dann folgt $\operatorname{Re} (x^*Ax) = 0$ und $B^*x = 0$, also $x^*B = 0$. Folglich ist

$$\lambda x^* = x^*AH.$$

Ds zeigt, dass die Matrix $[\lambda - AH, B]$ nicht surjektiv nach $\mathbb{C}^n$ abbildet. Also folgt aus dem Hautus-Kriterium für Stabilisierbarkeit, siehe Aufgabe 1 auf Blatt 7, dass (AH, B) nicht stabilisierbar ist. $\square$

4.5 Dynamische Rückkopplung

Wie Sie in den Übungen gesehen haben, lässt sich das System

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

auch dann, wenn (A, B) steuerbar und (A, C) beobachtbar ist, nicht unbedingt durch statisches Feedback stabilisieren. Eine solche Stabilisierung ist allerdings mit Hilfe von dynamischem Feedback möglich – das heißt, wenn man als Controller ebenfalls ein dynamisches System verwendet.

Wir benötigen dazu den folgenden Begriff:

Definition 4.5.1 (Asymptotische 0-Beobachtbarkeit). Sei $A \in \mathbb{K}^{n \times n}$ und $C \in \mathbb{K}^{p \times n}$. Das Paar (A, C) heißt **asymptotisch 0-beobachtbar** oder **entdeckbar**, falls für jedes $x_0 \in \mathbb{K}^n$ gilt: Wenn $Ce^{tA}x_0 = 0$ für alle $t \in [0, \infty)$ gilt, dann ist $\lim_{t \rightarrow \infty} e^{tA}x_0 = 0$.

In Satz 4.3.1(i) haben wir asymptotische 0-Steuerbarkeit definiert. Dieser Begriff verhält sich dual zu asymptotischer 0-Beobachtbarkeit:

Proposition 4.5.2 (Asympt. 0-Beobachtbarkeit und asympt. 0-Steuerbarkeit). Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$ und $C \in \mathbb{K}^{p \times n}$.

- (a) Es ist (A, B) asymptotisch 0-steuerbar genau dann, wenn (A^T, B^T) asymptotisch 0-beobachtbar ist.
- (b) Es ist (A, C) asymptotisch 0-beobachtbar genau dann, wenn (A^T, C^T) asymptotisch 0-steuerbar ist.

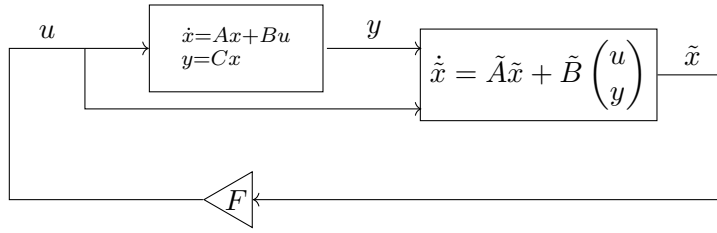
Beweis. (a) Den Beweis stellen wir als Übungsaufgabe.

(b) Das folgt aus (a) durch Vertauschen der Matrizen mit ihren Transponierten. $\square$

Definition 4.5.3 (Dynamisches Feedback). Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$. Wenn wir sagen, dass man das System

$$(\Sigma) \quad \begin{cases} \dot{x} &= Ax + Bu, \\ y &= Cx \end{cases}$$

durch einen dynamischen Regler stabilisieren kann, falls es eine Dimension $\tilde{n} \in \mathbb{N}$ und Matrizen $\tilde{A} \in \mathbb{K}^{\tilde{n} \times \tilde{n}}$, $\tilde{B} \in \mathbb{K}^{\tilde{n} \times (m+p)}$ und $F \in \mathbb{K}^{m \times \tilde{n}}$ gibt derart, dass die Systemmatrix $\hat{A} \in \mathbb{K}^{(n+\tilde{n}) \times (n+\tilde{n})}$ des geschlossenen Systems



(mit Zustandsvektor $(x, \tilde{x}) \in \mathbb{K}^{n+\tilde{n}}$) die Bedingung $s(\hat{A}) < 0$ erfüllt.

Satz 4.5.4 (Stabilisierung durch dynamische Rückkopplung). Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$. Dann sind äquivalent:

- (i) Das durch A , B und C gegebene System lässt sich durch einen dynamischen Regler stabilisieren.
- (ii) (A, B) ist asymptotisch 0-steuerbar und (A, C) ist asymptotisch 0-beobachtbar.

Beweis. Wir beobachten zunächst, dass die Systemmatrix $\hat{A}$ des geschlossenen Systems durch

$$\hat{A} = \begin{pmatrix} A & BF \\ \tilde{B}_2 C & \tilde{A} + \tilde{B}_1 F \end{pmatrix}$$

gegeben ist, wobei wir $\tilde{B}$ in der Form $\tilde{B} = (\tilde{B}_1, \tilde{B}_2)$ mit $\tilde{B}_1 \in \mathbb{K}^{\tilde{n} \times m}$ und $\tilde{B}_2 \in \mathbb{K}^{\tilde{n} \times p}$ schreiben.

“(i) $\Rightarrow$ (ii)” Sei $s(\hat{A}) < 0$. Betrachte ein $x_0 \in \mathbb{K}^n$. Seien $x(t)$ und $\tilde{x}(t)$ die beiden Komponenten von $e^{t\hat{A}}(x_0, 0)^T$ und sei $u(t) := F\tilde{x}(t)$ für jedes $t \geq 0$. Es gilt $x(t) \rightarrow 0$

für $t \rightarrow \infty$, die Funktion x erfüllt die Anfangsbedingung $x(0) = x_0$ und x löst die Differentialgleichung

$$\dot{x} = Ax + BF\tilde{x} = Ax + Bu.$$

Damit haben wir gezeigt, dass (A, B) asymptotisch 0-steuerbar ist.

Es gilt auch $s(\hat{A}^T) < 0$ und $\hat{A}^T$ hat die Form

$$\hat{A}^T = \begin{pmatrix} A^T & C^T \tilde{B}_2^T \\ F^T B^T & \tilde{A}^T + F^T \tilde{B}_1^T \end{pmatrix}$$

Indem wir dasselbe Argument wie oben wiederholen, sehen wir, dass (A^T, C^T) asymptotisch 0-steuerbar ist, und somit ist (A, C) laut Proposition 4.5.2 asymptotisch 0-beobachtbar.

“(ii) $\Rightarrow$ (i)” Wir wählen $\tilde{n} = n$. Weil (A, B) asymptotisch 0-steuerbar ist, gibt es laut Satz 4.3.1 eine Matrix $F \in \mathbb{K}^{m \times n}$ derart, dass $s(A + BF) < 0$ gilt. Dual dazu (Proposition 4.5.2) gibt es wegen der asymptotischen 0-Beobachtbarkeit von (A, C) eine Matrix $L \in \mathbb{K}^{p \times n}$ derart, dass $s(A + LC) < 0$ ist.¹⁰

Wir wählen nun $\tilde{B}_2 := -L$, $\tilde{B}_1 := B$ und $\tilde{A} := A + LC$. Damit ist

$$\begin{aligned} \hat{A} &= \begin{pmatrix} A & BF \\ -LC & A + LC + BF \end{pmatrix} \\ &= \begin{pmatrix} \text{id} & 0 \\ \text{id} & \text{id} \end{pmatrix} \begin{pmatrix} A + BF & BF \\ 0 & A + LC \end{pmatrix} \begin{pmatrix} \text{id} & 0 \\ -\text{id} & \text{id} \end{pmatrix} \end{aligned}$$

Die außenstehenden Matrizen in diesem Produkt sind inverse zueinander, also ist die Matrix in der Mitte ähnlich zu $\hat{A}$ und somit die selben Eigenwerte von $\hat{A}$. Deshalb ist $s(\hat{A}) < 0$. $\square$

Beispiel 4.5.5 (Stabilisierung durch einen Integrator). Betrachten Sie die Matrizen

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

Dann ist (A, B) steuerbar und (A, C) beobachtbar, aber sie wissen aus Aufgabe 2 auf Blatt 6, dass das System nicht durch eine statische lineare Rückkopplung stabilisiert werden kann.

Satz 4.5.4 sagt uns allerdings, dass wir das System durch eine dynamische Rückkopplung stabilisieren können. Der Beweis des Satz gibt uns zudem eine explizite Möglichkeit um das zu tun:

Dazu wählen wir $F = (-1, -1)$ und $L = (-1, -1)^T$. Dann haben die Matrizen

$$A + BF = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} \quad \text{und} \quad A + LC = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$$

¹⁰Warum genau gibt es dieses L ?

jeweils Spektralschranke < 0 .¹¹ Wir folgen dem Beweis von Satz 4.5.4 und wählen

$$\tilde{B} = (B, -L) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{und} \quad \tilde{A} = A + = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}.$$

Wir wissen aus dem Beweis des Satzes, dass diese Wahl von $\tilde{A}$, $\tilde{B}$ und F das System stabilisiert. Wer möchte, kann natürlich auch explizit nachrechnen (am einfachsten mit einem Computer), dass die Matrix

$$\hat{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & 0 & -2 & -1 \end{pmatrix}$$

die Bedingung $s(\hat{A}) < 0$ gilt (aber das wissen wir ja auch schon aus dem Beweis).

¹¹Wie können Sie das sofort sehen?

Kapitel 5

Beobachter und optimale Steuerung

5.1 Luenberger-Beobachter

Stellen Sie sich ein technisches System vor, das sich durch das mathematische System

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

modellieren lässt. Nehmen Sie außerdem an, dass folgendes erfüllt ist:

- Sie können zu jedem Zeitpunkt $t \geq 0$ das Eingangssignal $u(t)$ und das Ausgangssignal $y(t)$ messen; diese Messungen unterliegen aber Störungen, das heißt, sie messen in Wirklichkeit Signale $\tilde{u}(t)$ und $\tilde{y}(t)$, die von den tatsächlichen Signalen etwas abweichen können.
- Sie kennen möglicherweise den Startzustand x_0 nicht oder nicht genau, sondern müssen mit einem abweichenden Anfangszustand $\hat{x}_0$ arbeiten (den Sie irgendwie schätzen oder notfalls zum Beispiel einfach als 0 annehmen.)

In dieser Situation möchten Sie aus den zu jedem Zeitpunkt t gemessenen Werten $\tilde{u}(t)$ und $\tilde{y}(t)$ eine Schätzung $\hat{x}(t)$ des Zustands $x(t)$ zum Zeitpunkt t berechnen. Sie wünschen sich natürlich die folgenden Eigenschaften ihres Schätzers:

- Selbst wenn Ihre Schätzung von $x(0)$ völlig falsch wäre (zum Beispiel, weil Sie $x(0)$ nicht kennen), wäre es gut, wenn $\hat{x}(t)$ zumindest nach einiger Zeit (also für große t) nicht mehr weit von $x(t)$ entfernt liegt.
- Die Messfehler $\tilde{y}(t) - y(t)$ und $\tilde{u}(t) - u(t)$ sollen keinen allzu großen Fehler $\hat{x}(t) - x(t)$ verursachen.

Diese Ziele kann man mit Hilfe eines **Luenberger-Beobachters** umsetzen. Die wesentlichen Ideen dabei sind folgende:

- Während das reale System läuft, simulieren Sie das System zugleich (zum Beispiel auf einem Computer oder einem elektronischen Steuergerät) in Echtzeit mit denselben Matrizen A, B, C . Den Zustand des simulierten Systems bezeichnen wir als $\hat{x}$. Als Startwert verwenden wir einen Vektor $\hat{x}_0$; dessen Wahl hängt davon ab, welche Informationen wir über den tatsächlichen Startwert x_0 haben. Als Eingangssignal $\hat{u}$ verwenden wir das gemessene Eingangssignal $\tilde{u}$, das heißt wir setzen $\hat{u} := \tilde{u}$. Als Ausgangssignal erhalten wir $\hat{y} := C\hat{x}$. Damit lautet das simulierte System insgesamt

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + B\tilde{u}, \\ \hat{y} = C\hat{x}, \end{cases} \quad (5.1.1)$$

mit Startwert $\hat{x}(0) = \hat{x}_0$.

- Zu jedem Zeitpunkt vergleichen Sie den Ausgang $\hat{y}(t)$ des simulierten Systems mit dem gemessenen Ausgang $\tilde{y}(t)$ des tatsächlichen Systems. Die Abweichung $\hat{y}(t) - \tilde{y}(t)$ interpretieren Sie als Maß dafür, wie weit Ihr geschätzter Zustand $\hat{x}(t)$ vom tatsächlichen Zustand $x(t)$ entfernt ist. Gewichtet mit einer Feedbackmatrix $F \in \mathbb{K}^{n \times p}$ speisen Sie diese Abweichung wieder zurück in Ihr simuliertes System und nutzen es als Korrekturterm für die Differentialgleichung für $\hat{x}$. Anstelle von (5.1.1) simulieren Sie also die Differentialgleichung

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + B\tilde{u} + F(\hat{y} - \tilde{y}), \\ \hat{y} = C\hat{x}, \end{cases} \quad (5.1.2)$$

mit Startwert $\hat{x}(0) = \hat{x}_0$.

Die Differentialgleichung (5.1.2) ist ein **Luenberger-Beobachter** für das technische System, dessen Zustand $x(t)$ Sie schätzen möchten. Für den Fehler $e := \hat{x} - x$ erhält man die Differentialgleichung

$$\begin{aligned} \dot{e} &= Ae + B(\tilde{u} - u) + F(\hat{y} - \tilde{y}) \\ &= Ae + B(\tilde{u} - u) + FCe + F(y - \tilde{y}) \\ &= (A + FC)e + B(\tilde{u} - u) + F(y - \tilde{y}). \end{aligned}$$

Damit erhalten wir:

Satz 5.1.1 (Fehlerterm im Luenberger-Beobachter). *Sei (A, C) asymptotisch 0-beobachtbar und sei F derart, dass $s(A+FC) < 0$ gilt.¹ Dann gibt es eine Zahl $M \geq 0$ derart, dass für den Fehler e des soeben beschriebenen Luenberger-Beobachters² die Abschätzung $\limsup_{t \rightarrow \infty} \|e(t)\| \leq M \sup_{s \in [0, \infty)} (\|\tilde{u}(s) - u(s)\| + \|\tilde{y}(s) - y(s)\|)$ gilt.*

¹Die Existenz solcher F folgt, wie Sie wissen, aus der asymptotischen 0-Beobachtbarkeit.

²Unabhängig von den Startwerten x_0 und $\hat{x}_0$.

Beweis. Das folgt unmittelbar aus der Differentialgleichung

$$\dot{e} + (A + FC) + B(\tilde{u} - u) + F(y - \tilde{y})$$

und aus dem nachfolgenden Lemma. $\square$

Lemma 5.1.2 (Stabile Systeme mit beschränkter Störung). *Sei $s(A) < 0$. Dann gibt es eine Zahl $M \geq 0$ deart, dass für jedes messbare und beschränkte Funktion $u : [0, \infty) \rightarrow \mathbb{K}^m$ und jedes $x_0 \in \mathbb{K}^n$ die Lösung x des Anfangswertproblems*

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0$$

die Abschätzung $\limsup_{t \rightarrow \infty} \|x(t)\| \leq M \sup_{s \in [0, \infty)} \|u(s)\|$ erfüllt.

Beweis. Den Beweis stellen wir als Übungsaufgabe. $\square$

5.2 Lineare Optimalsteuerung mit quadratischen Kosten

In Abschnitt 2.2 haben wir uns darüber Gedanken gemacht wann ein System von einem Zustand x_0 zu einem anderen Zustand x_1 mit Hilfe einer Kontrollfunktion u gesteuert werden kann. In Abschnitt 4.1 und 4.5 haben wir dann gelernt mit welchen Rückkopplungen ein gewünschtes Langzeitverhalten des Systems zu erreichen ist. Dabei haben wir nur darauf geachtet, dass das Rückkopplung ein stabiles geschlossenes System erzeugt.

In physikalischen Systemen unterliegen die Zustände und Eingänge oft Nebenbedingungen, denen wir bei dieser Modellierung keine Rechnung tragen. So kann es sein, dass ein System durch einen Motor beeinflusst wird, der nur unter hohen (Strom-)Kosten ein gewisses Drehmoment erreichen kann. Oder wir haben es mit einer Stromversorgung in einem RLC-Netzwerk in dem die Stromversorgung (die als Steuerung fungiert) keine zu hohen Spannungen ausgeben kann. Ferner wollen wir gegebenenfalls Zustände mit großer Norm vermeiden, da diese mit physikalischen Zuständen korrelieren, die nicht erlangt werden können, da beispielweise durch zu hohen Stress Bauteile kaputt gehen können.

Ein mathematisches Modell, das diese Limitationen berücksichtigt, wollen wir nun im Folgenden betrachten. Dazu betrachten wir zunächst ein System der Form

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{5.2.1}$$

mit $A \in \mathbb{K}^{n \times n}$ und $B \in \mathbb{K}^{n \times m}$. Um ungewünschtes Verhalten bestrafen zu können definieren wir für alle Zeiten $t_0, t_1 \in \mathbb{R}$ mit $t_0 < t_1$, für alle Anfangswerte $x_0 \in \mathbb{K}^n$ und alle Kontrollfunktionen $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ das Kostenfunktional

$$J(t_0, t_1, x_0, u) := \int_{t_0}^{t_1} u(t)^* Ru(t) + x(t)^* Qx(t) dt + x(t_1)^* Sx(t_1), \tag{5.2.2}$$

wobei $x = x(\cdot, t_0, x_0, u)$ die Lösung von (5.2.1) zum Anfangswert $x(t_0) = x_0$ mit Steuerung u ist, $R \in \mathbb{K}^{m \times m}$ und $Q, S \in \mathbb{K}^{n \times n}$ hermitesche und positive semidefinite Matrizen sind und R vollen Rang hat, also positiv definit ist. Der erste Term des Integranden bestraft große Eingangswerte, die typischerweise mit hohen Kosten für die physikalischen Steuerungskomponenten einhergehen, während der zweite Term ein Abweichen des Zustandes von der 0 bestraft. Mit dem Term außerhalb des Integral können wir festlegen wie wichtig es ist, dass der Zustand am Ende des Zeithorizonts $[t_0, t_1]$ Nahe der 0 liegt. Die Matrizen R , Q und S dienen dabei als Gewichtung mit denen wir festlegen, in welchen Komponenten ein Abweichen von der 0 stärker zu bestrafen ist.

Das Optimalsteuerungsproblem lautet nun wie folgt:

Finde auf dem Zeithorizont $[t_0, t_1]$ und zum Anfangswert $x(t_0) = x_0$ eine Kontrollfunktion $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ mit minimalen Kosten $J(t_0, t_1, x_0, u)$.

Führt man die *Werte- oder Bellman-Funktion*

$$V(t_0, t_1, x_0) := \inf \{ J(t_0, t_1, x_0, u) : u \in L^2([t_0, t_1]; \mathbb{K}^m) \}$$

ein so suchen wir also eine Kontrolle $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ mit der Eigenschaft, dass

$$J(t_0, t_1, x_0, u) = V(t_0, t_1, x_0).$$

Eine Kontrollfunktion $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ mit diesen Eigenschaften nennen wir *optimal* für den Anfangswert $x(t_0) = x_0$ und den Zeithorizont $[t_0, t_1]$.

Wir wollen nun zeigen, dass das Optimalsteuerungsproblem immer eine Lösung hat und diese durch ein Feedback gegeben ist. Um das Feedback zu beschreiben betrachten wir das matrixwertige nichtlineare Riccati-Endwertproblem

$$\begin{aligned} \dot{P}(t) &= P(t)BR^{-1}B^*P(t) - P(t)A - A^*P(t) - Q, & t \in [t_0, t_1] \\ P(t_1) &= S, \end{aligned} \tag{5.2.3}$$

das gelöst wird durch eine stetig differenzierbare Funktion $P : [t_0, t_1] \rightarrow \mathbb{K}^{n \times n}$.

Bemerkung 5.2.1. (i) Das Riccati-Endwertproblem (5.2.3) hat eine lokal Lipschitz-stetige rechte Seite. Deshalb folgt aus dem Satz von Picard-Lindelöf, siehe die Bemerkungen nach Satz 1.3.1, die Existenz eines $t'_0 \in [t_0, t_1]$ und einer stetig differenzierbaren Funktion $P : [t'_0, t_1] \rightarrow \mathbb{K}^{n \times n}$, welche (5.2.3) löst. Ferner ist jede solche Lösung eindeutig bestimmt.

(ii) Die Lösung $P(\cdot)$ von (5.2.3) ist eine Funktion mit hermiteschen Werten (Übung).

(iii) Sei $\tilde{t}_0 := \inf \{ t'_0 \in [t_0, t_1] : (5.2.3) \text{ hat eine Lösung auf } [t'_0, t_1] \}$. Dann folgt aus einem Standardresultat über maximale Existenzintervalle von gewöhnlichen Differentialgleichung, dass entweder

$$t_0 = \tilde{t}_0 \quad \text{oder} \quad \lim_{t'_0 \searrow \tilde{t}_0} \|P(t)\| = \infty$$

gelten muss. Im ersten Fall existiert die Lösung von (5.2.3) auf ganz $[t_0, t_1]$. Um zu zeigen, dass dies der Fall ist müssen wir also nachweisen, dass die Norm der Lösung bei t'_0 nicht explodiert.

Im nächsten Theorem zeigen wir, dass das nichtlineare Problem (5.2.3) immer eine Lösung hat und dass wir mit Hilfe dieser die gewünschte Steuerung, welche als Rückkopplung realisiert werden kann, finden können.

Theorem 5.2.2 (Lineare Optimalsteuerung mit quadratischen Kosten). *Seien $R \in \mathbb{K}^{m \times m}$ hermitesch und positiv definit, $Q, S \in \mathbb{K}^{n \times n}$ hermitesch und positiv semidefinite, und $t_0, t_1 \in \mathbb{R}$ mit $t_0 < t_1$. Dann hat das Riccati-Endwertproblem (5.2.3) eine eindeutige stetig differenzierbare Lösung $P : [t_0, t_1] \rightarrow \mathbb{K}^{n \times n}$. Ferner ist $P(t)$ hermitesch und positiv semidefinit für alle $t \in [t_0, t_1]$.*

Ferner existiert für jeden Anfangswert $x_0 \in \mathbb{K}^n$ eine eindeutige optimale Steuerung $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ welche

$$u(t) = -R^{-1}B^*P(t)x(t) \quad (5.2.4)$$

für alle $t \in [t_0, t_1]$ erfüllt, wobei $x = x(\cdot, t_0, x_0, u)$ die Lösung von (5.2.1) zum Anfangswert $x(t_0) = x_0$ und zur optimalen Steuerung u ist. Die Kosten der optimalen Steuerung sind gegeben durch

$$V(t_0, t_1, x_0) = x_0^* P(t_0) x_0.$$

Man bemerke, die optimale Steuerung u hängt insofern von x_0 ab, als dass x_0 den momentanen Zustand $x(t)$ mitbestimmt. Das Theorem sagt ferner, dass die optimale Steuerung als Rückkopplung, mit zeitvarianter Rückkopplungsmatrix, realisiert werden kann.

Beweis von Theorem 5.2.2. Sei $P : [t'_0, t_1] \rightarrow \mathbb{K}^{n \times n}$ die eindeutige Lösung von (5.2.3) für ein passendes $t'_0 \in [t_0, t_1]$. Wir wissen bereits aus Bemerkung 5.2.1, dass ein solches t'_0 existiert und, dass die Werte von P hermitesch sind. Wir zeigen zunächst, dass die Rückkopplung (5.2.4) optimal ist auf dem Intervall $[t'_0, t_1]$. Es sei dazu $x_0 \in \mathbb{K}^n$ ein beliebiger Anfangswert und $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ eine beliebige Kontrollfunktion und bezeichne mit $x(\cdot) = x(\cdot, t_0, x_0, u)$ die Lösung von (5.2.1) zum gegebenen Anfangswert $x(t_0) = x_0$ und der Steuerung u . Wir differenzieren die skalarwertige Funktion $[t'_0, t_1] \ni t \mapsto x(t)^* P(t) x(t)$ und erhalten

$$\begin{aligned} \frac{d}{dt} x(t)^* P(t) x(t) &= x(t)^* \dot{P}(t) x(t) + \dot{x}(t)^* P(t) x(t) + x(t)^* P(t) \dot{x}(t) \\ &= x(t)^* (P(t) B R^{-1} B^* P(t) - P(t) A - A^* P(t) - Q) x(t) \\ &\quad + (x(t)^* A^* + u(t)^* B^*) P(t) x(t) + x(t)^* P(t) (A x(t) + B u(t)) \\ &= (u(t) + R^{-1} B^* P(t) x(t))^* R (u(t) + R^{-1} B^* P(t) x(t)) \\ &\quad - u(t)^* R u(t) - x(t)^* Q x(t). \end{aligned}$$

Dabei bilden die letzten beiden Terme gerade den negativen Integranden des Kostenfunktional (5.2.2). Damit und aus dem Hauptsatz der Differenzial- und Integralrechnung erhalten wir nach umstellen, dass

$$\begin{aligned}
 J(t'_0, t_1, x_0, u) &= -x(t_1)^* \underbrace{P(t_1)}_{=S} x(t_1) + x(t_0)^* P(t_0) x(t_0) + x(t_1)^* S x(t_1) + \\
 &\quad + \int_{t'_0}^{t_1} (u(t) + R^{-1} B^* P(t) x(t))^* R (u(t) + R^{-1} B^* P(t) x(t)) \, dt \\
 &= x_0^* P(t_0) x_0 \\
 &\quad + \int_{t'_0}^{t_1} \underbrace{(u(t) + R^{-1} B^* P(t) x(t))^* R (u(t) + R^{-1} B^* P(t) x(t))}_{\geq 0} \, dt.
 \end{aligned}$$

Da R invertierbar ist, ist somit ist die eindeutige optimale Steuerung auf $[t'_0, t_1]$ gerade die, welche $u(t) = -R^{-1} B^* P(t) x(t)$ für alle $t \in [t'_0, t_1]$ erfüllt. Also gerade die Steuerung ist, welche den Integranden gleich 0 setzt. Dann sind trivialerweise die Kosten der optimalen Steuerung gegeben durch

$$V(t'_0, t_1, x_0) = x_0^* P(t'_0) x_0.$$

Da $V(t'_0, t_1, x_0) \geq 0$ für alle $x_0 \in \mathbb{K}^n$ gilt, folgt, dass $P(t'_0)$ positiv semidefinit ist für alle $t'_0 \in [t_0, t_1]$ für welche die Lösung des Riccati-Endwertproblems auf $[t'_0, t_1]$ existiert (und natürlich auch für $t'_0 = t_1$, da $P(t_1) = S$).

Es bleibt noch zu zeigen, dass die Lösung des Riccati-Endwertproblems (5.2.3) auf ganz $[t_0, t_1]$ existiert und damit im obenstehenden ersten Teil des Beweises $t'_0 = t_0$ gewählt werden kann. Dazu genügt es nach Bemerkung 5.2.1 die Beschränktheit der Lösung auf dem Intervall $[t_0, t_1]$ zu zeigen. Diese folgt direkt aus der Abschätzung

$$\begin{aligned}
 \|P(t'_0)\| &= \sup_{\|x_0\|=1} x_0^* P(t'_0) x_0 \\
 &\leq \sup_{\|x_0\|=1} J(t'_0, t_1, x_0, 0) \\
 &= \sup_{\|x_0\|=1} \left(\int_{t'_0}^{t_1} x_0^* e^{(t-t'_0)A^*} Q e^{(t-t'_0)A} x_0 \, dt + x_0^* e^{(t_1-t'_0)A^*} S e^{2(t_1-t'_0)A} x_0 \right) \\
 &\leq \|Q\| \int_{t'_0}^{t_1} e^{2(t-t'_0)\|A\|} \, dt + \|S\| e^{2(t_1-t'_0)\|A\|} \|x_0\| \\
 &\leq e^{2(t_1-t_0)\|A\|} (\|Q\| (t_1 - t_0) + \|S\|).
 \end{aligned}$$

Diese obere Schranke ist endlich und von t'_0 unabhängig. Also kann die Lösung P auf $[t_0, t_1]$ nicht explodieren. Es folgt, dass (5.2.3) auf ganz $[t_0, t_1]$ eine eindeutige Lösung hat. $\square$

Bemerkung 5.2.3. Wir bemerken, dass wir das Lineare Optimalsteuerungsproblem mit quadratischen Kosten auch noch in allgemeiner Form eine eindeutige Lösung

hat. So können wir die Koeffizienten des Systems A und B durch messbare und beschränkte Funktionen $A : [t_0, t_1] \rightarrow \mathbb{K}^{n \times n}$ und $B : [t_0, t_1] \rightarrow \mathbb{K}^{n \times m}$ ersetzen, sodass wir als neues System das zeitvariante lineare System

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \quad \text{für alle } t \in [t_0, t_1]$$

bekommen. Diese hat aufgrund der Beschränktheit der Koeffizienten eine Lösung. Man beachte nur, dass von der Lösung P des Riccati-Endwertproblems dann nicht mehr erwartet werden kann, dass diese stetig differenzierbar ist. Sie ist jedoch immer absolut stetig.

Eine Zeitabhängigkeit können wir auch bei den Gewichten R , Q und S zulassen. Diese wollen wir dann als messbare beschränkte Funktionen auf $[t_0, t_1]$ annehmen. Dies wird beim Tracking-Problem eine Rolle spielen.

Anwendung: Das Tracking-Problem

Wir wenden die Lösung des lineare Optimalsteuerungsproblem nun auf ein den Ingenieuren bekanntes Problem an: dem Tracking-Problem. Zugrunde liegt uns das lineare System mit Ein- und Ausgang

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), & x(t_0) &= x_0 \in \mathbb{K}^n, \\ y(t) &= Cx(t), \end{aligned} \tag{5.2.5}$$

welches wir auf für die Zeiten $t \in [t_0, t_1]$ für ein fest gewähltes $t_1 > t_0$ betrachten. Hier sind gemäß Konvention $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$ und $C \in \mathbb{K}^{p \times n}$. Ferner haben wir ein Referenzsignal $r : [t_0, t_1] \rightarrow \mathbb{K}^p$ als messbare und beschränkte Funktion gegeben. Ziel ist es eine Kontrolle so zu finden, dass der Ausgang $y(t)$ des Systems bestmöglich dem Referenzsignal gleicht. Um die Güte der Annäherung an das Referenzsignal zu quantifizieren, definieren wir für jede Steuerung $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ das Kostenfunktional

$$\begin{aligned} J(t_0, t_1, x_0, u) &:= \int_{t_0}^{t_1} u(t)^* R u(t) + (y(t) - r(t))^* Q (y(t) - r(t)) \, dt \\ &\quad + x(t_1)^* S x(t_1), \end{aligned} \tag{5.2.6}$$

welches Abweichungen vom Referenzsignal mit Hilfe der Gewichte Q und S bestraft. Hier ist $y(\cdot) = y(\cdot, t_0, x_0, u)$ der Ausgang des Anfangswertproblems (5.2.5) und $x(\cdot) = x(\cdot, t_0, x_0, u)$ die Lösung. Wir wollen erneut annehmen, dass $R \in \mathbb{K}^{m \times m}$ hermitesch und positiv definit und $Q, S \in \mathbb{K}^{n \times n}$ hermitesch und positiv semidefinit sind.

Das Tracking-Problem lautet dann wie folgt:

Finde auf dem Zeithorizont $[t_0, t_1]$, zum Anfangswert x_0 und zur Referenz r eine Kontrollfunktion $u \in L^2([t_0, t_1]; \mathbb{K}^m)$ mit minimalen Kosten $J(t_0, t_1, x_0, u)$.

Wir liefern eine positive Antwort für das Tracking-Problem indem wir diese auf ein lineares Optimalsteuerungsproblem mit quadratischen Kosten zurückführen. Dazu definieren wir als System-Koeffizienten die $\mathbb{K}^{(n+1) \times (n+1)}$ - bzw. $\mathbb{K}^{(n+1) \times m}$ -Matrizen

$$\tilde{A} := \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}, \quad \tilde{B} := \begin{pmatrix} B \\ 0 \end{pmatrix}.$$

Dann erhalten wir ein System

$$\dot{\tilde{x}}(t) = \tilde{A}\tilde{x}(t) + \tilde{B}u(t) \quad \text{für alle } t \in [t_0, t_1]. \quad (5.2.7)$$

Ist nun $x(\cdot, t_0, x_0, u)$ eine Lösung von (5.2.5) zur Kontrolle u und $\tilde{x}(\cdot, t_0, x_0, u)$ eine Lösung von (5.2.7) zum Anfangswert $\tilde{x}(t_0) = \tilde{x}_0 := \begin{pmatrix} x_0 \\ 1 \end{pmatrix}$, so gilt

$$\tilde{x}(\cdot, t_0, \tilde{x}_0, u) = \begin{pmatrix} x(\cdot, t_0, x_0, u) \\ 1 \end{pmatrix}.$$

Als Koeffizienten des Kostenfunktional $\tilde{J}(t_0, t_1, \tilde{x}_0, u)$, definiert analog zu (5.2.2), wählen wir

$$\tilde{R} := R, \quad \tilde{Q}(t) := \begin{pmatrix} C^*QC & -C^*Qr(t) \\ -r(t)^*QC & r(t)^*Qr(t) \end{pmatrix}, \quad \tilde{S} := \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}$$

für alle $t \in [t_0, t_1]$. Es ist einfach nachzuprüfen, dass $\tilde{Q}$ hermitesche und positiv semidefinite Werte hat.³ Mit dieser Wahl erhalten wir, dass

$$J(t_0, t_1, x_0, u) = \tilde{J}\left(t_0, t_1, \begin{pmatrix} x_0 \\ 1 \end{pmatrix}, u\right).$$

Nach Theorem 5.2.2 existiert nun die eindeutige absolut stetige Lösung $P : [t_0, t_1] \rightarrow \mathbb{K}^{(n+1) \times (n+1)}$ des Riccati-Endwertproblems

$$\begin{aligned} \dot{\tilde{P}}(t) &= \tilde{P}(t)\tilde{B}\tilde{R}^{-1}\tilde{B}^*\tilde{P}(t) - \tilde{P}(t)\tilde{A} - \tilde{A}^*\tilde{P}(t) - \tilde{Q}(t), \quad t \in [t_0, t_1] \\ \tilde{P}(t_1) &= \tilde{S}. \end{aligned} \quad (5.2.8)$$

Wir zerlegen nun die hermitesche Lösung des Endwertproblems als

$$\tilde{P}(t) = \begin{pmatrix} P(t) & \beta(t) \\ \beta(t)^* & \alpha(t) \end{pmatrix}$$

mit $P : [t_0, t_1] \rightarrow \mathbb{K}^{n \times n}$. Eine einfache Rechnung zeigt, dass dann

$$\begin{pmatrix} \dot{P}(t) & \dot{\beta}(t) \\ \dot{\beta}(t)^* & \dot{\alpha}(t) \end{pmatrix} = \begin{pmatrix} P(t) & \beta(t) \\ \beta(t)^* & \alpha(t) \end{pmatrix} \begin{pmatrix} BR^{-1}B^* & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} P(t) & \beta(t) \\ \beta(t)^* & \alpha(t) \end{pmatrix}$$

³Wir erinnern, dass wir nach Bemerkung 5.2.3 auch zeitvariante Koeffizienten als Gewichte des Kostenfunktional erlauben können, solange sie messbar und beschränkt sind. Dies ist hier der Fall.

$$\begin{aligned}
 & - \begin{pmatrix} P(t)A & 0 \\ \beta(t)^*A & 0 \end{pmatrix} - \begin{pmatrix} A^*P(t) & A^*\beta(t) \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} C^*QC^* & -C^*Qr(t) \\ -r(t)^*QC & r(t)^*Qr(t) \end{pmatrix}, \\
 & \begin{pmatrix} P(t_1) & \beta(t_1) \\ \beta(t_1)^* & \alpha(t_1) \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}
 \end{aligned}$$

gilt. Also ist P gerade die Lösung des Riccati-Endwertproblems

$$\begin{aligned}
 \dot{P}(t) &= P(t)BR^{-1}B^*P(t) - P(t)A - A^*P(t) - Q, \quad t \in [t_0, t_1] \\
 P(t_1) &= S
 \end{aligned}$$

und damit unabhängig von r . Die Funktionen α und β sind ferner gegeben als Lösungen des entkoppelten Anfangswertproblems

$$\begin{aligned}
 \dot{\beta}(t) &= -(A - BR^{-1}B^*P(t))^* + C^*Qr(t) \\
 \dot{\alpha}(t) &= \beta(t)^*BR^{-1}B^*\beta(t) - r(t)^*Qr(t) \\
 (\beta(t_0), \alpha(t_0)) &= (0, 0).
 \end{aligned} \tag{5.2.9}$$

Die optimale Kontrolle erfüllt somit

$$\begin{aligned}
 u(t) &= -\tilde{R}^{-1}\tilde{B}^*\tilde{P}(t)\tilde{x}(t) \\
 &= -R^{-1} \begin{pmatrix} B^* & 0 \end{pmatrix} \begin{pmatrix} P(t) & \beta(t) \\ \beta(t)^* & \alpha(t) \end{pmatrix} \begin{pmatrix} x(t) \\ 1 \end{pmatrix} \\
 &= -R^{-1}B^*P(t)x(t) - R^{-1}B^*\beta(t)
 \end{aligned}$$

für alle $t \in [t_0, t_1]$.

Wir halten fest.

Theorem 5.2.4 (Tracking-Problem). *Seien $R \in \mathbb{K}^{m \times m}$ hermitesch und positiv definit, $Q, S \in \mathbb{K}^{n \times n}$ hermitesch und positiv semidefinit, und $t_0, t_1 \in \mathbb{R}$ mit $t_0 < t_1$. Dann existiert für jeden Anfangswert $x_0 \in \mathbb{K}^n$ eine eindeutige Steuerung $u \in L^2([t_0, t_1]; \mathbb{K}^m)$, welche das Kostenfunktional (5.2.6) des Tracking-Problems minimiert. Diese Kontrolle erfüllt*

$$u(t) = -R^{-1}B^*P(t)x(t) - R^{-1}B^*\beta(t)$$

für alle $t \in [t_0, t_1]$, wobei $x(\cdot) = x(\cdot, t_0, x_0, u)$ die Lösung von (5.2.1) zum Anfangswert $x(t_0) = x_0$ und zur kostenminimierenden Steuerung u ist, P die Lösung des Riccati-Endwertproblems (5.2.8) und β die Lösung des Anfangswertproblems (5.2.9) ist.

Die Kosten der kostenminimierenden Steuerung u sind gegeben durch

$$J(t_0, t_1, x_0, u) = x_0^*P(t_0)x_0 + 2x_0^*\beta(t_0) - \int_{t_0}^{t_1} \beta(t)^*BR^{-1}B^*\beta(t) - r(t)^*Qr(t) dt.$$

Beweis. Es ist lediglich da für die Formel für die minimalen Kosten der kostenminimierenden Steuerung zu zeigen. Dies stellen wir als Übungsaufgabe. $\square$

Kapitel 6

Lineare zeitinvariante Systeme

6.1 Die Laplace-Transformation

Wir verwenden für jedes $\alpha \in [-\infty, \infty)$ die Notationen

$$\mathbb{C}_{>\alpha} := \{\lambda \in \mathbb{C} \mid \operatorname{Re} \lambda > \alpha\}$$

und

$$\begin{aligned} L_{\alpha}^1([0, \infty); \mathbb{K}^{p \times m}) \\ := \{f \in L_{\text{loc}}^1([0, \infty); \mathbb{K}^{p \times m}) \mid e^{-r \cdot} f \in L^1([0, \infty); \mathbb{K}^{p \times m}) \text{ für alle } r > \alpha\}. \end{aligned}$$

Beachten Sie, dass diese Räume größer werden, wenn α größer wird.

Definition 6.1.1 (Laplace-Transformation). Sei $\alpha \in [-\infty, \infty)$. Für eine Funktion $f \in L_{\alpha}^1([0, \infty); \mathbb{K}^{p \times m})$ heißt die Funktion $\mathcal{L}f: \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$, die durch

$$(\mathcal{L}f)(\lambda) = \int_0^{\infty} e^{-\lambda t} f(t) dt$$

für alle $\lambda \in \mathbb{C}_{>\alpha}$ definiert ist, heißt die **Laplace-Transformierte** von f .

Die lineare Abbildung $\mathcal{L}$, die f auf $\mathcal{L}f$ schickt, nennt man die **Laplace-Transformation**.

Die folgenden Eigenschaften der Laplace-Transformation sind sehr nützlich. Wir werden Sie im Rahmen dieser Vorlesung nicht beweisen, sie aber häufig verwenden.

Satz 6.1.2 (Eigenschaften der Laplace-Transformation). Sei $\alpha \in [-\infty, \infty)$ und $f \in L_{\alpha}^1([0, \infty); \mathbb{K}^{p \times m})$.

- (a) Es gilt $\mathcal{L}f(\lambda) \rightarrow 0$ für $\operatorname{Re} \lambda \rightarrow \infty$.
- (b) Falls f stetig differenzierbar ist und seine Ableitung ebenfalls in L_{α}^1 liegt, gilt $\mathcal{L}\dot{f}(\lambda) = \lambda \mathcal{L}f(\lambda) - f(0)$ für alle $\lambda \in \mathbb{C}_{>\alpha}$.
- (c) Die Laplace-Transformierte $\hat{f}$ ist holomorph. Für $g(t) = t^n f(t)$ gilt $\mathcal{L}g(\lambda) = (-1)^n \left(\frac{d}{d\lambda}\right)^n \mathcal{L}f(\lambda)$.

- (d) Für $t_0 \geq 0$ gilt: $\mathcal{L}(f(\cdot - t_0)) = e^{-t_0\lambda} \mathcal{L}f(\lambda)$
- (e) Die Laplace-Transformation bildet die Faltung von zwei Funktionen auf das Produkt ihrer Laplace-Transformierten ab.

Satz 6.1.3 (Injektivität der Laplace-Transformation). Sei $\alpha \in [-\infty, \infty)$ und $f, g \in L^1_\alpha([0, \infty); \mathbb{K}^{p \times m})$. Wenn $(\mathcal{L}f)(\lambda) = (\mathcal{L}g)(\lambda)$ für alle $\lambda \in \mathbb{C}_{>\alpha}$ gilt, dann gilt $f = g$ fast überall.

6.2 Zeitinvarianz und Impulsantwort

Definition 6.2.1 (Konvergenz auf L^1_{loc}). Seien $f, f_n \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times m})$ für Indizes $n \in \mathbb{N}_0$. Wir sagen f_n **konvergiert in L^1_{loc}** gegen f , falls wir jede Zeit $t_1 > 0$ die Bedingung $\int_{[0, t_1]} |f_n(t) - f(t)| dt \rightarrow 0$ gilt.

Lemma 6.2.2 (Folgenstetigkeit der Faltung). Seien $f, f_n \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{m \times \ell})$ und $g, g_n \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times m})$ für $n \in \mathbb{N}$. Wenn $f_n \rightarrow f$ und $g_n \rightarrow g$ in L^1_{loc} gilt, dann gilt auch $g_n \star f_n \rightarrow g \star f$ in L^1_{loc} .

Beweis. Wir fixieren $t_1 > 0$. Wir verwenden die Notation $\tilde{f} := \mathbb{1}_{[0, t_1]} f$ sowie $\tilde{f}_n := \mathbb{1}_{[0, t_1]} f_n$ für alle n , und ebenso für g und g_n . Dann sind $\tilde{f}, \tilde{f}_n, \tilde{g}, \tilde{g}_n$ alle integrierbar und es gilt $f_n \rightarrow f$ und $g_n \rightarrow g$ in L^1 . Außerdem ist

$$\left\| \tilde{g}_n \star \tilde{f}_n - \tilde{g} \star \tilde{f} \right\|_1 \leq \left\| \tilde{g}_n - \tilde{g} \right\|_1 \left\| \tilde{f}_n \right\|_1 + \left\| \tilde{g} \right\|_1 \left\| \tilde{f}_n - \tilde{f} \right\|_1 \rightarrow 0$$

für $n \rightarrow \infty$. Auf dem Zeitintervall $[0, t_1]$ stimmt $\tilde{g}_n \star \tilde{f}_n$ aber mit $g_n \star f_n$ überein und $\tilde{g} \star \tilde{f}$ stimmt dort mit $g \star f$ überein. Also folgt

$$\left\| \mathbb{1}_{[0, t_1]} (g_n \star f_n - g \star f) \right\|_1 = \left\| \mathbb{1}_{[0, t_1]} (\tilde{g}_n \star \tilde{f}_n - \tilde{g} \star \tilde{f}) \right\|_1 \leq \left\| \tilde{g}_n \star \tilde{f}_n - \tilde{g} \star \tilde{f} \right\|_1 \rightarrow 0$$

für $n \rightarrow \infty$. □

Definition 6.2.3 (Lineares zeitinvariantes System). Ein **lineares zeitinvariantes System** (englisch **linear time invariant system**, kurz **LTI system**) ist eine Abbildung $T : L^1_{\text{loc}}([0, \infty), \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$ mit folgenden Eigenschaften:

- (I) T ist linear.
- (II) T ist folgenstetig, das heißt es schickt konvergente Folgen auf konvergente Folgen und vertauscht mit Grenzwerten.
- (III) T vertauscht für jedes $t \geq 0$ mit den Rechtsshift-Operatoren R_t , die jede Funktion um t Zeiteinheiten nach rechts verschiebt und von links die 0 nachschiebt.

Beispiel 6.2.4 (Rechtsshift als lineares zeitinvariantes System). Für jedes $t \geq 0$ ist der Rechtsshift R_t aus der vorangehenden Definition selbst ein lineares zeitinvariantes System.

Beispiel 6.2.5 (Faltung mit einem L^1_{loc} -Kern). Sei $k \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times n})$. Dann ist der Operator

$$\begin{aligned} T : L^1_{\text{loc}}([0, \infty), \mathbb{K}^m) &\rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p) \\ u &\mapsto k \star u, \end{aligned}$$

der mit k faltet,¹ ein lineares zeitinvariantes System, wie man leicht nachrechnet. Wir nennen k den **Faltungskern von T** .

Beweis. Die Linearität ist klar und das Kommutieren mit der Translation lässt sich leicht nachrechnen. Die L^1_{loc} -Folgenstetigkeit folgt aus Lemma 6.2.2. $\square$

Beispiel 6.2.6 (Lineare Differentialgleichungen erster Ordnung). Folgender Spezialfall von Beispiel 6.2.5 ist im Hinblick auf die vorangehenden Vorlesungskapitel interessant: Betrachten Sie das System

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

mit Startwert $x(0) = 0$. Dann für gegebenen Eingang u ist der der Ausgang y gegeben durch

$$y(t) = \int_0^t C e^{(t-s)A} B u(s) \, ds$$

für alle $t \in [0, \infty)$.

Die Abbildung $T : L^1_{\text{loc}}([0, \infty); \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$, die u auf y schickt, ist als eine Faltung mit dem Faltungskern k , der durch $k(t) = C e^{tA} B$ für $t \in [0, \infty)$ gegeben ist.

Definition 6.2.7 (Approximative Eins). Wir nennen eine Folge (f_n) in $L^1([0, \infty); \mathbb{K})$ eine **approximative Eins**, falls sie die folgenden Eigenschaften erfüllt:

- (I) Jedes f_n nimmt nur Werte ≥ 0 an.
- (II) Es gilt $\int_{[0, \infty)} f_n(t) \, dt = 1$ für jedes n .
- (III) Für jedes noch so kleine Zeit $t_0 > 0$ gilt $\int_{[t_0, \infty)} f_n(t) \, dt \rightarrow 0$ für $n \rightarrow \infty$.²

Wir nennen eine Folge (f_n) in $L^1([0, \infty); \mathbb{K}^{m \times m})$ eine **approximative Eins**, wenn jedes f_n in die Diagonalmatrizen abbildet und die Folge $((f_n)_{jj})$ in $L^1([0, \infty); \mathbb{K})$ für jedes $j \in \{1, \dots, m\}$ eine approximative Eins ist.

Proposition 6.2.8 (Faltung mit einer approximativen Eins).

¹Wir erinnern an die Definition $(k \star u)(t) := \int_0^t k(s)u(t-s) \, ds = \int_0^t k(t-s)u(s) \, ds$ für $t \in [0, \infty)$. Beachten Sie, dass die Faltung im matrix- und vektorwertigen Fall im Gegensatz zum skalarwertigen Fall nicht kommutativ ist (einfach, weil die Matrixmultiplikation nicht kommutativ ist).

²Das heißt, die Masse von f_n konzentriert sich für große n nahe bei 0.

- (a) Sei (f_n) in $L^1([0, \infty); \mathbb{K})$ eine approximative Eins und $u \in L^1_{\text{loc}}([0, \infty); \mathbb{K})$. Dann gilt $f_n \star u \rightarrow u$ in $L^1_{\text{loc}}([0, \infty); \mathbb{K})$ für $n \rightarrow \infty$.
- (b) Sei (f_n) in $L^1([0, \infty); \mathbb{K}^{m \times m})$ eine approximative Eins. Seien außerdem $u \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{m \times \ell})$ und $\tilde{u} \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times m})$. Dann gilt $f_n \star u \rightarrow u$ und $\tilde{u} \star f_n \rightarrow \tilde{u}$ in L^1_{loc} für $n \rightarrow \infty$.

Beweis. (a) Wir zeigen die Behauptung zunächst unter den Zusatzannahmen, dass u sogar in $L^1([0, \infty); \mathbb{K})$ liegt, stetig ist, in 0 verschwindet und beschränkten Träger hat. Dann können wir u zu einer stetigen Funktion auf $\mathbb{R}$ fortsetzen, die auf $(-\infty, 0]$ verschwindet und in $L^1(\mathbb{R}; \mathbb{K})$ liegt; wir bezeichnen diese Funktion ebenfalls mit u . Ebenso setzen wir die Funktionen f_n links der 0 durch 0 fort und erhalten somit Funktionen in $L^1(\mathbb{R}; \mathbb{K})$, die wir ebenfalls mit f_n bezeichnen.

Sei $\varepsilon > 0$. Da u auf $\mathbb{R}$ stetig ist und dort kompakten Träger hat, ist u sogar gleichmäßig stetig, das heißt wir finden ein $\delta > 0$ derart, dass $|u(s_2) - u(s_1)| \leq \varepsilon$ für alle $s_1, s_2 \in \mathbb{R}$ mit $|s_2 - s_1| \leq \delta$ gilt.

Wähle nun einen Index n_0 derart, dass $\int_{|t| \geq \delta} f_n(t) dt \leq \varepsilon$ für alle $n \geq n_0$ gilt. Für jedes n erhält man mit Hilfe der Eigenschaften $f_n \geq 0$ und $\int_{\mathbb{R}} f_n(s) ds = 1$ und mit dem Satz von Fubini–Tonelli

$$\begin{aligned} \|f_n \star u - u\|_1 &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f_n(s) u(t-s) ds - \int_{\mathbb{R}} f_n(s) u(t) ds \right| dt \\ &\leq \int_{\mathbb{R}} f_n(s) \int_{\mathbb{R}} |u(t-s) - u(t)| dt ds. \end{aligned}$$

Wir teilen das äußere Integral auf in ein Integral über den $[-\delta, \delta]$ und ein Integral über den Rest von $\mathbb{R}$ und erhalten somit

$$\|f_n \star u - u\|_1 \leq \|f_n\|_1 \varepsilon + 2\|u\|_{\infty} \varepsilon = (1 + 2\|u\|_{\infty}) \varepsilon.$$

Also gilt $f_n \star u \rightarrow u$ sogar in L^1 -norm.

Weil die Funktionen u mit den bisher gemachten Zusatzannahmen bezüglich der L^1 -Norm dicht in L^1 liegen, folgt aus der Youngschen Ungleichung $\|f \star g\|_1 \leq \|f\|_1 \|g\|_1$, die für alle $f, g \in L^1([0, \infty); \mathbb{K})$ gilt, mit eine 3ε -Argument, dass $f_n \star u \rightarrow u$ in L^1 -Norm sogar für alle $u \in L^1([0, \infty); \mathbb{K})$ gilt.

Zuletzt betrachten wir eine allgemeine Funktion $u \in L^1_{\text{loc}}([0, \infty); \mathbb{K})$ und fixieren eine Zeit $t_1 \in (0, \infty)$. Setze $\tilde{u} := \mathbb{1}_{[0, t_1]} u$. Ist $\tilde{u}$ in L^1 und $f_n \star \tilde{u}$ stimmt auf $[0, t_1]$ mit $f_n \star u$ überein. Somit ist

$$\int_{[0, \infty)} |f_n \star u - u|(t) dt = \int_{[0, \infty)} |f_n \star \tilde{u} - \tilde{u}|(t) dt \leq \|f_n \star \tilde{u} - \tilde{u}\|_1 \rightarrow 0$$

für $n \rightarrow \infty$.

(b) Das folgt aus (a) durch komponentenweise Betrachtung. $\square$

Wir möchten approximative Einsen nun benutzen um zu untersuchen, wann lineare zeitinvariante Systeme durch Faltung mit einem L^1_{loc} -Faltungskern gegeben sind. Dazu ist der folgende Integralbegriff sehr hilfreich:

Definition 6.2.9 (L^1_{loc} -wertige Integrale). Sei I ein reelles Intervall, sei $F : I \rightarrow L^1_{\text{loc}}([0; \infty); \mathbb{K}^{p \times m})$ folgenstetig und sei $f \in L^1_{\text{loc}}([0; \infty); \mathbb{K}^{p \times m})$. Wir schreiben

$$f = \int_I F(s) \, ds,$$

falls für jedes $t_1 \in (0, \infty)$ das uneigentliche $L^1([0, t_1]; \mathbb{K}^{p \times m})$ -wertige Riemann-Integral $\int_I F(s)|_{[0, t_1]} \, ds$ existiert³ und gleich $f|_{[0, t_1]}$ ist.

Beachten Sie: Wenn F sogar stetig von I nach $L^1([0, \infty); \mathbb{K}^{p \times m})$ abbildet und das uneigentliche L^1 -wertige Riemann-Integral $f := \int_I F(s) \, ds$ existiert, so gilt automatisch auch $f = \int_I F(s) \, ds$ im Sinne von Definition 6.2.9.

Beispiel 6.2.10 (Die Faltung als vektorwertiges Integral). Sei $u \in L^1_{\text{loc}}([0, \infty); \mathbb{K})$ und sei $v \in L^1([0, \infty))$ stetig. Für jedes $s \in [0, \infty)$ sei $R_s : L^1_{\text{loc}}([0, \infty); \mathbb{K}) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K})$ der Rechtshift um s Zeiteinheiten (der von links die 0 nachschiebt). Dann gilt

$$u \star v = \int_{[0, \infty)} (R_s u) v(s) \, ds$$

im Sinne von Definition 6.2.9. Falls u sogar $L^1([0, \infty); \mathbb{K})$ -wertig ist, ist die rechte Seite sogar ein uneigentliches Riemann-Integral in $L^1([0, \infty); \mathbb{K})$.

Lemma 6.2.11 (L^1_{loc} -wertige Integrale und lineare Operatoren). Sei I eine reelles Intervall, sei $F : I \rightarrow L^1([0; \infty); \mathbb{K})$ stetig, sei $f \in L^1_{\text{loc}}([0; \infty); \mathbb{K})$ und gelte $f = \int_I F(s) \, ds$ als uneigentliches L^1 -wertiges Riemann-Integral.

Wenn $T : L^1_{\text{loc}}([0; \infty); \mathbb{K}^{p \times m}) \rightarrow L^1_{\text{loc}}([0; \infty); \mathbb{K}^{\ell \times k})$ linear und folgenstetig ist, dann gilt $Tf = \int_I TF(s) \, ds$ im Sinne von Definition 6.2.9.

Beweis. Mit Hilfe von Definition 6.2.9 kann man die Aussage direkt auf die analoge Aussage für Banachraumwertige uneigentliche Riemann-Integrale und stetige lineare Operatoren zwischen Banachräumen zurückführen. $\square$

Mit dem soeben besprochenen Konzept von Integraloperatoren können wir in Proposition 6.2.13 eine nützliche Eigenschaft von linearen zeitinvarianten System zeigen. Zuerst vereinbaren wir noch, wie wir solche System auf matrixwertige Eingänge anwenden wollen:

Definition 6.2.12 (Anwendung von LTI-Systemen auf matrixwertige Funktionen). Sei $T : L^1_{\text{loc}}([0, \infty), \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$ ein lineares zeitinvariantes System. Für jede Funktion $u \in L^1_{\text{loc}}(\mathbb{K}^{m \times \ell})$ definieren wir $Tu \in L^1_{\text{loc}}(\mathbb{K}^{p \times \ell})$ spaltenweise.

Proposition 6.2.13 (Assoziativität von LTI-Systemen mit Faltungen). Wir betrachten ein lineares zeitinvariantes System $T : L^1_{\text{loc}}([0, \infty), \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$. Für alle $u \in L^1_{\text{loc}}(\mathbb{K}^{m \times \ell})$ und $v \in L^1_{\text{loc}}(\mathbb{K}^{\ell \times k})$ gilt $T(u \star v) = (Tu) \star v$.

³Beachten Sie, dass die Abbildung $s \mapsto F(s)|_{[0, t_1]}$ von I in den Banachraum $L^1([0, t_1]; \mathbb{K}^{p \times m})$ folgenstetig und somit stetig ist.

Beweis. Durch Betrachten der Matrixkomponenten sehen wir, dass es genügt, den Fall $m = p = \ell = 1$ zu betrachten. Wegen der L^1_{loc} -Folgenstetigkeit von T und von Faltungen (Lemma 6.2.2) genügt es die Aussage unter der Annahme zu zeigen, dass $u, v \in L^1([0, \infty); \mathbb{K})$ gilt und dass v sogar stetig ist.

Wie in Definition 6.2.3 bezeichnen wir für jedes $s \in [0, \infty)$ mit R_s den Rechtshift um s Zeiteinheiten auf $L^1_{\text{loc}}([0, \infty); \mathbb{R})$. Dann gilt laut Beispiel 6.2.10

$$u \star v = \int_{[0, \infty)} (R_s u) v(s) \, ds$$

als uneigentliches $L^1([0, \infty); \mathbb{K})$ -wertiges Riemann-Integral und

$$(Tu) \star v = \int_{[0, \infty)} (R_s Tu) v(s) \, ds = \int_{[0, \infty)} T(v(s) R_s u) \, ds$$

im Sinne von Definition 6.2.9, wobei die letzte Gleichung daraus folgt, dass T mit den Rechtshifts R_s kommutiert.

Aufgrund von Lemma 6.2.11 folgt also tatsächlich $T(u \star v) = (Tu) \star v$. $\square$

Als Konsequenz der vorangehenden Proposition erhält man die folgende Eigenschaft (die man aber auch direkt beweisen kann):

Korollar 6.2.14 (Faltung auf L^1_{loc} ist assoziativ). *Seien $k \in L^1_{\text{loc}}(\mathbb{K}^{p \times m})$, $u \in L^1_{\text{loc}}(\mathbb{K}^{m \times \ell})$ und $v \in L^1_{\text{loc}}(\mathbb{K}^{\ell \times k})$. Dann gilt*

$$(k \star u) \star v = k \star (u \star v).$$

Beweis. Wende Proposition 6.2.13 auf das lineare zeitinvariante System T an, das durch Faltung mit k gegeben ist. $\square$

Satz 6.2.15 (Antworten auf approximative Einsen und Darstellung für LTI-Systeme). *Wir betrachten ein lineares zeitinvariantes System $T : L^1_{\text{loc}}([0, \infty), \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$. Sei (f_n) eine approximative Eins in $L^1_{\text{loc}}([0, \infty); \mathbb{K}^{m \times m})$. Wenn (Tf_n) in $L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times m})$ gegen ein $k \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times m})$ konvergiert, dann gilt:*

- (a) *Für jedes $u \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^m)$ gilt $Tu = k \star u$.*
- (b) *Für jedes anderes approximative Eins (g_n) in $L^1_{\text{loc}}([0, \infty); \mathbb{K}^{m \times m})$ konvergiert auch (Tg_n) in L^1_{loc} gegen k .*

Beweis. (a) Für jeden Index n gilt laut Proposition 6.2.13

$$(Tf_n) \star u = T(f_n \star u).$$

Für $n \rightarrow \infty$ konvergiert die linke Seite in L^1_{loc} wegen der L^1_{loc} -Stetigkeit der Faltung gegen $k \star u$ und die rechte Seite wegen der L^1_{loc} -Stetigkeit von T gegen Tu , also ist $k \star u = Tu$.

(b) Wenn wir (a) auf $u = g_n$ anwenden, erhalten wir

$$Tg_n = k \star g_n$$

und die rechte Seite konvergiert für $n \rightarrow \infty$ in L^1_{loc} gegen k . $\square$

Aufgrund von Teil (b) des vorangehenden Satzes ist der folgende Begriff wohldefiniert.

Definition 6.2.16 (L^1_{loc} -Impulsantwort). Betrachten Sie ein lineares zeitinvariantes System $T : L^1_{\text{loc}}([0, \infty), \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$. Falls für eine (äquivalent: für jedes) approximative Eins (f_n) in $L^1_{\text{loc}}([0, \infty); \mathbb{K}^{m \times m})$ die Folge (Tf_n) in L^1_{loc} gegen ein $k \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times n})$ konvergiert, dann sagen wir **T besitzt eine L^1_{loc} -Impulsantwort** und wir nennen k die **Impulsantwort** von T .

6.3 Transferfunktionen

Wie zuvor verwenden wir für jedes $\alpha \in [-\infty, \infty)$ die Notation

$$\begin{aligned} L^1_\alpha([0, \infty); \mathbb{K}^{p \times m}) \\ := \{f \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times m}) \mid e^{-r \cdot} f \in L^1([0, \infty); \mathbb{K}^{p \times m}) \text{ für alle } r > \alpha\}. \end{aligned}$$

Beachten Sie erneut, dass diese Räume größer werden, wenn α größer wird.

Für die Faltung einer L^1_α -Funktion mit einer L^1_β -Funktion liegt in $L^1_{\max\{\alpha, \beta\}}$, wie man leicht nachrechnet. Deshalb gilt: Wenn T ein lineares zeitinvariantes System ist, das eine L^1_{loc} -Impulsantwort besitzt, die in L^1_α liegt, dann bildet T für jedes $\beta \in \mathbb{R}$ den Raum L^1_β nach $L^1_{\max\{\alpha, \beta\}}$ ab. Das motiviert die folgende Definition:

Definition 6.3.1 (Exponentiell beschränktes LTI-Systeme). Eine lineares zeitinvariantes System $T : L^1_{\text{loc}}([0, \infty); \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$ heißt **exponentiell beschränkt vom Typ $\alpha \in \mathbb{R}$** , wenn für jedes $\beta \in \mathbb{R}$ die Inklusion

$$TL^1_\beta([0, \infty); \mathbb{K}^m) \subseteq L^1_{\max\{\alpha, \beta\}}([0, \infty); \mathbb{K}^p)$$

gilt.

Für $\alpha \in [-\infty, \infty)$ setzen wir

$$\mathbb{C}_{>\alpha} := \{\lambda \in \mathbb{C} \mid \operatorname{Re} \lambda > \alpha\}.$$

Satz 6.3.2 (Existenz einer Transferfunktion). *Betrachten Sie ein lineares zeitinvariantes System $T : L^1_{\text{loc}}([0, \infty); \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$, das exponentiell beschränkt ist vom Typ $\alpha \in [-\infty, \infty)$. Dann gibt es eine holomorphe Funktion $G : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$ mit folgenden Eigenschaften:*

- (a) *Für jedes $\gamma > \alpha$ ist G auf $\mathbb{C}_{>\gamma}$ beschränkt.*
- (b) *Für jedes $\beta \in \mathbb{R}$ und jedes $u \in L^1_\beta([0, \infty); \mathbb{K}^m)$ gilt $(\mathcal{L}Tu)(\lambda) = G(\lambda)(\mathcal{L}u)(\lambda)$ für alle $\lambda \in \mathbb{C}_{>\max\{\alpha, \beta\}}$.*

Beweis. Sei (f_n) eine approximative Eins in $L^1([0, \infty); \mathbb{K}^{m \times m})$ derart, dass jedes f_n Träger in $[0, \frac{1}{n}]$ hat. Dann liegt jedes f_n in L^1_β für jedes $\beta \in \mathbb{R}$ und somit liegt Tf_n in L^1_α für jedes n . Deshalb ist die Laplace-Transformation $\mathcal{L}Tf_n$ für jedes n auf der Halbebene $\mathbb{C}_{>\alpha}$ definiert.

Sei $\gamma > \alpha$ fest. Wir zeigen, dass die Folge $(\mathcal{L}Tf_n)$ auf $\mathbb{C}_{>\gamma}$ in Supremumsnorm beschränkt ist. Sei λ_γ das mit $e^{-\gamma \cdot}$ gewichtete Lebesgue-Maß auf $[0, \infty)$ und analog dazu sei λ_α definiert.

Dann gilt $L^1(d\lambda_\alpha) \subseteq L^1_\alpha \subseteq L^1(d\lambda_\gamma)$, also bildet T den Banachraum $L^1(d\lambda_\alpha)$ in den Banachraum $L^1(d\lambda_\gamma)$ ab. Wegen der L^1_{loc} -Folgenstetigkeit von T und des Satzes vom abgeschlossenen Graphen folgt, dass T von $L^1(d\lambda_\alpha)$ nach $L^1(d\lambda_\gamma)$ stetig ist; seine Operatornorm zwischen diesen Banachräumen bezeichnen wir mit $M \geq 0$.

Da die Funktionen f_n in L^1 normiert sind und Träger in $[0, 1]$ haben, ist die Folge (f_n) in $L^1(d\lambda_\alpha)$ beschränkt, also ist (Tf_n) in $L^1(d\lambda_\gamma)$ beschränkt. Daraus folgt sofort, dass $(\mathcal{L}Tf_n)$, wie behauptet, in Supremumsnorm auf $\mathbb{C}_{>\gamma}$ beschränkt ist.

Somit ist $(\mathcal{L}Tf_n)$ eine Folge holomorpher Funktionen auf $\mathbb{C}_{>\alpha}$, die lokal beschränkt ist. Wegen des Satzes von Montel besitzt diese Folge eine Teilfolge $(\mathcal{L}Tf_{n_j})$, die lokal gleichmäßig gegen eine holomorphe Funktion $G : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$ konvergiert. Natürlich ist auch G für jedes $\gamma > \alpha$ auf $\mathbb{C}_{>\alpha}$ beschränkt.

Wir zeigen nun, dass G auch die Eigenschaft (b) hat. Sei nun $\beta \in \mathbb{R}$ und $u \in L^1_\beta([0, \infty); \mathbb{K}^m)$. Wir können ohne Einschränkung $\beta \geq \alpha$ annehmen. Sei $\lambda \in \mathbb{C}_{>\beta}$ und wähle zwei Zahlen $\gamma_1, \gamma_2 \in \mathbb{R}$ mit $\beta < \gamma_1 < \gamma_2 < \operatorname{Re} \lambda$. Da die Folge (f_{n_j}) eine approximative Eins ist, konvergiert $(f_{n_j} \star u)$ aufgrund des nachfolgenden Lemmas in $L^1(d\lambda_{\gamma_1})$ gegen u . Mit demselben Argument wie zu Anfang des Beweis sehen wir, dass T den Banachraum $L^1(d\lambda_{\gamma_1})$ stetig in den Banachraum $L^1(d\lambda_{\gamma_2})$ abbildet. Somit konvergiert $(T(f_{n_j} \star u))$ in $L^1(d\lambda_{\gamma_2})$ gegen Tu . Darauf folgt $\mathcal{L}(T(f_{n_j} \star u))(\lambda) \rightarrow \mathcal{L}(Tu)(\lambda)$. Andererseits erhalten wir aber aus Proposition 6.2.13 $T(f_{n_j} \star u) = (Tf_{n_j}) \star u$ und somit

$$(\mathcal{L}(Tf_{n_j} \star u))(\lambda) = (\mathcal{L}Tf_{n_j})(\lambda)(\mathcal{L}u)(\lambda) \rightarrow G(\lambda)u(\lambda).$$

Dies zeigt die behauptete Gleichheit $\mathcal{L}(Tu)(\lambda) = G(\lambda)u(\lambda)$. $\square$

Im vorangehenden Beweis haben wir das folgende Hilfsresultat verwendet:

Lemma 6.3.3 (Approximative Eins in gewichteten L^1 -Räumen). *Sei (f_n) eine approximative Eins in $L^1([0, \infty); \mathbb{K}^{m \times m})$ und jedes f_n lebe nur auf $[0, \frac{1}{n}]$. Sei $\gamma \in \mathbb{R}$ und bezeichne λ_γ das mit $e^{-\gamma \cdot}$ gewichtete Lebesgue-Maß auf $[0, \infty)$. Für jedes $u \in L^1_{\text{loc}}([0, \infty); d\lambda_\gamma; \mathbb{K}^m)$ gilt dann $f_n \star u \rightarrow u$ in $L^1_{\text{loc}}([0, \infty); d\lambda_\gamma; \mathbb{K}^m)$.*

Beweis. Es genügt die Behauptung für $m = 1$ zu beweisen.

Setze $\rho_n := \|e^{-\gamma \cdot} f_n\|_1 > 0$ für jeden Index n . Dann ist $(g_n) := (\frac{e^{-\gamma \cdot} f_n}{\rho_n})$ ebenfalls eine approximative Eins. Weil $e^{-\gamma \cdot} u$ in L^1 liegt, gilt, wie wir im Beweis von Proposition 6.2.8 gesehen haben $g_n \star (e^{-\gamma \cdot} u) \rightarrow e^{-\gamma \cdot} u$ in L^1 -Norm. Zugleich gilt für

jeden Index n

$$g_n \star (e^{-\gamma \cdot} u) = \frac{1}{\rho_n} e^{-\gamma \cdot} (f_n \star u).$$

Also konvergiert $\frac{1}{\rho_n}(f_n \star u)$ in $L^1(d\lambda_\gamma)$ -Norm gegen u .

Wir müssen also nur noch zeigen, dass $\rho_n \rightarrow 1$ gilt. Dazu berechnen wir

$$\begin{aligned} |\rho_n - 1| &= \left| \|e^{-\gamma \cdot} f_n\|_1 - \|f_n\|_1 \right| \leq \|e^{-\gamma \cdot} f_n - f_n\|_1 \\ &\leq \|f_n\|_1 \|e^{-\gamma \cdot} - 1\|_{L^\infty([0, \frac{1}{n}])} \rightarrow 0 \end{aligned}$$

für $n \rightarrow \infty$, wobei die letzte Ungleichung daraus folgt, dass f_n nur auf $[0, \frac{1}{n}]$ lebt. $\square$

Proposition 6.3.4 (Eindeutigkeit der Transferfunktion). *Betrachten Sie ein lineares zeitinvariantes System $T : L^1_{\text{loc}}([0, \infty); \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$, das exponentiell beschränkt ist vom Typ $\alpha \in [-\infty, \infty)$.*

Es gibt nur eine Funktion $G : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$, die die Eigenschaft (b) aus Satz (b) erfüllt.

Beweis. Seien $G, \tilde{G}$ zwei solcher Funktionen.

Sei $j \in \{1, \dots, m\}$ und bezeichne $e_j \in \mathbb{K}^m$ den j -ten kanonischen Einheitsvektor und sei $u : [0, \infty) \rightarrow \mathbb{K}^m$ durch $u(t) = e^{t\alpha}$ für alle $t \in [0, \infty)$ definiert. Dann ist $u \in L^1_\alpha([0, \infty); \mathbb{K}^m)$ und $(\mathcal{L}u)(\lambda) = \frac{1}{\lambda - \alpha} e_j$ gilt für alle $\lambda \in \mathbb{C}_{>\alpha}$. Zugleich gilt für jedes solche λ

$$G(\lambda)(\mathcal{L}u)(\lambda) = (\mathcal{L}Tu)(\lambda) = \tilde{G}(\lambda)(\mathcal{L}u)(\lambda).$$

Durch multiplizieren mit $\lambda - \alpha$ sehen wir, dass die j -ten Spalten von $G(\lambda)$ und $\tilde{G}(\lambda)$ übereinstimmen. $\square$

Aufgrund der vorangehenden Proposition ergibt die folgende Definition Sinn.

Definition 6.3.5 (Transferfunktion). *Betrachten Sie ein lineares zeitinvariantes System $T : L^1_{\text{loc}}([0, \infty); \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p)$, das exponentiell beschränkt ist vom Typ $\alpha \in [-\infty, \infty)$.*

*Die (laut Proposition 6.3.4 eindeutig bestimmte) Funktion $G : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$, die die Eigenschaft (b) in Satz 6.3.2 erfüllt, heißt die **Transferfunktion** oder **Übergangsfunktion** von T .*

Beispiel 6.3.6 (Transferfunktion für Systeme mit L^1_{loc} -Impulsantwort). Wenn ein lineares zeitinvariantes System T eine L^1_{loc} -Impulsantwort $k \in L^1_{\text{loc}}([0, \infty); \mathbb{K}^{p \times m})$ besitzt und sogar $k \in L^1_\alpha$ für ein $\alpha \in \mathbb{R}$ gilt, dann ist T exponentiell beschränkt vom Typ α . Die Transferfunktion G von T ist dann gleich der Laplace-Transformierten $\mathcal{L}k$, dann für alle $\beta \in \mathbb{R}$, alle Eingänge $u \in L^1_\beta$ und alle Zahlen $\lambda \in \mathbb{C}_{>\max\{\alpha, \beta\}}$ gilt

$$(\mathcal{L}Tu)(\lambda) = (\mathcal{L}(k \star u))(\lambda) = (\mathcal{L}k)(\lambda)(\mathcal{L}u)(\lambda).$$

Beispiel 6.3.7 (Transferfunktion einer Differentialgleichung). Betrachten Sie das System

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

für Matrizen $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$ und $D \in \mathbb{K}^{p \times m}$ und mit Startwert $x(0) = 0$. Das zugehörige lineare zeitinvariante System mit Startwert ist (siehe Beispiel 6.2.6 – mit dem Unterschied, das wir nun auch noch die Durchgriffsmatrix D in der Gleichung haben) gegeben durch

$$(Tu)(t) = C \int_0^t e^{(t-s)A} Bu(s) ds + Du(t) \quad \text{für } t \in [0, \infty)$$

für alle $u \in L_{\text{loc}}^1([0, \infty); \mathbb{K}^m)$.

Eine kurze Rechnung zeigt, dass T exponentiell beschränkt vom Typ $s(A)$ ist und dass die Transferfunktion G durch

$$G(\lambda) = C(\lambda - A)^{-1}B + D$$

für $\lambda \in \mathbb{C}_{>s(A)}$ gegeben ist.

Aus der Cramer-Regel für Matrixinverse folgt somit, dass (jeder Eintrag von) $G(\lambda)$ eine rationale Funktion bezüglich λ ist. Insbesondere besitzt G eine holomorphe Fortsetzung auf die komplexe Ebene mit nur endlich vielen Definitionslücken; in jedem Eintrag ist der Grad des Zählers kleiner oder gleich dem Grad des Nenners; genau dann hat man in jedem Eintrag eine echte Ungleichung der Grade, wenn $D = 0$ ist. Das folgt ebenfalls aus der Cramerschen Regel.

Man beachte, dass man die Durchgriffsmatrix D aus der Transferfunktion G bestimmen kann: Wegen $(\lambda - A)^{-1} \rightarrow 0$ für $\text{Re } \lambda \rightarrow \infty$ gilt $D = \lim_{\text{Re } \lambda \rightarrow \infty} G(\lambda)$.

Satz 6.3.8 (Polstellen der Transferfunktion und Eigenwerte). *Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$, $D \in \mathbb{K}^{p \times m}$. Es sei (A, B) steuerbar und (A, C) beobachtbar. Dann sind die Eigenwerte von A genau die nicht-hebbaren Pole der (Fortsetzung der) Transferfunktion*

$$\begin{aligned} G: \mathbb{C} \setminus \sigma(A) &\rightarrow \mathbb{C}^{p \times m}, \\ \lambda &\mapsto C(\lambda - A)^{-1}B + D. \end{aligned}$$

Genauer: Die beiden Funktionen G und $\mathcal{R}(\cdot, A) := (\cdot - A)^{-1}$ besitzen dieselben Polstellen und in jedem Pol stimmen ihre Polordnungen überein.

Den Satz sollte man folgendermaßen verstehen: Die Multiplikation der Resolvente $(\cdot - A)^{-1}$ mit C und B für keinen Pol von $(\cdot - A)^{-1}$ die Polordnung verkleinert; insbesondere werden durch Multiplikation mit C und B keine Pole zu hebbaren Singularitäten.

Zum Beweis verwenden wir das folgende Lemma.

Lemma 6.3.9. Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$ und sei $A \neq 0$. Wenn

$$\text{bild}(AB) = \text{bild}(A) \quad \text{und} \quad \text{bild}(A^T C^T) = \text{bild}(A)$$

gilt, dann ist $CAB \neq 0$.

Beweis. Weil A nicht 0 ist und das gleiche Bild wie AB hat, gibt es ein $x \in \mathbb{K}^m$ mit $ABx \neq 0$, also ist $y^T ABx \neq 0$ für ein $y \in \mathbb{K}^n$. Weil A^T dasselbe Bild wie $A^T C^T$ hat, gibt es zudem ein $z \in \mathbb{K}^p$ mit $A^T y = A^T C^T z$, also $y^T A = z^T C A$. Somit ist

$$0 \neq y^T ABx = z^T CABx$$

und folglich $CAB \neq 0$. □

Beweis von Satz 6.3.8. Ist $\lambda \in \mathbb{C}$ eine Pol von G mit Polordnung $p \in \mathbb{N}$, dann ist λ offensichtlich auch ein Pol von $\mathcal{R}(\cdot, A)$ mit Polordnung mindestens p .

Sei also umgekehrt $\lambda \in \mathbb{C}$ ein Pol von $\mathcal{R}(\cdot, A)$ mit Polordnung $p \in \mathbb{N}$. Dann kann man mithilfe von Lemma 6.3.9 zeigen, dass λ auch ein Pol von G mit Polordnung p ist. [Die Details wurden in der Vorlesung besprochen.] □

Wir schließen den Abschnitt mit der folgenden einfachen, aber wichtigen Beobachtung über lineare zeitinvariante Systeme.

Proposition 6.3.10 (Die Transferfunktion bestimmt das LTI-System eindeutig). Seien $T, \tilde{T}: L_{\text{loc}}^1([0, \infty); \mathbb{K}^m) \rightarrow L_{\text{loc}}^1([0, \infty); \mathbb{K}^p)$ lineare zeitinvariante Systeme, die beide exponentiell beschränkt vom Typ $\alpha \in [-\infty, \infty)$ sind. Seien $G, \tilde{G}: \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$ die zugehörigen Transferfunktionen. Wenn $G = \tilde{G}$ ist, dann auch $T = \tilde{T}$.

Beweis. Sei $u \in L_{\alpha}^1([0, \infty); \mathbb{K}^m)$. Dann gilt $\mathcal{L}(Tu) = G\mathcal{L}u = \tilde{G}\mathcal{L}u = \mathcal{L}(\tilde{T}u)$. Wegen der Eindeutigkeit der Laplace-Transformation folgt $Tu = \tilde{T}u$. Weil $L_{\alpha}^1([0, \infty); \mathbb{K}^m)$ folgen-dicht in $L_{\text{loc}}^1([0, \infty); \mathbb{K}^m)$ ist und die Operatoren T und $\tilde{T}$ beide L_{loc}^1 -stetig sind, folgt $T = \tilde{T}$. □

6.4 Realisierungstheorie

Definition 6.4.1 (Realisierbarkeit). (a) Sei $T: L_{\text{loc}}^1([0, \infty); \mathbb{K}^m) \rightarrow L_{\text{loc}}^1([0, \infty); \mathbb{K}^p)$ ein lineares zeitinvariantes System. Wir nennen T **realisierbar** (durch eine lineare gewöhnliche Differentialgleichung über $\mathbb{K}$), wenn es Matrizen $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$, $D \in \mathbb{K}^{p \times m}$ gibt derart, dass

$$Tu = (Ce^{\cdot A}B) \star u + Du$$

für alle $u \in L_{\text{loc}}^1([0, \infty); \mathbb{K}^m)$ gilt.

In diesem Fall nennen wir das Tupel (A, B, C, D) eine **Realisierung** von T .

- (b) Sei $\alpha \in [-\infty, \infty)$. Wir nennen eine Funktion $G: \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times n}$ **realisierbar** (durch eine lineare gewöhnliche Differentialgleichung über $\mathbb{K}$), wenn es Matrizen $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$, $D \in \mathbb{K}^{p \times m}$, gibt derart, dass

$$G(\lambda) = C(\cdot - A)^{-1}B + D$$

für alle $\lambda \in \mathbb{C}_{>\alpha}$ gilt.

In diesem Fall nennen wir das Tupel (A, B, C, D) eine **Realisierung** von G .

Der Einfachheit halber betrachten wir zunächst dem Fall, in dem der Eingangs- und der Ausgangsraum jeweils ein-dimensional sind.

Proposition 6.4.2 (Realisierbarkeit im Single-Input-Single-Output-Fall). *Wir betrachten den Fall $m = p = 1$.*

- (a) *Ein lineares zeitinvariantes System $T: L_{\text{loc}}^1([0, \infty); \mathbb{K}) \rightarrow L_{\text{loc}}^1([0, \infty); \mathbb{K})$ ist genau dann realisierbar über, wenn es eine Zahl $D \in \mathbb{K}$ und eine Funktion*

$$k \in \text{span}_{\mathbb{K}}\{(t \mapsto t^\ell e^{\mu t}) \mid \ell \in \mathbb{N}_0, \mu \in \mathbb{C}\} \subseteq L_{\text{loc}}^1([0, \infty); \mathbb{C})$$

gibt, die nach $\mathbb{K}$ abbildet, derart, dass $Tu = k \star u + Du$ für alle $u \in L_{\text{loc}}^1([0, \infty); \mathbb{K})$ gilt.

- (b) *Sei $\alpha \in [-\infty, \infty)$. Eine Funktion $G: \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}$ ist genau dann realisierbar, wenn es Polynome p, q über $\mathbb{K}$ mit $\deg p \leq \deg q$ gibt derart, dass q auf $\mathbb{C}_{>\alpha}$ nicht verschwindet und dass $G(\lambda) = \frac{p(\lambda)}{q(\lambda)}$ für alle $\lambda \in \mathbb{C}_{>\alpha}$ gilt.*

Beweis. (b) Das folgt aus Aufgabe 4 auf Blatt 11.

(a) Das folgt aus (b). □

Lassen Sie uns nun den Fall für allgemeines p und n betrachten. Um Funktionentheorie verwenden zu können, betrachten wir im folgenden Satz 6.4.3 lediglich den komplexen Skalarkörper. Ähnliches gilt aber auch über $\mathbb{R}$, siehe zum Beispiel [Son98, Abschnitt 6.5].

Wir konzentrieren uns in Satz 6.4.3 außerdem auf Realisierungen, bei denen die Durchgriffsmatrix D gleich 0 ist.⁴ Wir benötigen die folgende Beobachtung:

Sei $r \geq 0$ und sei $G: \{\lambda \in \mathbb{C} \mid |\lambda| > r\} \rightarrow \mathbb{C}^{p \times m}$ eine holomorphe Funktion ist, die für $|\lambda| \rightarrow \infty$ gegen 0 konvergiert. Dann besitzt G eine Laurentreihen-Entwicklung mit Mittelpunkt 0. Genauer gibt es eine Folge von eindeutig bestimmten Matrizen $M_0, M_1, \dots \in \mathbb{C}^{p \times m}$ gibt derart, dass

$$G(\lambda) = \sum_{k=0}^{\infty} \frac{1}{\lambda^{k+1}} M_k \tag{6.4.1}$$

⁴Den Fall $D \neq 0$ zu berücksichtigen ist nicht viel schwieriger; aber im Falle $D = 0$ wird die Notation ein bisschen einfacher.

für alle $\lambda \in \mathbb{C}$ mit $|\lambda| > r$ gilt,⁵ wobei die Reihe absolut konvergiert. Die Matrizen $M_0, M_1, \dots$ nennt man die **Laurentkoeffizienten** von G (um den Entwicklungspunkt 0).

Satz 6.4.3 (Realisierbarkeit im Multiple-Input-Multiple-Output-Fall). *Sei $\mathbb{K} = \mathbb{C}$. Sei $\alpha \in [-\infty, \infty)$ und sei $G: \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$. Dann sind äquivalent:*

- (i) *Die Funktion G ist realisierbar mit Durchgriffsmatrix $D = 0$.*
- (ii) *Alle Einträge von G sind rationale Funktionen, wobei in jedem Eintrag der Zählergrad kleiner als der Nennergrad ist.*
- (iii) *Die Funktion G lässt sich außerhalb einer kompakten Teilmenge von $\mathbb{C}$ zu einer holomorphen Funktion fortsetzen für $|\lambda| \rightarrow \infty$ gegen 0 konvergiert; außerdem gibt es ein $n \in \mathbb{N}$ und Matrizen $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times m}$, $C \in \mathbb{C}^{p \times n}$ derart, dass die Koeffizienten $M_0, M_1, \dots$ in der Laurent-Entwicklung (6.4.1) die Bedingung $M_k = CA^k B$ für alle $k \geq 0$ erfüllen.⁶*
- (iv) *Die Funktion G lässt sich außerhalb einer kompakten Teilmenge von $\mathbb{C}$ zu einer holomorphen Funktion fortsetzen für $|\lambda| \rightarrow \infty$ gegen 0 konvergiert; außerdem ist die Folge der Koeffizienten $M_0, M_1, \dots$ in der Laurent-Entwicklung (6.4.1) **rekursiv**, das heißt, es gibt ein $n \in \mathbb{N}$ und Zahlen $\alpha_0, \dots, \alpha_{n-1} \in \mathbb{C}$ derart, dass*

$$M_{j+n} + \alpha_{n-1}M_{j+n-1} + \dots + \alpha_1 M_{j+1} + \alpha_0 M_j = 0$$

für alle $j \geq 0$ gilt.

Beweis von Satz 6.4.3. “(i) $\Rightarrow$ (ii)” Das folgt aus der Formel für die Transferfunktion (Beispiel 6.3.7) und der Cramerschen Regel für die Inverse von Matrizen.

“(ii) $\Rightarrow$ (i)” Diese Aussage lässt sich leicht aus dem Single-Input-Single-Output-Fall in Proposition (b) herleiten, indem man die Realisierungen von mp Single-Input-Single-Output-Systemen geeignet zusammensetzt.

“(i) $\Rightarrow$ (iii)” Das folgt aus der Formel für die Transferfunktion (Beispiel 6.3.7) und der Neumann-Reihen-Darstellung

$$(\lambda - A)^{-1} = \sum_{k=0}^{\infty} \frac{A^k}{\lambda^{k+1}}$$

der Resolvente für $|\lambda| > \|A\|$.

“(iii) $\Rightarrow$ (iv)” Das folgt aus dem Satz von Cayley-Hamilton.

“(iv) $\Rightarrow$ (ii)” Das folgt durch komponentenweises Anwenden des skalaren Resultats im nachfolgenden Lemma 6.4.4. $\square$

⁵Beachten Sie, dass wir wieder eine Notation gewählt haben, bei der der Index der Koeffizienten im Vergleich zur Potenz um 1 verschoben ist.

⁶Manchmal werden die Matrizen $CB, CAB, CA^2B, CA^3B, \dots$, die **Markov-Parameter** des Systems (A, B, C) genannt.

Lemma 6.4.4 (Wann sind Laurent-Reihen rationale Funktionen?). *Sei $r \geq 0$ und sei $G : \{\lambda \in \mathbb{C} \mid |\lambda| > r\} \rightarrow \mathbb{C}$ eine holomorphe Funktion, die für $|\lambda| \rightarrow \infty$ gegen 0 konvergiert. Sei*

$$G(\lambda) = \sum_{k=0}^{\infty} \frac{1}{\lambda^{k+1}} M_k$$

für $|\lambda| > r$ die Laurent-Reihenentwicklung von G mit Koeffizienten $M_0, M_1, \dots \in \mathbb{C}$. Falls es Zahlen $\alpha_0, \dots, \alpha_{n-1} \in \mathbb{C}$ gibt derart, dass

$$M_{j+n} + \alpha_{n-1} M_{j+n-1} + \dots + \alpha_1 M_{j+1} + \alpha_0 M_j = 0$$

für alle $j \geq 0$ gilt, dann ist G eine rationale Funktion (mit Zählergrad $<$ Nennergrad).

Beweis. [Siehe Vorlesung.] □

Als nächstes untersuchen wir, wie klein man die Dimension n des Zustandsraums wählen kann, auf dem man eine Transferfunktion G realisiert. Wir beschränken uns der Einfachheit halber weiterhin auf den Fall $\mathbb{K} = \mathbb{C}$, obwohl ähnliches auch für den Körper $\mathbb{R}$ gilt (siehe zum Beispiel [Son98, Abschnitt 6.5]).

Die folgenden Objekte sind sehr nützliche Hilfsmittel:

Definition 6.4.5 (Hankel-Matrizen). Sei $M = (M_k)_{k \in \mathbb{N}_0}$ eine Folge von Matrizen in $\mathbb{K}^{p \times m}$. Für alle $s, t \in \mathbb{N}_0$ heißt die Matrix

$$H_{s,t}(M) := \begin{pmatrix} M_0 & M_1 & \dots & M_t \\ M_1 & M_2 & \dots & M_{t+1} \\ \vdots & \ddots & \dots & \vdots \\ M_s & M_{s+1} & \dots & M_{s+t} \end{pmatrix} \in \mathbb{K}^{p(s+1) \times m(t+1)}$$

die (s, t) -te **Hankel-Matrix** zu M .⁷

Proposition 6.4.6. *Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$ und $M := (CA^k B)_{k \in \mathbb{N}_0}$ für alle $k \in \mathbb{N}_0$. Für alle $s, t \in \mathbb{N}_0$ gilt dann*

$$\underbrace{\begin{pmatrix} C \\ CA \\ \vdots \\ CA^s \end{pmatrix}}_{=: O_s(A, C)} \underbrace{\begin{pmatrix} B & AB & \dots & A^t B \end{pmatrix}}_{=: R_t(A, B)} = H_{s,t}(M).$$

Beweis. Das folgt sofort aus der Definition von $H_{s,t}(M)$. □

⁷Bitte beachten Sie, dass sich unsere Indizierung geringfügig von der in [Son98, Seite 285] unterscheidet.

Satz 6.4.7 (Minimalität von Realisierungen). *Sei $\mathbb{K} = \mathbb{C}$. Seien $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$, $C \in \mathbb{K}^{p \times n}$, $D \in \mathbb{K}^{p \times m}$. Die Transferfunktion des Systems (A, B, C, D) bezeichnen wir mit G . Dann sind äquivalent:*

- (i) *Die Realisierung (A, B, C, D) von G ist **minimal**, das heißt, es gibt keine Realisierung von G auf einem Zustandsraum der Dimension $< n$.*
- (ii) *Es ist (A, B) steuerbar und (A, C) beobachtbar.⁸*

Beweis. Ohne Einschränkung sei $D = 0$. Wir verwenden die Notation aus Proposition 6.4.6.

“(i) $\Rightarrow$ (ii)” [Siehe Vorlesung.]

“(ii) $\Rightarrow$ (i)” Da die Matrizen $R_{n-1}(A, B) \in \mathbb{K}^{n \times nm}$ und $O_{n-1}(A, C) \in \mathbb{K}^{np \times n}$ wegen (ii) Rang n haben, besitzt die erste von beiden eine Rechtsinverse $R^\# \in \mathbb{K}^{nm \times n}$ und die zweite eine Linksinverse $O^\# \in \mathbb{K}^{n \times np}$. Wegen Proposition 6.4.6 gilt

$$O^\# H_{n-1, n-1}(M) R^\# = \text{id} \in \mathbb{K}^{n \times n},$$

Also hat $H_{n-1, n-1}(M)$ mindestens Rang n .

Sei nun $(\tilde{A}, \tilde{B}, \tilde{C})$ eine weitere Realisierung derselben Transferfunktion G mit Zustandsraum der Dimension $\tilde{n}$. Seien $M_0, M_1, \dots$ die Laurentkoeffizienten von G zum Entwicklungspunkt 0. Dann gilt $M_k = CA^k B = \tilde{C} \tilde{A}^k \tilde{B}$ für alle $k \in \mathbb{N}_0$, also sind die Hankelmatrizen für beide Systeme gleich. Proposition 6.4.6 (angewendet auf $(\tilde{A}, \tilde{B}, \tilde{C})$) zeigt, dass $\text{rg } H_{s, t}(M) \leq \tilde{n}$ for all $s, t \in \mathbb{N}_0$. Also gilt insbesondere

$$n \leq H_{n-1, n-1}(M) \leq \tilde{n}.$$

Das zeigt die behauptete Minimalität. □

Satz 6.4.8 (Ähnlichkeit von minimalen Realisierungen). *Seien $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times m}$, $C \in \mathbb{C}^{p \times n}$, $D \in \mathbb{C}^{p \times m}$ sowie $\tilde{A} \in \mathbb{C}^{\tilde{n} \times \tilde{n}}$, $\tilde{B} \in \mathbb{C}^{\tilde{n} \times m}$, $\tilde{C} \in \mathbb{C}^{p \times \tilde{n}}$, $\tilde{D} \in \mathbb{C}^{p \times m}$.*

Seien (A, B) und $(\tilde{A}, \tilde{B})$ steuerbar sowie (A, C) und $(\tilde{A}, \tilde{C})$ beobachtbar. Wenn die Systeme (A, B, C, D) und $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ dieselbe Transferfunktion haben,⁹ dann sind sie ähnlich, das heißt es gilt $n = \tilde{n}$ und es gibt eine invertierbare Matrix $S \in \mathbb{C}^{n \times n}$ derart, dass

$$A = S^{-1} \tilde{A} S, \quad B = S^{-1} \tilde{B}, \quad C = \tilde{C} S, \quad D = \tilde{D}$$

gilt.

Beweis. [Siehe Vorlesung.] □

⁸In diesem Fall bezeichnet man das Tupel (A, B, C) übrigens als **kanonisch**.

⁹Auf einer Halbebene, die genügend weit rechts liegt – und somit auch überall nach dem Heben eventueller Definitionslücken.

6.5 Realisierungen positiver Systeme

Definition 6.5.1 (Positive LTI-Systeme). Ein lineares Zeit-invariantes System $T : L_{\text{loc}}^1([0, \infty); \mathbb{K}^m) \rightarrow L_{\text{loc}}^1([0, \infty))$ heißt **positiv**, falls $Tf \geq 0$ für alle $f \geq 0$ gilt, wobei die Ungleichungen fast überall zu verstehen sind.

Proposition 6.5.2 (Positivität via Impulsantwort). Sei $k \in L_{\text{loc}}^1([0, \infty); \mathbb{K}^{p \times m})$ und sei T sei LTI-System, dass durch Faltung mit k gegeben ist. Es ist T genau dann positiv, wenn $k \geq 0$ gilt.¹⁰

Beweis. “ $\Leftarrow$ ” Diese Implikation ist klar.

“ $\Rightarrow$ ” Die sind man, indem man T auf eine approximative Eins anwendet. $\square$

Definition 6.5.3 (Vollständige monotone Funktion). Sei $\alpha \in [-\infty, \infty)$. Eine holomorphe Funktion $f : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}$ heißt **vollständig monoton**,¹¹ falls für alle reellen Zahlen $\lambda \in (\alpha, \infty)$ und alle $n \in \mathbb{N}_0$ die Bedingung $(-1)^n f^{(n)}(\lambda) \geq 0$ für die n -te Ableitung von f gilt.

Beispiele 6.5.4 (Einige vollständig monotone Funktionen). (a) Die konstante Funktion $\mathbb{1}$ ist vollständig monoton.

(b) Für jedes $\beta \geq 0$ ist die ganze Funktion $\exp(-\beta \cdot)$ vollständig monoton.

(c) Für jedes $\beta \in \mathbb{R}$ ist die Funktion $\frac{1}{\cdot - \beta}$ auf $\mathbb{C}_{>\beta}$ vollständig monoton.

Proposition 6.5.5 (Verhalten vollständig monotoner Funktionen). Sei $\alpha \in [-\infty, \infty)$.

- (a) Wenn eine Folge vollständig monotoner Funktionen auf $\mathbb{C}_{>\alpha}$ lokal gleichmäßig konvergiert, dann ist auch die Grenzfunktion vollständig monoton.
- (b) Sei $f : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}$ holomorph. Wenn es ein $\beta \in [\alpha, \infty)$ gibt deart, dass f auf $\mathbb{C}_{>\beta}$ vollständig monoton ist, dann ist f sogar auf $\mathbb{C}_{>\alpha}$ vollständig monoton.
- (c) Sei $f : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}$ vollständig monoton. Dann gilt für jedes $\lambda \in \mathbb{C}_{>\alpha}$ die Abschätzung $|f(\lambda)| \leq f(\text{Re } \lambda)$.

Beweis. [Siehe Vorlesung.] $\square$

Proposition 6.5.6 (Algebraische Konsistenzeigenschaften vollständig monotoner Funktionen). Sei $\alpha \in [-\infty, \infty)$ und seien $f, g : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}$ holomorph und vollständig monoton.

- (a) Es ist auch $(-1)^n f^{(n)}$ vollständig monoton für jedes $n \in \mathbb{N}_0$.
- (b) Es sind auch fg und $\beta f + \gamma g$ vollständig monoton für alle $\beta, \gamma \in [0, \infty)$.
Insbesondere ist damit auch f^n vollständig monoton für jedes $n \in \mathbb{N}_0$.

¹⁰Hierbei ist die Ungleichung wiederum fastüberall zu verstehen und wie in Abschnitt 3.2 nennen wir eine Matrix M positiv und schreiben $M \geq 0$, falls alle ihre Einträge ≥ 0 sind.

¹¹Auf Englisch: **completely monotonic**

- (c) Ist $\beta > 0$ eine reelle Zahl mit $\beta < f(\lambda)$ für alle $\lambda \in (\alpha, \infty)$, dann ist auch $\lambda \mapsto \frac{1}{\beta - f(\lambda)}$ auf $\mathbb{C}_{>\alpha}$ wohldefiniert und vollständig monoton.

Beweis. [Siehe Vorlesung.] □

Satz 6.5.7 (Positivität via Laplace-Transformation). Sei $\alpha \in [-\infty, \infty)$ und $k \in L^1_\alpha([0, \infty); \mathbb{C})$. Es gilt $k \geq 0$ genau dann, wenn die Laplace-Transformierte $\mathcal{L}k$ vollständig monoton ist.

Beweis. [Siehe Vorlesung.] □

Satz 6.5.8 (Positivität via Transferfunktion). Sei $T : L^1_{\text{loc}}([0, \infty); \mathbb{K}^m) \rightarrow L^1_{\text{loc}}([0, \infty))$ ein lineares Zeit-invariantes System, dass exponentiell beschränkt vom Typ $\alpha \in [-\infty, \infty)$ ist.

Es ist T genau dann positiv, wenn jede Komponente seiner Transferfunktion vollständig monoton ist.

Beweis. Es genügt den Fall $m = p = 1$ zu betrachten.

“ $\Rightarrow$ ” Verwende die Approximation der Transferfunktion durch die Laplacetransformierten einer geeigneten approximativen Eins (siehe den Beweis von Satz 6.3.2).

“ $\Leftarrow$ ” Für jedes $0 \leq u \in L^1_\alpha([0, \infty); \mathbb{C})$ gilt $\mathcal{L}(Tu) = G\mathcal{L}u$ und diese Funktion ist vollständig monoton als Produkt von zwei vollständig monotonen Funktionen. Also ist $Tu \geq 0$ laut Satz 6.5.7. Weil $L^1_\alpha([0, \infty); \mathbb{C})_+$ dicht in $L^1_{\text{loc}}([0, \infty); \mathbb{C})_+$ ist, folgt die Behauptung. □

6.6 Externe und interne Stabilität

Wir betrachten erneut ein lineares System der Form

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), & x(0) = x_0, \\ y(t) = Cx(t) \end{cases} \quad \Sigma(A, B, C)$$

wobei $t \geq 0$, $A \in \mathbb{K}^{n \times n}$, $B \in \mathbb{K}^{n \times m}$ und $C \in \mathbb{K}^{p \times n}$. In Satz 3.1.1 haben wir den Begriff der Stabilität eines Systems als eine Eigenschaft der Matrix A eingeführt. Die Eingangs- und Ausgangsmatrizen B und C , die beschreiben wie ein System Extern wirkt, waren für diesen Begriff der Stabilität irrelevant. Wir wollen nun einen Begriff der Stabilität einführen der diesem Umstand Rechnung trägt.

Definition 6.6.1. (i) Wir nennen das System $\Sigma(A, B, C)$ *intern stabil*, falls $s(A) < 0$, also falls alle Eigenwerte von A in der linken offenen Halbebene liegen.

- (ii) Wir nennen $\Sigma(A, B, C)$ *Eingangs-Ausgangs stabil*, *extern stabil* oder *BIBO-stabil*¹², falls es eine konstante $\gamma > 0$ gibt, sodass für alle Lösungen von $\Sigma(A, B, C)$ mit $x_0 = 0$ und $u \in L^\infty([0, \infty); \mathbb{K}^m)$ gilt, dass

$$\|y\|_\infty \leq \gamma \|u\|_\infty.$$

¹²BIBO steht für die Englische Bezeichnung “Bounded Input Bounded Output”

Hierbei ist

$$\|u\|_\infty := \operatorname{ess\,sup}_{t \in [0, \infty)} \|u(t)\|.$$

Wir erinnern, dass das zu $\Sigma(A, B, C)$ zugehörige lineare zeitinvariante System (LTI) gegeben war durch die Abbildung

$$\begin{aligned} T : L^1_{\text{loc}}([0, \infty); \mathbb{K}^m) &\rightarrow L^1_{\text{loc}}([0, \infty); \mathbb{K}^p), \\ u &\mapsto \int_0^t k(t-s)u(s) \, ds, \end{aligned}$$

wobei k die L^1_{loc} -Impulsantwort des Systems gegeben war durch

$$k : [0, \infty) \rightarrow \mathbb{K}^{p \times m}, \quad t \mapsto Ce^{tA}B.$$

Es ist nun leicht zu sehen, dass $\Sigma(A, B, C)$ genau dann Eingangs-Ausgangs stabil ist, wenn T den Raum $L^\infty([0, \infty); \mathbb{K}^m)$ nach $L^\infty([0, \infty); \mathbb{K}^p)$ abbildet und die Einschränkung $T : L^\infty([0, \infty); \mathbb{K}^m) \rightarrow L^\infty([0, \infty); \mathbb{K}^p)$ stetig ist. Die Stetigkeit folgt sogar automatisch:

Satz 6.6.2. *Das System $\Sigma(A, B, C)$ ist genau dann Eingangs-Ausgangs stabil, wenn $Tu \in L^\infty([0, \infty); \mathbb{K}^p)$ für jedes $u \in L^\infty([0, \infty); \mathbb{K}^m)$*

Beweis. Das folgt direkt aus dem Satz vom abgeschlossenen Graphen. $\square$

Eine weitere Charakterisierung der Eingangs-Ausgangs Stabilität des Systems erhalten als Kriterium an die L^1_{loc} -Impulsantwort k des Systems.

Satz 6.6.3. *Das System $\Sigma(A, B, C)$ ist genau dann Eingangs-Ausgangs stabil, wenn $k \in L^1([0, \infty); \mathbb{K}^{p \times m})$, d.h.*

$$\int_0^\infty \|Ce^{tA}B\| \, dt < \infty.$$

Beweis. “ $\Rightarrow$ ”: Sei $u \in L^\infty([0, \infty); \mathbb{K}^m)$. Dann folgt

$$\begin{aligned} \|(Tu)(t)\| &= \left\| \int_0^t Ce^{(t-s)A}Bu(s) \, ds \right\| \\ &\leq \int_0^t \|Ce^{(t-s)A}B\| \|u(s)\| \, ds \\ &\leq \left(\int_0^t \|Ce^{(t-s)A}B\| \, ds \right) \|u\|_\infty \\ &= \left(\int_0^t \|Ce^{(s)A}B\| \, ds \right) \|u\|_\infty \\ &\leq \underbrace{\|g\|_1}_{\gamma:=} \|g\|_\infty. \end{aligned}$$

und somit ist $\|Tu\|_\infty \leq \|u\|_\infty$ for all $u \in L^\infty([0, \infty); \mathbb{K}^m)$.

“ $\Leftarrow$ ”: Ohne Einschränkung können wir vereinfacht annehmen, dass $m = p = 1$ gilt. Andernfalls zeigen wir die Aussage komponentenweise. Sei $t > 0$ nun beliebig. Wir definieren die Abbildung

$$\text{sign} : \mathbb{K} \rightarrow \mathbb{K}, \quad \text{sign}(x) := \begin{cases} \frac{\bar{x}}{|x|}, & x \neq 0, \\ 0, & x = 0, \end{cases}$$

und betrachten den Eingang

$$u_t : [0, \infty) \rightarrow \mathbb{K}, \quad u_t(s) := \begin{cases} \text{sign}(Ce^{(t-s)A}B), & s \in [0, t], \\ 0, & \text{sonst.} \end{cases}$$

Dann gilt $\|u_t\|_\infty \leq 1$ und es folgt aus der Eingangs-Ausgangs Stabilität mit Schranke γ , dass

$$\begin{aligned} \int_0^\infty |Ce^{sA}B| \, ds &= \sup_{t>0} \int_0^t |Ce^{sA}B| \, ds = \sup_{t>0} \int_0^t |Ce^{(t-s)A}B| \, ds \\ &= \sup_{t>0} |(Tu_t)(t)| \leq \sup_{t>0} \|Tu_t\|_\infty \\ &\leq \gamma \cdot \sup_{t>0} \|u_t\|_\infty \leq \gamma. \end{aligned}$$

Also ist k integrierbar. □

Lässt man den Eingang in der Zeit abklingen so haben Eingangs-Ausgangs stabile System die Eigenschaft, dass auch der Ausgang langsam abklingt.

Satz 6.6.4. *Ist $\Sigma(A, B, C)$ Eingangs-Ausgangs stabil und $u \in L^\infty([0, \infty); \mathbb{K}^m)$ ein Eingang mit $\lim_{t \rightarrow \infty} u(t) = 0$, so folgt $\lim_{t \rightarrow \infty} (Tu)(t) = 0$.*

Beweis. Es sei $\varepsilon > 0$ beliebig und sei k die Impulsantwort des Systems. Wir suchen ein $t_0 > 0$ so, dass für fast alle $t \geq t_0$ gilt, dass $\|(Tu)(t)\| \leq \varepsilon$. Nach Satz 6.6.3 ist k integrierbar und $\|k\|_1$ endlich. Ferner existiert nach Annahme ein $t_1 > 0$ so, dass

$$\|u(t)\| \leq \frac{\varepsilon}{2\|k\|_1}$$

für fast alle $t \geq t_1$ und

$$\int_{t_1}^\infty \|k(t)\| \, dt \leq \frac{\varepsilon}{2\|u\|_\infty}.$$

Sei nun $t_0 := 2t_1$. Mit dieser Wahl erhalten wir für $t > t_0$, dass

$$\begin{aligned} \|(Tu)(t)\| &\leq \|u\|_\infty \int_0^{t_1} \|k(t-s)\| \, ds + \frac{\varepsilon}{2\|k\|_1} \int_{t_1}^t \|k(t-s)\| \, ds \\ &= \|u\|_\infty \int_{t-t_1}^t \|k(s)\| \, ds + \frac{\varepsilon}{2\|k\|_1} \int_0^{t-t_1} \|k(s)\| \, ds \leq \varepsilon. \end{aligned} \quad \square$$

Wir sehen ferner einfach, dass die interne Stabilität die Eingangs-Ausgangs Stabilität impliziert.

Satz 6.6.5. *Ist das System $\Sigma(A, B, C)$ intern stabil, so ist es auch Eingangs-Ausgangs stabil.*

Beweis. Da $\Sigma(A, B, C)$ intern stabil ist, existieren nach Satz 3.1.1 Konstanten $M \geq 1$ und $\omega > 0$ derart, dass

$$\|e^{tA}\| \leq M e^{-\omega t}$$

für alle $t \geq 0$. Somit gilt

$$\begin{aligned} \int_0^\infty \|C e^{tA} B\| dt &\leq \int_0^\infty \|C\| \|e^{tA}\| \|B\| dt \\ &\leq \|C\| M \|B\| \int_0^\infty e^{-\omega t} dt \leq \frac{\|C\| M \|B\|}{\omega}. \end{aligned}$$

Also ist die L^1_{loc} -Impulsantwort des Systems $\Sigma(A, B, C)$ integrierbar. Die Eingangs-Ausgangs Stabilität folgt nun aus Satz 6.6.3. $\square$

Bemerkung 6.6.6. Die Umkehrung der Aussage von Satz 6.6.5 gilt nicht. Ein einfaches Gegenbeispiel erhält man, wenn wir $C = 0$ setzen. In diesem Fall ist das zugehörige LTI-System trivialerweise Eingangs-Ausgangs stabil, denn es gilt $Tu = 0$ für alle $u \in L^\infty([0, \infty); \mathbb{K}^m)$. Es ist hier egal ob $s(A) < 0$ oder nicht.

Unter der zusätzlichen Voraussetzungen erhalten wir allerdings die umgekehrte Implikation des Satzes 6.6.5. Dafür benötigen wir zunächst den folgenden Hilfssatz.

Wir erinnern, dass ein Matrixpaar (A, C) *asymptotisch 0-beobachtbar* heißt, wenn für alle $x_0 \in \mathbb{K}^n$ gilt, dass

$$(\forall t \in [0, \infty) : C e^{tA} x_0 = 0) \quad \Rightarrow \quad \lim_{t \rightarrow \infty} e^{tA} x_0 = 0.$$

Lemma 6.6.7. *Ein Matrixpaar (A, C) mit $A \in \mathbb{K}^{n \times n}$ und $C \in \mathbb{K}^{p \times n}$ ist genau dann asymptotisch 0-beobachtbar, wenn für alle $x_0 \in \mathbb{K}^n$ gilt, dass*

$$\lim_{t \rightarrow \infty} C e^{tA} x_0 = 0 \quad \Rightarrow \quad \lim_{t \rightarrow \infty} e^{tA} x_0 = 0.$$

Beweis. “ $\Leftarrow$ ”: Dies ist klar.

“ $\Rightarrow$ ”: Nach Proposition 4.5.2(b) und Satz 4.3.1 existiert eine Matrix $F \in \mathbb{K}^{n \times p}$ so, dass $s(A + FC) < 0$. Wir betrachten den Luenberger-Beobachter

$$\begin{cases} \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + F(\hat{y}(t) - y(t)) \\ \quad = (A + FC)\hat{x}(t) + (B - F) \begin{pmatrix} u(t) \\ y(t) \end{pmatrix}, \quad \hat{x}(0) = 0, \\ \hat{y}(t) = C\hat{x}(t), \end{cases} \quad (6.6.1)$$

wobei $u(\cdot)$ das Eingangs- und $y(\cdot)$ das Ausgangssignal von $\Sigma(A, B, C)$ ersetzen. Wir betrachten nun das System (6.6.1) mit dem alternativen Ausgang $\tilde{y} = \hat{x}$ und folgern aus Satz 6.6.5, dass dieses System Eingangs-Ausgangs stabil ist.

Sei nun $x_0 \in \mathbb{K}^n$ so, dass $\lim_{t \rightarrow \infty} Ce^{tA}x_0 = 0$. Für die Daten $u \equiv 0$ und $x(0) = x_0$ liefert das System $\Sigma(A, B, C)$ gerade den Ausgang $y(t) = Ce^{tA}x_0$. Es folgt nun aus Satz 6.6.4, dass das System (6.6.1) die Eingänge $y(t) = Ce^{tA}x_0$ und $u \equiv 0$, auf einen Ausgang $\tilde{y} = \hat{x}$ abbildet mit $\lim_{t \rightarrow \infty} \hat{x}(t) = 0$.

Da (6.6.1) ein Luenberger-Beobachter für $\Sigma(A, B, C)$ ist, folgt aus Satz 5.1.1 ferner, dass

$$\limsup_{t \rightarrow \infty} \|\hat{x}(t) - x(t)\| = 0.$$

Insgesamt erhalten wir $\lim_{t \rightarrow \infty} e^{tA}x_0 = \lim_{t \rightarrow \infty} x(t) = 0$, was zu zeigen war. $\square$

Satz 6.6.8. *Ist das System $\Sigma(A, B, C)$ Eingangs-Ausgangs stabil, asymptotisch 0-stabilisierbar und asymptotisch 0-beobachtbar, so ist $\Sigma(A, B, C)$ auch intern stabil.*

Beweis. Nach Satz 3.1.1 genügt es zu zeigen, dass $\lim_{t \rightarrow \infty} e^{tA}x_0$ für alle $x_0 \in \mathbb{K}^n$. Sei dafür $x_0 \in \mathbb{K}^n$ beliebig. Wir wählen zunächst eine Matrix $F \in \mathbb{K}^{n \times m}$ so, dass $s(A + BF) < 0$. Setze nun

$$\tilde{u}(t) := -Fe^{t(A+BF)}x_0$$

für alle $t \geq 0$. Dann gilt $\tilde{u} \in L^\infty([0, \infty); \mathbb{K}^m)$ und $\lim_{t \rightarrow \infty} \tilde{u}(t) = 0$. Da $\Sigma(A, B, C)$ Eingangs-Ausgangs stabil ist, folgt für den Ausgang y zum Anfangswert $x(0) = 0$ und Eingang $u = \tilde{u}$ mit Satz 6.6.4, dass $\lim_{t \rightarrow \infty} y(t) = 0$. Wir behaupten, dass die eindeutige Lösung $x(\cdot)$ zu ebendiesen Daten gegeben ist durch

$$x(t) = e^{tA}x_0 - e^{t(A+BF)}x_0 \quad \text{für alle } t \geq 0.$$

Dies verifizieren wir durch einfaches Differenzieren:

$$\begin{aligned} \dot{x}(t) &= Ae^{tA}x_0 - (A + BF)e^{t(A+BF)}x_0 \\ &= A(e^{tA}x_0 - e^{t(A+BF)}x_0) + B(-Fe^{t(A+BF)}x_0) \\ &= Ax(t) + B\tilde{u}(t). \end{aligned}$$

Also folgt für alle $t \geq 0$, dass

$$y(t) = Ce^{tA}x_0 - Ce^{t(A+BF)}x_0$$

und aus der Tatsache, dass wegen $s(A + BF) < 0$ der Term $Ce^{t(A+BF)}x_0$ für $t \rightarrow \infty$ gegen 0 konvergiert, auch $\lim_{t \rightarrow \infty} Ce^{tA}x_0 = 0$. Mit Lemma 6.6.7 folgt nun die Aussage. $\square$

Wir erinnern uns, dass nach Satz 6.6.3 das System $\Sigma(A, B, C)$ genau dann Eingangs-Ausgangs stabil ist, wenn die Impulsantwort k integrierbar ist. Aus praktischer Sicht ist dieses Kriterium schwer zu überprüfen, da hierfür das uneigentliche Integral

$$\int_0^\infty \|Ce^{tA}B\| \, ds$$

nach oben abgeschätzt werden muss. Praktischer wäre es, wenn wir zusätzlich noch ein Kriterium hätten, mit dem durch Betrachtung der Übergangsfunktion G die Eingangs-Ausgangs Stabilität feststellen könnten. Wir erinnern uns, dass die Übertragungsfunktion $G : \mathbb{C}_{>\alpha} \rightarrow \mathbb{C}^{p \times m}$ des Systems $\Sigma(A, B, C)$ gegeben ist durch

$$G(\lambda) = C(\lambda I - A)^{-1}B,$$

und dass $\alpha \geq s(A)$ gewählt werden kann. Präziser kann G sogar auf dem Bereich $\mathbb{C} \setminus \sigma(A)$ fortgesetzt werden. Ferner erinnern wir uns, dass G gerade die Laplace-Transformierte der Impulsantwort k ist.

Ferner haben wir in Satz 6.4.3 bereits gelernt, dass G genau dann realisierbar ist, wenn die Einträge von G rationale Funktionen sind, deren Zählergrade kleiner sind als deren Nennergrade. Wenn man die rationalen Zahlen alle auf einen gemeinsamen Nenner bringt, erhält man also eine Darstellung

$$G(\lambda) = \frac{1}{q(\lambda)}P(\lambda)$$

für alle $\lambda \in \mathbb{C}_{>\alpha}$, wobei $q \in \mathbb{K}[\lambda]$ und $P \in \mathbb{K}[\lambda]^{p \times m}$, sodass die Grade der Einträge von P kleiner sind als der Grad von q . Im Folgenden lernen wir, dass wir q speziell wählen können.

Lemma 6.6.9. *Sei G die Transferfunktion des Systems $\Sigma(A, B, C)$ und sei q das charakteristische Polynom von A . Dann existiert ein matrixwertiges Polynom $P \in \mathbb{K}[\lambda]^{p \times m}$ mit Einträgen vom Grad höchstens $n - 1$ so, dass*

$$G(\lambda) = \frac{1}{q(\lambda)}P(\lambda)$$

für alle $\lambda \in \mathbb{C} \setminus \sigma(A)$.

Beweis. Wir schreiben die Transferfunktion G als Laurentreihe

$$G(\lambda) = \sum_{k=0}^{\infty} \frac{1}{\lambda^{k+1}} M_k$$

für Koeffizienten $M_k = CA^k B \in \mathbb{K}^{p \times m}$ (siehe Satz 6.4.3). Somit folgt aus dem Satz von Cayley-Hamilton, dass

$$M_n + \alpha_{n-1}M_{n-1} + \cdots + \alpha_1 M_1 + \alpha_0 M_0 = 0,$$

wobei $q(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \dots + \alpha_1\lambda + \alpha_0$. Für $j \in \{0, \dots, n-1\}$ definieren wir uns die Matrixen $P_j \in \mathbb{K}^{p \times m}$ durch

$$P_j := M_{n-j-1} + \sum_{k=1}^{n-j-1} \alpha_{n-k} M_{n-j-1-k}$$

Dann folgt durch einen einfachen Koeffizientenvergleich und aus der Eindeutigkeit der Koeffizienten der Laurentreihe, dass

$$(\lambda^n + \alpha_{n-1}\lambda^{n-1} + \dots + \alpha_1\lambda + \alpha_0) \left(\sum_{k=0}^{\infty} \frac{1}{\lambda^{k+1}} M_k \right) = P_0 + P_1\lambda + \dots + P_{n-1}\lambda^{n-1}.$$

Setzen wir nun $P(\lambda) := P_0 + P_1\lambda + \dots + P_{n-1}\lambda^{n-1}$, so erhalten wir die Aussage. $\square$

Bemerkung 6.6.10. Es kann gezeigt werden, dass das System

$$\begin{cases} \dot{x}(t) = \begin{pmatrix} 0 & I_{m \times m} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & I_{m \times m} \\ -a_0 I_{m \times m} & -a_1 I_{m \times m} & \dots & \dots & -a_{n-1} I_{m \times m} \end{pmatrix} x(t) + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ I_{m \times m} \end{pmatrix} u(t), \\ y(t) = (P_0 \ P_1 \ \dots \ P_{n-1}) x(t) \end{cases}$$

eine Realisierung von G ist (vgl. Aufgabe 4 auf Übungsblatt 11 und Aufgabe 2 auf Übungsblatt 13).

Definition 6.6.11 (Polstelle). Sei G die Transferfunktion von $\Sigma(A, B, C)$. Ein Punkt $\lambda^* \in \mathbb{C}$ heißt (nicht-hebbare) *Polstelle von G* , wenn λ^* eine (nicht-hebbare) Polstelle für mindestens eine der Komponentenfunktionen ist.

In einer Darstellung der Transferfunktion G in der die Zähler und Nennerpolynome der Einträge nicht koprim sind, was zum Beispiel in der Darstellung von Lemma 6.6.9

$$G(\lambda) = \frac{1}{q(\lambda)} P(\lambda)$$

der Fall sein kann, ist λ^* genau dann eine nicht-hebbare Polstelle von G , wenn $j, k \in \mathbb{N}_0$ existieren mit $j < k$ so, dass λ^* eine k -fache Nullstelle des Nennerpolynoms und eine j -fache Nullstelle des Zählerpolynoms ist. (Der Fall $j = 0$ heißt hier, dass λ^* keine Nullstelle ist.)

Wir bemerken außerdem, dass λ^* genau dann eine nicht-hebbare Polstelle ist, wenn $\|G(\cdot)\|$ in jeder Umgebung von λ^* unbeschränkt ist.

Satz 6.6.12. Es sei $G : \mathbb{C} \setminus \sigma(A) \rightarrow \mathbb{C}^{p \times m}$ die (Fortsetzung der) Transferfunktion von $\Sigma(A, B, C)$. Dann sind äquivalent:

- (a) Das System $\Sigma(A, B, C)$ ist Eingangs-Ausgangs stabil.
- (b) Die nicht-hebbaren Polstellen von G liegen in $\mathbb{C}_{<0} \cap \sigma(A)$.
- (c) Es gilt

$$\sup_{\substack{\lambda \in \mathbb{C}, \\ \operatorname{Re}(\lambda) \geq 0}} \|G(\lambda)\| < \infty.$$

Bevor wir dies beweisen bemerken wir, dass uns ein Teil der Aussage bereits in ähnlicher Form bereits begegnet ist. So haben wir in Satz 6.3.8 die nicht-hebbaren Pole der (Fortsetzung der) Transferfunktion auf $\mathbb{C} \setminus \sigma(A)$ bereits als die Pole der Resolvente von A identifiziert, sofern das zugrundeliegende System $\Sigma(A, B, C)$ die stärkere Voraussetzung der Minimalität erfüllt (vgl. Satz 6.4.7).

Bemerkung 6.6.13. Wir bemerken, dass für $\mu \in \mathbb{N}$ und $\lambda^* \in \mathbb{K}$ die Funktion

$$k : [0, \infty) \rightarrow \mathbb{K}, \quad t \mapsto \frac{t^{\mu-1}}{(\mu-1)!} e^{\lambda^* t}$$

die Laplace-Transformierte

$$G(\lambda) = \int_0^\infty e^{-\lambda t} k(t) dt = \frac{1}{(\mu-1)!} \int_0^\infty t^{\mu-1} e^{-(\lambda-\lambda^*)t} dt = \frac{1}{(\lambda-\lambda^*)^\mu}$$

für $\operatorname{Re}(\lambda) > |\lambda^*|$ hat. Die letzte Gleichheit folgt durch wiederholtes partielles Integrieren.

Beweis von Satz 6.6.12. Die Äquivalenz von (c) und (c) ist klar. Wir zeigen noch die weitere Äquivalenz.

“(a) $\Rightarrow$ (c)”: Ist $\Sigma(A, B, C)$ Eingangs-Ausgangs stabil, so ist die zugehörige L^1_{loc} -Impulsantwort k integrierbar (siehe Satz 6.6.3). Somit erhalten wir, aus der Tatsache, dass G die Laplace-Transformierte von k ist, für alle $\lambda \in \mathbb{C}$ mit $\operatorname{Re}(\lambda) \geq 0$ die Ungleichung

$$\|G(\lambda)\| \leq \int_0^\infty \|k(t)\| \underbrace{\left| e^{-\lambda t} \right|}_{\leq 1} dt \leq \int_0^\infty \|k(t)\| dt < \infty.$$

“(c) $\Rightarrow$ (a)”: Nach Lemma 6.6.9 lässt sich die Transferfunktion G schreiben als

$$G(\lambda) = \frac{1}{\det(\lambda I - A)} P(\lambda)$$

für ein matrixwertiges Polynom $P \in \mathbb{K}[\lambda]^{p \times m}$ mit Einträgen vom Grad höchstens $n-1$. Wir schreiben nun $G = (G_{ij})_{i,j}$, $P = (P_{ij})_{i,j}$ und $k = (k_{ij})_{i,j}$, wobei $k : [0, \infty) \rightarrow \mathbb{K}^{p \times m}$ die L^1_{loc} -Impulsantwort des Systems $\Sigma(A, B, C)$ ist.

Sei nun $i \in \{1, \dots, p\}$ und $j \in \{1, \dots, m\}$. Dann gilt

$$G_{ij}(\lambda) = \frac{P_{ij}(\lambda)}{\det(\lambda I - A)} = \hat{k}_{ij}(\lambda)$$

für $\lambda \in \mathbb{C}_{>\alpha}$ für ein $\alpha \in \mathbb{R}$, wobei $\hat{k}_{ij}$ die Laplace-Transformierte des Eintrags k_{ij} der L^1_{loc} -Impulsantwort ist.¹³ Nach Satz 6.6.3 genügt es zu zeigen, dass $k_{ij} \in L^1([0, \infty); \mathbb{K})$. Dazu schreiben wir

$$G_{ij}(\lambda) = \frac{\tilde{p}_{ij}(\lambda)}{\prod_{\nu=1}^l (\lambda - \lambda_{\nu}^*)^{r_{\nu}}}$$

für $\lambda \in \mathbb{C} \setminus \sigma(A)$ mit Eigenwerten $\lambda_1^*, \dots, \lambda_l^* \in \mathbb{C}$ von A , $r_{\nu} \in \mathbb{N}$ und einem Polynom $\tilde{p}_{ij} \in \mathbb{K}[\lambda]$ derart, dass $\tilde{p}_{ij}$ und $\prod_{\nu=1}^l (\cdot - \lambda_{\nu}^*)^{r_{\nu}}$ teilerfremd sind. Man beachte, dass die gewählten Zähler- und Nennerpolynome und damit auch die Eigenwerte $\lambda_1, \dots, \lambda_l$ und die Vielfachheiten $r_1, \dots, r_l \in \mathbb{N}$ von i und j abhängen können.

Da nach Voraussetzung $\sup_{\text{Re}(\lambda) \geq 0} |G_{ij}(\lambda)| < \infty$ gilt, folgt $\{\lambda_1^*, \dots, \lambda_l^*\} \subseteq \mathbb{C}_{<0}$. Nach einer Partialbruchzerlegung hat G_{ij} die Form

$$G_{ij}(\lambda) = \sum_{\nu=1}^l \sum_{\mu=1}^{r_{\nu}} \frac{a_{\nu\mu}}{(\lambda - \lambda_{\nu}^*)^{\mu}}$$

mit $a_{\nu\mu} \in \mathbb{K}$ für alle $\lambda \in \mathbb{C}$ mit $\text{Re}(\lambda) \geq 0$.

Nach Bemerkung 6.6.13 und der Tatsache, dass G_{ij} die Laplace-Transformierte von k_{ij} ist, folgt

$$k_{ij}(t) = \sum_{\nu=1}^l \sum_{\mu=1}^{r_{\nu}} \frac{t^{\mu-1}}{(\mu-1)!} e^{\lambda_{\nu}^* t}$$

für fast alle $t \geq 0$. Aus der Tatsache, dass $\text{Re}(\lambda_{\nu}^*) < 0$ folgt nun, dass $k_{ij} \in L^1([0, \infty); \mathbb{K})$. $\square$

Wir halten fest:

Bemerkung 6.6.14. Sei G die Transferfunktion von $\Sigma(A, B, C)$. Dann gilt:

- (i) Alle Polstellen von G sind auch Eigenwerte von A . Im Allgemeinen sind nicht alle Eigenwerte von A auch Polstellen von G . (Im Falle $C = 0$ oder $B = 0$ hat G keine Polstellen.)
- (ii) Das System $\Sigma(A, B, C)$ ist genau dann intern stabil, wenn alle Eigenwerte von A in $\mathbb{C}_{<0}$ liegen (siehe Satz 3.1.1) und $\Sigma(A, B, C)$ ist genau dann Eingangs-Ausgangs stabil, wenn alle Polstellen von G in $\mathbb{C}_{<0}$ liegen (siehe Satz 6.6.12).

¹³Wir bemerken, dass wir $\text{Re}(\lambda) > \alpha$ benötigen um die Konvergenz der Laplace-Transformation zu erhalten. Natürlich können wir die Gleichheit dann a posteriori holomorph auf $\mathbb{C} \setminus \sigma(A)$ fortsetzen.

- (iii) Ist $\Sigma(A, B, C)$ asymptotisch 0-steuerbar und asymptotisch 0-beobachtbar, so sind alle instabilen Eigenwerte von A (also diejenigen mit nicht-negativem Realteil) auch Pole von G (ohne Beweis). In diesem Fall sind interne Stabilität und Eingangs-Ausgangs-Stabilität äquivalent (siehe Satz 6.6.8).
- (iv) Ist $\Sigma(A, B, C)$ steuerbar und beobachtbar, so sind alle Eigenwerte von A auch Pole von G (siehe Satz 6.3.8).

Literaturverzeichnis

- [Heu09] Harro Heuser. *Gewöhnliche Differentialgleichungen. Einführung in Lehre und Gebrauch*. Vieweg+Teubner, 6th updated ed. edition, 2009.
- [PW10] Jan W. Prüss and Mathias Wilke. *Gewöhnliche Differentialgleichungen und dynamische Systeme*. Grundstud. Math. Birkhäuser, 2010.
- [Son98] Eduardo D. Sontag. *Mathematical control theory. Deterministic finite dimensional systems.*, volume 6 of *Texts Appl. Math.* New York, NY: Springer, 2nd ed. edition, 1998.