Universität Bayreuth
Fakultät für Mathematik und Physik

Diplomarbeit

Bilineare Systeme

MATTHIAS ZERNDL

28. Juni 2007

Betreut durch
PROF. DR. FRANK LEMPIO
DR. ROBERT BAIER

Danksagung

An dieser Stelle möchte ich mich bei Prof. Dr. Frank Lempio und Dr. Robert Baier für ihre umfangreiche Betreuung herzlichst bedanken.
Ein besonderer Dank gilt auch meiner Familie für ihre moralische Unterstützung.

Inhaltsverzeichnis

Symbolverzeichnis

$\mathbb{N}$, $\mathbb{N}_0$	natürliche Zahlen ohne/mit 0
$\mathbb{Z}$	ganze Zahlen
$\mathbb{R}$	reelle Zahlen
$\mathbb{C}$	komplexe Zahlen
$\mathbb{R}_+$	positive reelle Zahlen
$\mathbb{R}^{n\times m}$	Vektorraum aller $n \times m-$Matrizen mit Koeffizienten aus $\mathbb{R}$
I	reelles Intervall
I^c	$\mathbb{R}\backslash I$
$\emptyset$	leere Menge
$B_r(p), B(p,r)$	offene Kugel $\{q \in \mathbb{R}^n \mid \|p-q\| < r\}$ mit Zentrum $p \in \mathbb{R}^n$ und Radius r
$\overline{M}$	abgeschlossene Hülle der Menge M
int $M, \delta M$	Inneres und Rand der Menge $M \subseteq \mathbb{R}^n$
K	abgeschlossene Einheitskugel $\overline{B_1(0)}$
E_m	Einheitswürfel 62
$x(\cdot),\ t \mapsto x(t)$	Funktion über einem reellen Intervall
$\dot{x}, f', \frac{\mathrm{d}}{\mathrm{d}t}f$	erste (komponentenweise) Ableitung der Funktion $x(\cdot)$ bzw. $f(\cdot)$ 135
$\dot{x} = f(x,u,t)$	Kurzform von $\dot{x}(t) = f(x(t),u(t),t)$
$\partial_i f(s), \frac{\partial}{\partial s_i} f(s)$	partielle Ableitung von f bezüglich der i-ten Koordinatenrichtung 49
$\mathrm{D}f$	Funktionalmatrix von der Funktion f
$x^{(k)}$	k-te Ableitung der Funktion $x(\cdot)$
$\mathrm{GL}(n,\mathbb{R})$	Gruppe der invertierbaren $n \times n-$Matrizen mit Koeffizienten aus $\mathbb{R}$
$\mathrm{GL}^+(n,\mathbb{R})$	Untergruppe der Matrizen aus $\mathrm{GL}(n,\mathbb{R})$ mit positiver Determinante 40

$\mathrm{O}(n) \subset \mathrm{GL}(n,\mathbb{R})$	Untergruppe der orthogonalen Matrizen
$\mathrm{SO}(n) = \mathrm{GL}^+(n,\mathbb{R}) \cap \mathrm{O}(n)$	spezielle orthogonale Gruppe
$\mathrm{Alt}(n,\mathbb{R})$	Vektorraum aller schiefsymmetrischen Matrizen aus $\mathbb{R}^{n\times n}$
S_n	symmetrische Gruppe aller Permutationen von Ordnung n
I	Einheitsmatrix
trA	Spur der Matrix $A \in \mathbb{R}^{n\times n}$ 155
$B = (b^1 \vert b^2 \vert \ldots \vert b^m)$	Matrix $B \in \mathbb{R}^{n\times m}$ mit Spalten $b^1, b^2, \ldots, b^m \in \mathbb{R}^n$ 10
$B = (b_{ij})$	Matrix B mit Koeffizienten b_{ij} an der $(i,j)-$ten Stelle
$e^A, \exp(A)$	Matrixexponentialfunktion von $A \in \mathbb{R}^{n\times n}$
$\mathrm{ad}_A^k N$	ad-Operator 28
$[A,N] = AN - NA$	Lie-Klammer des $\mathbb{R}^{n\times n}$
c^*, B^*	Transponierte des Vektors $c \in \mathbb{R}^n$ bzw. der Matrix $B \in \mathbb{R}^{n\times m}$
$\mathrm{Im}B = \{Bu \mid u \in \mathbb{R}^m\}$	Bild der Matrix $B \in \mathbb{R}^{n\times m}$
e_k	k-ter kanonischer Einheitsvektor
$\{S\}_{\mathrm{AA}}$	die kleinste (assoziative) Unteralgebra von $\mathbb{R}^{n\times n}$, welche die Teilmenge $S \subseteq \mathbb{R}^{n\times n}$ enthält 48
Σ	System 2,3
$\mathcal{T}$	Zeitmenge 1
$\mathcal{X}$	Zustandsraum 2
$\mathcal{U}$	Steuerbereich 2
$\mathcal{Y}$	Messbereich 3
ϕ	Zustandsübergangsfunktion 2
$t \in \mathcal{T}$	Zeit
$u \in \mathcal{U}$	Eingangsgröße 3
$x \in \mathcal{X}$	Zustand 3
$y \in \mathcal{Y}$	Ausgangsgröße 3
Σ^{-1}	zeitumgekehrte System von Σ 36
$\mathcal{U}^I$	Menge aller Abbildungen von I nach $\mathcal{U}$
$u \in \mathcal{U}^I$	Kontrollfunktion 3
$u_1 \&_s u_2(t)$	Konkatenation 2
$u^s(t) := u(t-s)$	Translation 4

$u\|_J$	Restriktion der Funktion $u \in \mathcal{U}^I$ auf ein Teilintervall $J \subseteq I$
u_k	k-te Komponente des Vektors $u \in \mathbb{R}^m$
$\mathcal{U}_u^I$	Klasse der messbaren, lokal (essentiell) beschränkten Funktionen aus $\mathcal{U}^I$ 11
$\mathcal{V}_r^I$	Klasse der messbaren Funktionen auf I, deren Bilder in einer kompakten Teilmenge $\mathcal{V} \subseteq \mathcal{U}$ sind 11
$\mathcal{V}_b^I$	Klasse der stückweise konstanten Funktionen auf I mit Werten in der kompakten Teilmenge $\mathcal{V} \subseteq \mathcal{U}$ 11
Λ	Ein-Ausgangsverhalten 6
λ	Responsefunktion 6
Ψ	Ein-Ausgangsfunktion 7
$\mathcal{A}_t(p)$	erreichbare Menge von p zur Zeit t 6,35
$\mathcal{A}(p)$	erreichbare Menge von p 6,35
$\mathcal{C}_t(p), \mathcal{C}(p)$	erreichbare Mengen bzgl. Σ^{-1} 36
$\mathcal{A}_t, \mathcal{A}, \mathcal{C}_t$	Abkürzungen für $\mathcal{A}_t(\mathrm{I}), \mathcal{A}(\mathrm{I}), \mathcal{C}_t(\mathrm{I})$ 37
$\mathcal{A}_{\leq T} := \bigcup_{0\leq t\leq T} \mathcal{A}_t$	60
$\mathcal{A}_t^b, \mathcal{A}_t^b(p)$	erreichbare Mengen bzgl. $\mathcal{V}_b^{[0,T]}$ 63
Φ	Fundamentallösung 23,150
χ_I	charakteristische Funktion von I 65
sgn	Signumfunktion
$O(t^k), o(\|u\|_\infty^l)$	Landau-Symbole 77,114
$\|x\|, \|x\|_2$	euklidische Norm für einen Vektor $x \in \mathbb{R}^n$ 130
$\|A\|$	Operatornorm für eine Matrix $A \in \mathbb{R}^{n\times m}$ 131
$\langle x, y\rangle$	euklidisches Skalarprodukt der Vektoren $x, y \in \mathbb{R}^n$
$\|x\|_\infty$	Maximumsnorm des Vektors $x \in \mathbb{R}^n$ 130
$C^0([a,b], \mathbb{R}^n)$	Raum der stetigen Funktionen von $[a,b]$ nach $\mathbb{R}^n$ 130
$\|f\|_\infty, \|f\|_0$	Supremumsnorm einer stetigen Funktion $f : [a,b] \to \mathbb{R}^n$ 130
$\mathcal{L}^\infty(I, \mathbb{R}^n)$	Raum der messbaren, essentiell beschränkten Funktionen von I nach $\mathbb{R}^n$ 131
$L^\infty(I, \mathbb{R}^n)$	Banachraum der Äquivalenzklassen aus $\mathcal{L}^\infty(I, \mathbb{R}^n)$ 131
$\|f\|_\infty$	essentielle Supremumsnorm von $f \in \mathcal{L}^\infty(I, \mathbb{R}^n)$ 131
ess. sup	essentielles Supremum 131
$\mathcal{L}_2([a,b], \mathbb{R}^n)$	Raum der quadratintegrablen Funktionen 131
$L_2([a,b], \mathbb{R}^n)$	Hilbertraum der Äquivalenzklassen aus $\mathcal{L}_2([a,b], \mathbb{R}^n)$ 131

$\langle f, g\rangle_{L_2}$	Skalarprodukt des $\mathcal{L}_2([a,b],\mathbb{R}^n)$ 131
$\|f\|_{L_2} := \langle f, f\rangle_{L_2}^{1/2}$	induzierte Norm
$L_2[0,T]$	Kurzform von $L_2([0,T],\mathbb{R}^m)$ 57
d_H	Hausdorffabstand 131
$\operatorname{dist}(x, C)$	Abstand des Vektors $x \in \mathbb{R}^n$ zu der Menge $C \subseteq \mathbb{R}^n$ 160
$\mathcal{L}^1([a,b])$	Menge aller über $[a,b]$ Lebesgue-integrierbaren Funktionen 134
$\operatorname{aff}(M)$	affine Hülle einer nichtleeren Teilmenge M des $\mathbb{R}^n$ oder $\mathbb{R}^{n\times n}$ 48
$\operatorname{ext}\mathcal{U}$	Menge aller extremen/extremalen Punkte von $\mathcal{U}$ 62
$\operatorname{cvx}\mathcal{N}$	konvexe Hülle einer nichtleeren Menge $\mathcal{N} \subseteq \mathbb{R}^m$ 62
$\forall$	für alle
$\forall'$	für fast alle
f.ü.	fast überall 130
$\mathcal{N}$	Menge aller komplett unkontrollierbaren Zustände 78
$\langle A; U\rangle$	kleinster A-invarianter Unterraum von $\mathbb{R}^n$, der die Teilmenge $U \subseteq \mathbb{R}^n$ enthält 78
$\mathcal{L}^\perp$	Orthogonalraum eines linearen Raums $\mathcal{L}$
$\sigma(p)$	Typ von p 85
$\omega(p)$	Konvexitätsausdehnung von p 85
$\mathrm{H}(p,\infty)$	Menge aller zulässigen Lösungen auf $[0,\infty)$ bzgl. des Anfangszustands p 83
$\omega_0, \omega_{i_1\dots i_k}$	Volterra-Kerne der Ordnung 0 bzw. k 102
ds^k	Kurzform von $\mathrm{ds_k ds_{k-1}} \dots \mathrm{ds_1}$
$\omega_{i_1\dots i_k}(t, s^k)$	Abkürzung für Volterra-Kern $\omega_{i_1\dots i_k}(t, s_1, \dots, s_k)$
H	Hilbertraum
$f_k \xrightarrow{w} f$, $f_k \xrightarrow{s} f$	schwache und starke Konvergenz einer Folge (f_k) in H gegen ein $f \in H$ 136
τ_w	schwache Topologie 136

Einleitung

Bilineare Systeme modellieren zahlreiche relevante Systeme aus Physik, Biologie und den Wirtschaftswissenschaften. Ganz gleich, ob die Neutronenpopulation in einem Kernreaktor, das Bremsverhalten eines Automobils, die Leistung eines Gleichstrommotors, die Regulierung des Hormons T4 im menschlichen Stoffwechsel, oder die Gewinnung von Weideland zur Viehzucht zu untersuchen sind—ein bilineares System bildet häufig die Grundlage für die Analyse und Optimierung solcher Prozesse [18].

Um die Bedeutung bilinearer Systeme nachvollziehen zu können, müssen wir verstehen, wie sich diese innerhalb der Klasse der nichtlinearen Kontrollsysteme einordnen lassen. Die zeitliche Entwicklung der Zustandsgrössen eines typischen bilinearen Systems ist gegeben durch gewöhnliche Differentialgleichungen der Form

$$\dot{x}_i(t) = \sum_{j=1}^{n} a_{ij} x_j(t) + \sum_{k=1}^{m} \sum_{j=1}^{n} n_{ij}^k u_k(t) x_j(t) + \sum_{k=1}^{m} b_{ik} u_k(t) \ , \quad a_{ij}, n_{ij}^k, b_{ik} \in \mathbb{R} \ ,$$

wobei x_i den i-ten Zustand ($1 \leq i \leq n$) und u_k die k-te Eingangsgröße bezeichnet. Die Eingangsgrößen gehen also sowohl additiv über den Term „$b_{ij}u_k$“ als auch multiplikativ über den Term „$n_{ij}^k u_k x_j$“ in die Zustandsgleichung ein. Sind die Koeffizienten b_{ij} oder n_{ij}^k ungleich Null, so nennen wir u_k demzufolge eine *additive* bzw. *multiplikative* Eingangsgröße. Die Tatsache, dass bei bilinearen Systemen neben den additiven noch multiplikative Eingangsgrößen berücksichtigt werden, ist der entscheidende Vorteil und Unterschied zu den linearen Systemen. So kann beispielsweise die Wachstumsrate einer exponentiell wachsenden Population (Beispiel 2.13), die Reaktivität eines Kernreaktors [19], die Kreisfrequenz der erzeugten Sinusschwingung eines Oszillators (Beispiel 2.15) oder der Erregerstrom eines Gleichstrommotors (Beispiel 2.16) durch eine multiplikative Eingangsgröße innerhalb eines bilinearen Systems modelliert werden.

Leider verhalten sich allgemeine bilineare Systeme nicht so angenehm und

durchschaubar wie lineare Kontrollsysteme: Die Analyse ist oft schwierig und führt zu schwächeren Resultaten. Die Problematik zeigt sich im Vergleich des Ein-Ausgangsverhaltens—angebrachter wäre hier der englische Begriff „Input-to-State Verhalten“—dieser beiden Systemklassen. Während sich die Lösung eines linearen Systems der Form

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(t) \in \mathbb{R}^n, \;\; u(t) \in \mathbb{R}^m, \;\; x(t_0) = x_0$$

über ein Faltungsintegral

$$x(t) = e^{At}x_0 + \sum_{i=1}^{m} \int_0^t e^{A(t-s)} b^i u_i(s) \mathrm{ds} \qquad (b^i = B\mathrm{e}_i) \tag{1}$$

darstellen läßt, ist die Lösung eines bilinearen Systems der Form

$$\dot{x}(t) = Ax(t) + \sum_{k=1}^{m} N_k u_k(t) x(t) + Bu(t), \;\; x(t) \in \mathbb{R}^n, \; u(t) \in \mathbb{R}^m, \; x(t_0) = x_0$$

(mit zusätzlichen $n \times n-$Matrizen $N_k = (n_{ij}^k)$) durch eine sog. Volterra-Reihe

$$\begin{aligned} x(t) = {} & e^{At}x_0 + \sum_{i=1}^{m} \int_0^t e^{A(t-s)} b^i u_i(s) \mathrm{ds} + \\ & + \sum_{k=2}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \int_0^t \int_0^{s_1} \cdots \int_0^{s_k} e^{A(t-s_1)} N_{i_1} e^{A(s_1-s_2)} N_{i_2} \cdots \\ & \qquad \cdots e^{A(s_{k-2}-s_{k-1})} N_{i_{k-1}} e^{A(s_{k-1}-s_k)} b^{i_k} u_{i_1}(s_1) \ldots u_{i_k}(s_k) \mathrm{ds_k} \ldots \mathrm{ds_1} \\ & + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \int_0^t \int_0^{s_1} \cdots \int_0^{s_k} e^{A(t-s_1)} N_{i_1} e^{A(s_1-s_2)} N_{i_2} \cdots \\ & \qquad \cdots e^{A(s_{k-1}-s_k)} N_{i_k} e^{As_k} x_0 \, u_{i_1}(s_1) \ldots u_{i_k}(s_k) \mathrm{ds_k} \ldots \mathrm{ds_1} \end{aligned}$$

gegeben. Offensichtlich ist die erste Darstellung leichter auszuwerten als die letzte—vor allem wenn die Volterra-Reihe nicht abbricht. Falls alle stückweise stetigen Kontrollfunktionen mit Werten in ganz $\mathbb{R}^m$ zugelassen sind, so können wir direkt aus (1) ablesen, dass die erreichbaren Mengen eines linearen Systems zu jedem Zeitpunkt $t > 0$ konvex sind und einen affinen Untervektorraum des $\mathbb{R}^n$ bilden. Die erreichbaren Mengen allgemeiner bilinearer Systeme dagegen müssen diese Eigenschaften zu keinem Zeitpunkt $t > 0$ besitzen (Beispiel 3.26). Deswegen zieht man sich beim Studium bilinearer Systeme gerne auf kleinere Teilklassen mit einfacherer Struktur zurück—wie z.B. Systeme von Rang 1 [10] oder Systeme mit quasikommutativen Matrizen

[14][15]—in der Hoffnung, diese besser charakterisieren zu können.

Jemand, der sich in Fachzeitschriften und -bücher über bilineare Systeme informieren möchte, wird feststellen, dass diese häufig nur als Spezialfall von Systemen der Form

$$\dot{x}(t) = f(x(t)) + \sum_{i=1}^{m} u_i(t) g_i(x(t)) \tag{2}$$

erwähnt werden, wobei $f, g_1, \ldots, g_m$ analytische Vektorfelder des $\mathbb{R}^n$ sind. Systeme vom Typ (2) heißen auch *Systeme mit linearer Kontrolle*. Neben ihrer Größe haben sie gegenüber bilinearen Systeme den Vorteil, dass man beim Verwenden von Feedback-Kontrollen wieder ein System der Form (2) erhält. Mit den Worten von Roger W. Brockett [4] gesprochen: Die Klasse der Systeme mit linearer Kontrolle ist abgeschlossen unter Komposition und Regelung. Viele Ergebnisse in Bezug auf das Ein-Ausgangsverhalten bilinearer Systeme sind ebenso für Systeme mit linearer Kontrolle gültig und von dort übernommen. Insbesondere hat die Ein-Ausgangsfunktion solcher Systeme eine Volterra-Reihen-Repräsentation [4][16].
Definitionsgemäß ist die rechte Seite der Zustandsgleichung $\dot{x} = f(x, u)$ eines bilinearen Systems nicht nur linear in der Kontrolle, sondern auch (affin-) linear im Zustand. (Dies erklärt die Verwendung des Adjektivs „bilinear“.) D.h. bei fest vorgegebener Kontrolle u ist die Zustandsgleichung durch ein lineares Differentialgleichungssystem der Form $\dot{x} = A(t)x + b(t)$ gegeben. Dies hat zur Folge, dass man ein—in der Regel nichtautonomes—lineares System lösen muss, um die Lösung eines bilinearen Systems bzgl. einer Kontrollfunktion zu gewinnen. Konzepte aus der hochentwickelten Theorie linearer Systeme wie Variation der Konstanten sind dabei sehr hilfreich und finden regelmäßig Verwendung in dieser Arbeit.

Diese Diplomarbeit soll dem Leser als Einführung in die Klasse der bilinearen Systeme dienen. Es werden vorwiegend theoretische Aspekte angesprochen, die anhand zahlreicher Beispiele erläutert werden und die Grundlage für Anwendungen in der Optimierung oder Numerik bilden könnten. Schwerpunkte setze ich bei der Untersuchung der erreichbaren Mengen und des Ein-Ausgangsverhaltens bilinearer Systeme. Ein besonderer Augenmerk liegt dabei auf der vom Mathematiker Otomar Hàjek eingeführten und äußert relevanten Teilklasse der „bilinearen Systeme von Rang 1“.

Zusammenfassung

Die Arbeit ist folgendermaßen aufgebaut:

In Kapitel 1 lernen wir Grundbegriffe aus der Systemtheorie, basierend auf der Axiomatik von Eduardo D. Sontag [23], kennen. Leser, die keinen Wert auf exakte Wortwahl legen oder bereits vertraut sind mit Begriffen wie „Ein-Ausgangsverhalten“, können dieses Kapitel mit ruhigem Gewissen überspringen.

Inhaltlich gesehen müssten auf das erste Kapitel die beiden Teile des Anhangs folgen. In Teil A werden elementare Konzepte aus der Funktionalanalysis (z.B. absolute Stetigkeit, Satz von Arzelà-Ascoli, schwache Konvergenz) aufgelistet, welche der Beweisführung dienlich sind.
Teil B beschäftigt sich mit kontinuierlichen Systemen. Im ersten Abschnitt führen wir den Lösungsbegriff im Sinne von Caratheodory zusammen mit Existenz- und Eindeutigkeitsaussagen ein. Der zweite Abschnitt ist den linearen Systemen gewidmet. Es wurde bereits angedeutet, dass wir vor allem an den Darstellungen und Eigenschaften der Lösungen nichtautonomer linearer Systeme interessiert sind. Die Resultate zu den autonomen Kontrollsystemen aus den letzten beiden Abschnitten werden in Abschnitt 2.3 auf autonome bilineare Systeme übertragen. Besonders interessant sind die Beispiele B.37 und B.39, die zeigen, dass die erreichbaren Mengen von autonomen Systemen mit kompaktem konvexen Steuerbereich weder beschränkt noch abgeschlossen sein müssen. Auf Beweise wird nicht verzichtet—selbst der Satz von Filippov, der ursprünglich ein wichtiger Bestandteil dieser Diplomarbeit sein sollte (bis er sich als obsolet erwiesen hat), ist weitgehend bewiesen.

Im ersten Abschnitt von Kapitel 2 stellen wir fest, dass die Lösungen von Anfangswertproblemen zu bilinearen Systemen—falls erforderlich—auf ganz $\mathbb{R}$ definiert und dort eindeutig bestimmt sind. Wir nutzen den zweiten Abschnitt dieses Kapitels, um anhand ausgewählter Beispiele das Leistungsvermögen bilinearer Systeme in der Modellierung physikalischer Systeme zu demonstrieren. Zu Beginn des dritten Abschnitt verallgemeinern wir das aus der Theorie linearer Systeme stammende Konzept der Fundamentallösung mit dem Ziel, die Lösungen bilinearer Systeme zu beschreiben. Später in diesem Abschnitt wollen wir kurz auf spezielle bilineare Systeme, nämlich Systeme mit quasikommutativen Matrizen und Systeme von Rang 1, eingehen.

Die erreichbaren Mengen bilinearer Systeme sind das Thema von Kapitel 3. In den ersten beiden Abschnitten werden wichtige Notationen und algebraische Eigenschaften eingeführt. Von fundamentaler Bedeutung sind die erreichbaren Mengen $\mathcal{A}_t$ von der Einheitsmatrix I bzgl. des assoziierten Ma-

trixsystems in $\mathbb{R}^{n\times n}$. Wie wir in Abschnitt 3.2 sehen werden, lassen sich die Elemente von $\mathcal{A}_t$ in Abhängigkeit von der Systemgleichung bestimmten Lie-Gruppen wie $\mathrm{GL}(n,\mathbb{R})$ oder $\mathrm{SO}(n)$ zuweisen.
Ein Ziel des dritten Abschnitts ist das Auffinden affiner Unterräume, welche die erreichbare Menge $\mathcal{A}_t(p)$ eines autonomen Systems enthalten. Die hergeleiteten Aussagen werden es uns ermöglichen, die affine Hülle von $\mathcal{A}_t(p)$ zu berechnen.
Die topologischen Eigenschaften aus dem vierten Abschnitt gelten ebenso für Systeme mit linearer Kontrolle [9, p.90ff]. Wir beweisen, dass die erreichbaren Mengen von bilinearen Systemen mit kompaktem konvexen Steuerbereich kompakt und wegzusammenhängend sind. Anhand von unterschiedlichen Bang-Bang Prinzipien charakterisieren wir die erreichbaren Mengen, die man bei Restriktion der zulässigen Kontrollen auf stückweise konstante Bang-Bang Funktionen erhält.
Der letzte Abschnitt konzentriert sich auf Systeme von Rang 1. Gemessen an den dort erzielten Resultaten (Maximumprinzip, schwaches Bang-Bang Prinzip) würde man den Abschnitt eher dem größeren Gebiet „Kontrolltheorie“ zuordnen. In diesem Zusammenhang stellt er eine Verallgemeinerung der linearen Kontrolltheorie dar—zumal sich autonome lineare Kontrollsysteme als Systeme von Rang 1 interpretieren lassen (Beispiel 2.34). Im ersten Abschnitt wird gezeigt, dass die erreichbaren Mengen von sog. Spalten-Kontrollsystemen ein nichtleeres Inneres besitzen (sofern sie nicht einelementig sind) und zumindest für kleine Zeiten konvex sind. Das Pontryaginsche Maximumprinzip liefert eine Charakterisierung extremaler Kontrollen, die speziell für Single-Input Systeme zu stärkeren Aussagen im nächsten Teilabschnitt führt. Dort sehen wir, dass Single-Input Systeme von Rang 1 das schwache Bang-Bang Prinzip erfüllen und zu kleinen Zeiten strikt konvexe erreichbare Mengen besitzen.

In Kapitel 4 wir das Ein-Ausgangsverhalten bilinearer beschrieben. Dazu bestimmen wir die Volterra-Reihen-Repräsentation der Ein-Ausgangsfunktion, indem wir die Kerne der zugehörigen Volterra-Reihe berechnen. Im dritten Abschnitt wird gezeigt, inwieweit diese Kerne eindeutig bestimmt sind. Der vierte Abschnitt beinhaltet ein Kriterium für die Endlichkeit der Volterra-Reihe unserer Repräsentation. Anwendungen und Beispiele befinden sich im letzten Abschnitt.

Zum Abschluss wird ein Ausblick auf Anwendungs- und Erweiterungsmöglichkeiten zu dieser Diplomarbeit gegeben.

Kapitel 1

Grundbegriffe aus der Systemtheorie

Die Notation und Terminologie aus diesem Kapitel sind weitgehend von Eduardo D. Sontag [23, Kapitel 2] übernommen. Missverständnisse im Umgang mit elementaren Begriffen wie „System“ oder „Ein-Ausgangsverhalten“ in späteren Kapiteln sollen verhindert werden.

1.1 Was ist ein System?

Zumindest für Systeme, die deterministisch, dynamisch und kausal sind, wird diese Frage in Definition 1.2 beantwortet. Stochastische Systeme, z.B., werden nicht berücksichtigt.

Definition 1.1.

- Eine *Zeitmenge* $\mathcal{T}$ ist eine Untergruppe von $(\mathbb{R}, +)$. Sie enthält sämtliche Zeitpunkte, zu denen sich der Zustand eines Systems ändern kann. (Beispiel: $\mathcal{T} = h\mathbb{Z}$ für eine *Schrittweite* $h \in \mathbb{R}$.)

- Für $a < b$ in $\mathbb{R} \cup \{+\infty\}$ sei $[a, b) := \{t \in \mathcal{T} \mid a \leq t < b\}$.

- Für $a < b$ in $\mathbb{R} \cup \{+\infty\}$ und einer nichtleeren Menge $\mathcal{U}$ sei

$$\mathcal{U}^{[a,b)} := \begin{cases} \{u \mid u : [a, b) \to \mathcal{U}\} \ , & \text{falls } a < b \\ \epsilon \ , & \text{falls } a = b \end{cases}$$

  die Menge aller Abbildungen von $[a, b)$ nach $\mathcal{U}$. ϵ nennt man die *leere Sequenz*.

- Für die Elemente a, s, b aus $\mathcal{T}$ mit $a \leq s \leq b$ sei die *Konkatenation* $u \in \mathcal{U}^{[a,b)}$ von $u_1 \in \mathcal{U}^{[a,s)}$ und $u_2 \in \mathcal{U}^{[s,b)}$ wie folgt definiert:

$$u(t) := u_1 \&_s u_2(t) := \begin{cases} u_1(t) \ , & \text{falls } a \leq t < s \\ u_2(t) \ , & \text{falls } s \leq t < b \end{cases}$$

Definition 1.2. Ein *System* Σ ist ein 4-Tupel $(\mathcal{T}, \mathcal{X}, \mathcal{U}, \phi)$ bestehend aus

- einer Zeitmenge $\mathcal{T}$;
- einer nichtleeren Menge $\mathcal{X}$, dem sog. *Zustandsraum*;
- einer nichtleeren Menge $\mathcal{U}$, dem sog. *Steuerbereich*;
- einer Funktion

$$\begin{array}{rccc} \phi: & D(\phi) & \longrightarrow & \mathcal{X} \\ & (t, t_0, x, u) & \longmapsto & \phi(t; t_0, x, u) \end{array}$$

  mit nichtleerer Teilmenge $D(\phi)$ von

$$\left\{(t, t_0, x, u) \mid t, t_0 \in \mathcal{T}, t_0 \leq t, x \in \mathcal{X}, u \in \mathcal{U}^{[t_0,t)}\right\}$$

  als Definitionsbereich, der sog. *Zustandsübergangsfunktion*;

welches die folgenden Axiome erfüllt:

1. Die *Nichttrivialität*, d.h. zu jedem *Zustand* $x \in \mathcal{X}$ existiert ein Paar $t_0 < t$ in $\mathcal{T}$ und eine Funktion $u \in \mathcal{U}^{[t_0,t)}$, so dass $(t, t_0, x, u) \in D(\phi)$. Man sagt, u ist *zulässig* für x.

2. Die *Zulässigkeit von Restriktionen*, d.h. falls $u \in \mathcal{U}^{[t_0,t)}$ zulässig für x ist, so ist für alle $t_1 \in [t_0, t)$ die Restriktion $u_1 := u|_{[t_0,t_1)}$ zulässig für x und $u_2 := u|_{[t_1,t)}$ ist zulässig für $\phi(t_1; t_0, x, u_1)$.

3. Die *Identitätseigenschaft* der Übergangsfunktion, d.h. für alle $t \in \mathcal{T}$ und alle $x \in \mathcal{X}$ ist die leere Sequenz ϵ zulässig für x und $\phi(t; t, x, \epsilon) = x$.

4. Die *Halbgruppeneigenschaft* der Übergangsfunktion, d.h. für alle t_0, t_1, t in $\mathcal{T}$ mit $t_0 < t_1 < t$ und alle $x \in \mathcal{X}$ gilt: Falls $u_1 \in \mathcal{U}^{[t_0,t_1)}$ zulässig für x ist und $u_2 \in \mathcal{U}^{[t_1,t)}$ zulässig für $\phi(t_1; t_0, x, u_1)$, so gilt bezüglich $x_1 := \phi(t_1; t_0, x, u_1)$ und $x_2 := \phi(t; t_1, x_1, u_2)$, dass $u := u_1 \&_{t_1} u_2$ zulässig für x ist und $\phi(t; t_0, x, u) = x_2$.

Definition 1.3. Die Elemente aus $\mathcal{U}$ nennt man *Eingangsgrößen*. Sie wirken auf den Zustand des Systems ein, ohne selbst beeinflusst zu werden.
Die Funktionen aus $\mathcal{U}^{[t_0,t)}$ heissen *Eingangsfunktionen* oder *Kontrollfunktionen* (kurz: *Kontrolle* oder *Steuerung*).
Die Elemente aus $\mathcal{X}$ nennt man *Zustandsgrößen* (kurz: *Zustände*). Sie ermöglichen es, zukünftige Zustände anhand der Eingangsfunktion zu bestimmen.
Man sagt, $x(t) := \phi(t; t_0, x_0, u)$ ist der Zustand zum Zeitpunkt t, den man ausgehend vom Anfangszustand x_0 und dem Anfangszeitpunkt t_0 mittels der Eingangsfunktion $u \in \mathcal{U}^{[t_0,t)}$ erhält. (Insbesondere $x_0 = x(t_0)$ nach Axiom 3.)

Bei der Modellierung technischer Systeme wird häufig berücksichtigt, dass es in der Praxis unmöglich oder zu aufwendig ist, die Werte einiger Zustandsgrössen zu bestimmen.
Grössen, die vom Zustand des Systems abhängen und gemessen werden, nennt man *Ausgangsgrößen*.

Beispiel 1.4. Als Beispiel wähle ich das klassische Quecksilberthermometer. Es besteht im Wesentlichen aus einer Glaskapillare, die mit Quecksilber gefüllt ist, und einer Skala. Zur Beschreibung der Funktionsweise könnte man die Umgebungstemperatur als Eingangsgröße, die Quecksilbertemperatur und das Quecksilbervolumen als Zustandsgrößen, und die Skalenanzeige als Ausgangsgröße verwenden.

Die folgende Definition berücksichtigt das Konzept der Systemsausgangsgrössen:

Definition 1.5. Ein *System* ***mit Ausgang*** ist gegeben durch ein System $\Sigma = (\mathcal{T}, \mathcal{X}, \mathcal{U}, \phi)$, zusammen mit

- einer Menge $\mathcal{Y}$, dem sog. *Messbereich*;
- einer Abbildung $y : \mathcal{T} \times \mathcal{X} \to \mathcal{Y}$, der sog. *Ausgangsfunktion*.

Die Elemente aus $\mathcal{Y}$ nennt man *Ausgangsgrössen*.

Ein System mit Ausgang ist ein 6-Tupel $(\mathcal{T}, \mathcal{X}, \mathcal{U}, \phi, \mathcal{Y}, y)$ und wird ebenfalls mit Σ bezeichnet.

1.2 Klassifikation

Die nachfolgenden Eigenschaften dienen der Klassifikation von Systemen.

Definition 1.6. Ein System Σ ist *ohne Kontrolle*, falls dessen Steuerbereich $\mathcal{U}$ einelementig ist.

Bei Systemen ohne Kontrolle gibt es für alle t_0, t genau ein mögliches $u \in \mathcal{U}^{[t_0,t)}$. Daher ist die Übergangsfunktion $\phi(t; t_0, x, u)$ unabhängig von der letzten Koordinate; wir schreiben dann kurz $\phi(t; t_0, x)$.

Beispiel 1.7. Natürlich gehören Systeme, die über klassische Differentialgleichungen der Form $\dot{x} = f(x, t)$ definiert sind, zur Klasse der Systeme ohne Kontrolle (siehe Satz B.17). Aber auch Feedback-Systeme, deren Eingangsgrößen $u(t)$ über eine Abbildung $F : \mathcal{X} \times \mathcal{T} \to \mathcal{U}$ mittels $u(t) := F(x(t), t)$ gegeben sind, lassen sich als Systeme ohne Kontrolle interpretieren.

Definition 1.8. Ein System Σ ist *endlich-dimensional* (über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$), falls $\mathcal{U}$ und $\mathcal{X}$ Teilmengen von endlich-dimensionalen Vektorräumen über $\mathbb{K}$ sind. Bei Systemen mit Ausgang gilt dies ebenso für $\mathcal{Y}$.
Falls $\mathcal{X} = \mathbb{K}^n$, so sagen wir auch, dass Σ ein „System im $\mathbb{K}^n$“ ist.

Definition 1.9. Ein System Σ ist *zeitinvariant* oder *autonom*, wenn für alle $u \in \mathcal{U}^{[t_0,t)}$, $x \in \mathcal{X}$ und $s \in \mathcal{T}$ die folgende Aussage gilt:
Falls u zulässig für x ist, so ist die Translation

$$u^s \in \mathcal{U}^{[t_0+s,t+s)},\ u^s(t) := u(t-s)$$

zulässig für x und

$$\phi(t; t_0, x, u) = \phi(t+s; t_0+s, x, u^s).$$

Bei Systemen mit Ausgang wird noch gefordert, dass die Ausgangsfunktion $y(t, x)$ unabhängig von t ist.

Definition 1.10. Ein System Σ mit einer abzählbaren unendlichen Zeitmenge $\mathcal{T}$ heißt *diskret*. (Ohne Einschränkung darf dann $\mathcal{T} := \mathbb{Z}$ gesetzt werden.)

Beispiel 1.11. Diskrete bilineare Systeme mit Zustandsraum $\mathcal{X} = \mathbb{R}^n$ und Steuerbereich $\mathcal{U} \subseteq \mathbb{R}^m$ sind durch eine Differenzengleichung der Form

$$x(t+1) = Ax(t) + \sum_{k=1}^{m} N_k u_k(t) x(t) + Bu(t), \quad t \in \mathbb{Z},\ \ u(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{pmatrix} \in \mathcal{U}$$

gegeben, wobei $B \in \mathbb{R}^{n\times m}$ und $A, N_k \in \mathbb{R}^{n\times n}$ $(k = 1, \ldots, m)$.

Definition 1.12. Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$. Unter einem *kontinuierlichen* System verstehen wir ein System $\Sigma = (\mathbb{R}, \mathcal{X}, \mathcal{U}, \phi)$ mit einer offenen Menge $\mathcal{X} \subseteq \mathbb{K}^n$

und einem metrischen Raum $\mathcal{U}$, dessen Dynamik durch eine Differentialgleichung der Form

$$\dot{x}(t) = f(t, x(t), u(t)), \quad f : \mathbb{R} \times \mathcal{X} \times \mathcal{U} \to \mathbb{K}^n \tag{1.1}$$

bestimmt ist.
Bei Systemen mit Ausgang wird noch gefordert, dass die Ausgangsfunktion $y : \mathcal{T} \times \mathcal{X} \to \mathcal{Y}$ stetig und $\mathcal{Y}$ ein metrischer Raum ist.

Bemerkung 1.13. Um aus der Differentialgleichung (1.1) eine wohldefinierte Übergangsfunktion ϕ zu gewinnen, muss die rechte Seite $f(t, x, u)$ sogenannte Regularitätsbedingungen erfüllen, welche in der Regel die eindeutige Existenz einer Lösung $t \mapsto x(t; t_0, x_0, u)$ von (1.1) bei vorgegebener Anfangsbedingung $x(t_0) = x_0$ und zulässiger Kontrollfunktion $u \in \mathcal{U}^{[t_0,t)}$ auf einem Intervall $[t_0, t]$ für ein $t > t_0$ sicherstellen. Dabei ist mit „zulässiger Kontrollfunktion“ häufig nur die Angehörigkeit der Kontrolle u zu einer Klasse von messbaren Funktionen gemeint, wie etwa der Menge der stückweise stetigen Funktionen. Der Begriff „Lösung“ ist—zumindest in dieser Arbeit—im Sinne Caratheodorys zu verstehen (Definition B.5). Setze dann

$$\phi(t; t_0, x_0, u) := x(t; t_0, x_0, u) \ .$$

Die Caratheodory-Bedingungen B.2 und die lokale Lipschitz-Bedingung in Lemma B.13 sind ein typisches Beispiel solcher Regularitätsbedingungen, wie wir in Satz 2.12 und B.17 sehen können. Allgemeinere Bedingungen findet man in [23, p.43].

1.3 Erreichbarkeit

Eine typische Problemstellung in der Kontrolltheorie ist die Bestimmung einer Steuerung oder Regelung, die einen Anfangszustand $p \in \mathcal{X}$ auf eine noch festzulegende optimale Weise (z.B. zeitoptimal) in einen Zustand $q \in \mathcal{Z}$ überführt, der zu einer vorgegebenen Zielmenge $\mathcal{Z} \subseteq \mathcal{X}$ gehört. Dabei stellt sich die Frage, ob und wann ein solcher Punkt q von p aus erreicht werden kann.

Ein System $\Sigma = (\mathcal{T}, \mathcal{X}, \mathcal{U}, \phi)$ sei vorgegeben.

Definition 1.14. Es seien $x, x_0 \in \mathcal{X}$ und $t, t_0 \in \mathcal{T}$ mit $t - t_0 := T \geq 0$. Angenommen, es existiert ein $u \in \mathcal{U}^{[t_0,t)}$, so dass

$$(t, t_0, x_0, u) \in D(\phi) \quad \text{und} \quad x = \phi(t, t_0, x_0, u) \ .$$

Dann sagen wir:
Die Kontrolle u *steuert* x_0 nach x zur Zeit T. Oder umgekehrt, x kann von x_0 zur Zeit T *erreicht* werden.

Speziell für autonome Systeme führen wir eine Notation ein, die sich auch auf nichtautonome Systeme übertragen lässt, indem man den Zeitpunkt t_0 aus Definition 1.14 fest vorgibt.

Definition 1.15. Es sei Σ autonom. Dann bezeichnet

$$\mathcal{A}_t(x_0) := \{x \in \mathcal{X} \mid x \text{ kann von } x_0 \text{ zur Zeit } t \text{ erreicht werden}\}$$

die *erreichbare Menge von x_0 zur Zeit t.*

Die *erreichbare Menge von x_0* ist

$$\mathcal{A}(x_0) := \bigcup_{\substack{t \in \mathcal{T}: \\ t \geq 0}} \mathcal{A}_t(x_0) \ .$$

Lemma 1.16. Es sei Σ autonom. Wenn x_1 nach x_2 (zur Zeit t_1) und x_2 nach x_3 (zur Zeit t_2) gesteuert werden kann, so auch x_1 nach x_3 (zur Zeit $t_1 + t_2$). Kurz:

$$x_2 \in \mathcal{A}_{t_1}(x_1), x_3 \in \mathcal{A}_{t_2}(x_2) \Longrightarrow x_3 \in \mathcal{A}_{t_1+t_2}(x_1) \tag{1.2}$$

Beweis. Diese Aussage folgt direkt aus der Halbgruppeneigenschaft des Systems Σ. Die Konkatenation zweier zulässiger Kontrollen führt zur Konkatenation der zugehörigen Lösungstrajektorien. □

1.4 Ein-Ausgangsverhalten

Bei vielen realen Systemen fällt es dem Mathematiker leichter, das System über sein Ein-Ausgangsverhalten zu beschreiben, als durch ein konzeptionelles Modell mit Zustandsraum und Übergangsfunktion.

Definition 1.17. (i) Ein *initialisiertes System* ist ein Paar (Σ, x_0), bestehend aus einem System Σ (mit/ohne Ausgang) und einem Anfangszustand $x_0 \in \mathcal{X}$.

(ii) Unter dem *Ein-Ausgangsverhalten* eines initialisierten Systems (Σ, x_0) mit Ausgang verstehen wir ein Viertupel $\Lambda = (\mathcal{T}, \mathcal{U}, \mathcal{Y}, \lambda)$ mit $\mathcal{T}, \mathcal{U}, \mathcal{Y}$ aus $\Sigma = (\mathcal{T}, \mathcal{X}, \mathcal{U}, \phi, \mathcal{Y}, y)$ und einer Funktion, *Responsefunktion* genannt, die wie folgt definiert ist:

$$\begin{aligned} \lambda: \qquad & D(\lambda) \longrightarrow \mathcal{Y} \\ & \lambda(t; t_0, u) \ := \ y(t; \phi(t; t_0, x_0, u)) \end{aligned}$$

mit Definitionsmenge $D(\lambda) := \{(t, t_0, u) \mid (t, t_0, x_0, u) \in D(\phi)\}$.

Bemerkung 1.18. $\lambda(t; t_0, u)$ berechnet diejenige Ausgangsgröße $y(t)$ zum Zeitpunkt t, welche ausgehend vom Anfangszustand $x(t_0) = x_0$ durch die Eingangsfunktion u erwirkt wird.

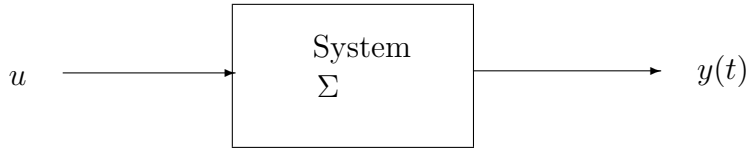

Abbildung 1.1: Schematische Darstellung des Ein-Ausgangsverhaltens

Alternativ zur Responsefunktion könnte man das Ein-Ausgangsverhalten auch durch eine Abbildung beschreiben, die jeder Eingangsfunktion u eine „Ausgangsfunktion“ $y \in \mathcal{Y}^{[t_0,T]}$ zuordnet, welche sämtliche Ausgangsgrößen im Intervall $[t_0, T]$ als Funktion der Zeit spezifiziert.

Definition 1.19. Die *Ein-Ausgangsfunktion* Ψ zum Ein-Ausgangsverhalten Λ ist die Abbildung

$$\begin{aligned} \Psi: \qquad & D(\lambda) \longrightarrow \bigcup_{t_0 \le T} \mathcal{Y}^{[t_0,T]} \\ \Psi(T, t_0, u)(t) \; & := \; \lambda(t; t_0, u|_{[t_0,t)}) \qquad \forall t \in [t_0, T] \; . \end{aligned} \tag{1.3}$$

Bemerkung 1.20. Die Responsefunktion lässt sich aus der Ein-Ausgangsfunktion zurückgewinnen, da

$$\lambda(t; t_0, u) = \Psi(t, t_0, u)(t) \; .$$

Kapitel 2

Bilineare Systeme

2.1 Existenz und Eindeutigkeit

Wir beschäftigen uns in dieser Arbeit mit kontinuierlichen endlich-dimensionalen Kontrollsystemen Σ, die durch eine Differentialgleichung der Form

$$\dot{x}(t) = f(x(t), u(t), t), \qquad f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n \tag{2.1}$$

modelliert werden, wobei $x(t) \in \mathbb{R}^n$ der Zustand und $u(t) \in \mathcal{U}$ der Kontrollwert zum Zeitpunkt t ist. Die Zustandsmenge $\mathcal{X}$ ist zunächst der gesamte $\mathbb{R}^n$ und der Steuerbereich $\mathcal{U}$ ist eine (nichtleere) Teilmenge des $\mathbb{R}^m$.

Definition 2.1. Wir nennen Σ ein *bilineares System*, falls die rechte Seite $f(x, u, t)$ von (2.1)

- affin-linear in x ist, d.h. für jedes Tupel $(u, t) \in \mathbb{R}^m \times \mathbb{R}$ gibt es eine Matrix $G(u, t) \in \mathbb{R}^{n \times n}$ und einen Translationsvektor $g(u, t) \in \mathbb{R}^n$, so dass

$$f(x, u, t) = G(u, t)x + g(u, t) \ , \tag{2.2}$$

- affin-linear in u ist, d.h. für jedes Tupel $(x, t) \in \mathbb{R}^n \times \mathbb{R}$ gibt es eine Matrix $H(x, t) \in \mathbb{R}^{n \times m}$ und einen Translationsvektor $h(x, t) \in \mathbb{R}^n$, so dass

$$f(x, u, t) = H(x, t)u + h(x, t) \ . \tag{2.3}$$

Systeme, welche durch rechte Seiten der Form (2.2) oder (2.3) bestimmt sind, nennt man auch *linear im Zustand* bzw. *linear in der Kontrolle*. Bilineare Systeme sind also linear im Zustand und in der Kontrolle—getrennt, aber nicht notwendigerweise gleichzeitig.

Das dynamische Verhalten eines bilinearen Systems über einem Zeitintervall I ist im Allgemeinen gegeben durch eine Differentialgleichung der Form

$$\dot{x}(t) = A(t)x(t) + \sum_{k=1}^{m} u_k(t)N_k(t)x(t) + B(t)u(t), \quad u(t) \in \mathcal{U}, \;\; t \in I \tag{2.4}$$

mit matrixwertigen Funktionen $A, N_1, \dots N_m : I \to \mathbb{R}^{n\times n}$ bzw. $B : I \to \mathbb{R}^{n\times m}$. Indem wir diese Funktionen durch Null auf $I^c = \mathbb{R}\backslash I$ fortsetzen und $u(t) := u_0$ für ein $u_0 \in \mathcal{U}$ und alle $t \in I^c$ setzen, könnten wir ohne Einschränkung $I = \mathbb{R}$ voraussetzen. Mit $u_k(t)$ bezeichnen wir die k-te Komponente von $u(t)$—entsprechend der folgenden Notation:

Bezeichnung 2.2. Für Vektoren $z \in \mathbb{R}^l$ verwenden wir die Schreibweise

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_l \end{pmatrix} .$$

Eine Matrix $Z \in \mathbb{R}^{l\times k}$ mit Spalten $z^1, z^2, \dots, z^k \in \mathbb{R}^l$ läßt sich folgendermaßen darstellen:

$$Z = (z^1|z^2|\dots|z^k)$$

Bemerkung 2.3. Offenbar ist die rechte Seite von (2.4)

$$f(x,u,t) = \left(A(t) + \sum_{k=1}^{m} u_k N_k(t) \right) x + B(t)u$$

affin-linear in x. Mit einer kleinen Umformung können wir auch die Linearität in der Kontrolle verdeutlichen:

$$\begin{aligned} f(x,u,t) &= A(t)x + \sum_{k=1}^{m} u_k N_k(t)x + B(t)u \\ &= A(t)x + (N_1(t)x|N_2(t)x|\dots|N_m(t)x)\cdot \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} + B(t)u \\ &= \underbrace{A(t)x}_{:=h(x,t)} + \underbrace{\big((N_1(t)x|N_2(t)x|\dots|N_m(t)x) + B(t)\big)}_{:=H(x,t)} u \end{aligned}$$

$f(x,u,t)$ genügt daher den Ansprüchen aus Definition 2.1.

Definition 2.4. Falls $B \equiv 0$ in der Gleichung (2.4), d.h.

$$\dot{x}(t) = \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) x(t), \quad u(t) \in \mathcal{U}, \;\; t \in I \;, \tag{2.5}$$

so sprechen wir von einem *homogenen* bilinearen System. Sind stattdessen die Funktionen $N_1, \ldots, N_m$ identisch Null, so wird durch $\dot{x} = Ax + Bu$ ein *lineares* Kontrollsystem gegeben.
Ein bilineares System ist autonom, wenn in der Systemgleichung (2.4) sämtliche matrixwertigen Funktionen konstant sind.

Bemerkung 2.5. Per Zustandsraumvergrößerung können wir jedes bilineare System, das durch (2.4) definiert ist, formal der Klasse der homogenen bilinearen Systeme zuordnen. Wir erhalten aus (2.4) die homogenen Form

$$\begin{pmatrix} \dot{x}(t) \\ \dot{\varphi}(t) \end{pmatrix} = \begin{pmatrix} A(t) & 0 \\ 0^* & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ \varphi(t) \end{pmatrix} + \sum_{k=1}^{m} u_k(t) \begin{pmatrix} N_k(t) & b^k(t) \\ 0^* & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ \varphi(t) \end{pmatrix} \;, \tag{2.6}$$

wobei $b^k(t)$ die k-te Spalte von $B(t)$ bezeichnet, mit neuer Zustandsvariable (x, φ) im vergrößerten Zustandsraum $\mathbb{R}^{n+1}$. Das Ausgangssystem taucht wieder in der Hyperebene $\varphi = 1$ auf. Auf diese Weise können wir uns in vielen Fällen auf das Studium der formal einfacheren homogenen Systeme beschränken.

Voraussetzung 2.6. Um die eindeutige Lösbarkeit von Anfangswertproblemen zur Differentialgleichung (2.4) sicherstellen zu können, fordern wir, dass die Funktionen $A(\cdot), N_1(\cdot), \ldots, N_m(\cdot), B(\cdot)$ messbar und lokal (essentiell) beschränkt sind, und dass die Kontrollfunktion $u : I \to \mathcal{U}$ zu einer Klasse von zulässigen Steuerfunktionen gehören.
Wir unterscheiden zwischen drei Klassen von zulässigen Steuerfunktionen:

(I) $\mathcal{U}_u^I$ ist die Klasse der messbaren, lokal (essentiell) beschränkten Funktionen von einem reellen Intervall I nach $\mathcal{U}$.

(II) $\mathcal{V}_r^I$ ist die Klasse der messbaren Funktionen auf I, deren Bilder in einer kompakten Teilmenge $\mathcal{V} \subseteq \mathcal{U}$ sind.

(III) $\mathcal{V}_b^I$ ist die Klasse der stückweise konstanten Funktionen auf I mit Werten in der kompakten Teilmenge $\mathcal{V} \subseteq \mathcal{U}$.

Die Buchstaben u, r und b stehen für „unrestricted“„restricted“ und „Bang-Bang“. Es gilt offenbar $\mathcal{U}_u^I \supseteq \mathcal{V}_r^I \supseteq \mathcal{V}_b^I$. Aussagen können also hierarchisch zwischen den Klassen vererbt werden.

Es wäre sehr lästig die soeben gestellte Voraussetzung in den folgenden Sätzen und Lemmas zu wiederholen. Deshalb verlangen wir, dass eine derartige Voraussetzung stets bis zum Ende des laufenden Abschnitts ihre Gültigkeit behält.

Definition 2.7. Es sei eine Klasse aus (I)-(III) vorgegeben. Eine messbare Abbildung $u : I \mapsto \mathcal{U}$ heißt *zulässige* Kontrollfunktion, falls u zu dieser Klasse gehört.

Lemma 2.8. Ist $u \in \mathcal{U}_u^I$, so erfüllt die Funktion

$$g : \mathbb{R}^n \times I \to \mathbb{R}^n$$
$$g(x,t) := A(t)x + \sum_{k=1}^{m} u_k(t)N_k(t)x + B(t)u(t)$$

die Caratheodory-Bedingungen (C1)-(C3) aus B.2 und die globale Lipschitzbedingung in x aus Folgerung B.16. In diesem Sinne ist (2.4) eine Caratheodory-Gleichung. Die Theorie aus Abschnitt B.1 lässt sich übertragen.

Beweis. 1. $g(x,t)$ ist affin-linear und somit stetig in x für alle t. Die Caratheodory-Bedingung (C1) ist also erfüllt.

2. Die Komponenten von $g(x,t)$ lassen sich als (endliche) Summen von Produkten darstellen, deren Faktoren Komponenten von

$$A(t), N_1(t), \ldots, N_m(t), B(t), u(t), x \tag{2.7}$$

sind. Nach Voraussetzung 2.6 sind die Komponenten der Funktionen aus (2.7) messbar. Natürlich sind auch Summen und Produkte messbarer reellwertiger Funktionen messbar, woraus die Messbarkeit von $g(x,t)$ in t für alle x folgt. Es gilt daher auch die Caratheodory-Bedingung (C2).

3. Da die globale Lipschitzbedingung in x erfüllt sein wird, genügt es anstelle von (C3) die abgeschwächte Bedingung (C3') aus Lemma B.13 nachzuweisen. Wir tun dies, indem wir zeigen, dass für jedes feste $x \in \mathbb{R}^n$ die Funktion $m(t) := \|g(x,t)\|$ messbar und lokal essentiell beschränkt ist auf I.
Dazu sei J irgendein beschränktes Teilintervall von I. Da die Funktionen aus (2.7) gemäß Voraussetzung 2.6 lokal essentiell beschränkt sind, existiert eine obere Schranke $M > 0$, so dass mit Hilfe der Dreiecksungleichung gilt

$$m(t) = \|g(x,t)\| \le \|A(t)\| \cdot \|x\| + \sum_{k=1}^{m} |u_k(t)| \cdot \|N_k(t)\| \cdot \|x\| + \|B(t)\| \cdot \|u(t)\|$$
$$\le M \qquad \forall' t \in J \ .$$

Es fehlt noch die Messbarkeit von $m(\cdot)$. Diese folgt aus der Messbarkeit von $t \mapsto g(x,t)$ und der Stetigkeit der Normfunktion.

4. Schließlich soll $g(x,t)$ noch die globale Lipschitzbedingung in x erfüllen, d.h. es gibt eine lokal integrierbare Funktion $\mu : I \to \mathbb{R}_+$, so dass

$$\|g(x,t) - g(y,t)\| \leq \mu(t)\|x - y\| \qquad \forall x, y \in \mathbb{R}^n, t \in I \ .$$

Dazu setze einfach

$$\mu(t) := \left\| A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right\| .$$

Analog zu 3. sehen wir, dass μ messbar und lokal (essentiell) beschränkt ist. Folglich ist μ eine lokal integrierbare Funktion und

$$\begin{aligned} \|g(x,t) - g(y,t)\| &= \left\| \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) \cdot (x - y) \right\| \\ &\leq \underbrace{\left\| A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right\|}_{\mu(t)} \cdot \|x - y\| \quad \forall x, y \in \mathbb{R}^n, t \in I \ . \quad \square \end{aligned}$$

Bemerkung 2.9. Für ein fest vorgegebenes $u \in \mathcal{U}_u^I$ ist durch die rechte Seite $g(x,t)$ ein inhomogenes lineares System ohne Kontrolle der Form (B.26) gegeben. Letztendlich leiten wir in diesem Abschnitt nur Existenz- und Eindeutigkeitsaussagen für solche linearen Systeme her. Die Ergebnisse sollten daher nicht überraschen.

Definition 2.10. Eine Funktion $x : J \to \mathbb{R}^n$ heißt *zulässige* Lösung von (2.4) auf $J \subseteq I$ bezüglich einer (zulässigen) Kontrolle $u : J \to \mathcal{U}$, falls $x(\cdot)$ eine Caratheodory-Lösung von (2.4) auf J gemäß Definition B.5 ist. Das heißt, eine der beiden äquivalenten Bedingungen wird von $x(\cdot)$ erfüllt:

(i) $x : J \to \mathbb{R}^n$ ist eine lokal absolut stetige Funktion mit

$$\dot{x}(t) = A(t)x(t) + \sum_{k=1}^{m} u_k(t) N_k(t) x(t) + B(t)u(t)$$

für fast alle $t \in J$.

(ii) Die Abbildung $t \mapsto A(t)x(t) + \sum_{k=1}^{m} u_k(t)N_k(t)x(t) + B(t)u(t)$ ist lokal integrierbar über J, und es gilt

$$x(t) = x(a) + \int_a^t A(s)x(s) + \sum_{k=1}^{m} u_k(s)N_k(s)x(s) + B(s)u(s)\,\mathrm{ds}$$

für alle $a, t \in J$. (Dabei sei $\int_a^t = -\int_t^a$ für $a > t$.)

Wir sagen, $x : J \to \mathbb{R}^n$ ist eine (zulässige) Lösung des Anfangswertproblems

$$\dot{x} = A(t)x + \sum_{k=1}^{m} u_k N_k(t)x + B(t)u, \quad t \in I, \;\; x(t_0) = x_0 \qquad \text{(AWP)}$$

auf dem Intervall $J \subseteq I$ bezüglich der zulässigen Kontrolle $u : I \to \mathcal{U}$, falls $x(\cdot)$ eine zulässige Lösung von (2.4) bzgl. der Restriktion $u|_J$ ist, welche den Bedingungen $t_0 \in J$ und $x(t_0) = x_0$ genügt.
Sie heißt maximal, wenn sie die folgende Eigenschaft besitzt:
Ist $\tilde{x} : \tilde{J} \to \mathbb{R}^n$ eine weitere Lösung von (AWP) auf $\tilde{J} \subseteq I$, so folgt $\tilde{J} \subseteq J$ und $x(t) = \tilde{x}(t)$ für alle $t \in \tilde{J}$.

Satz 2.11 (Existenz- und Eindeutigkeitssatz)**.** Es sei $u \in \mathcal{U}_u^I$. Dann existiert eine (eindeutige) maximale Lösung $\rho : I \to \mathbb{R}^n$ von (AWP), die auf ganz I definiert ist. Das bedeutet, falls $x(\cdot)$ eine zulässige Lösung von (AWP) auf einem Teilintervall $J \subseteq I$ ist, so folgt $x(t) = \rho(t)$ für alle $t \in J$.

Beweis. Da für ein vorgegebenes $u \in \mathcal{U}_u^I$ die Funktion

$$g(x,t) := A(t)x + \sum_{k=1}^{m} u_k(t)N_k(t)x + B(t)u(t)$$

den Caratheodory-Bedingungen und der globalen Lipschitzbedingung in x genügt (Lemma 2.8), sind alle Voraussetzungen von Folgerung B.16 erfüllt. Es existiert daher eine maximale Lösung $\rho : I \to \mathbb{R}^n$ von

$$\dot{x}(t) = g(x(t),t) = A(t)x(t) + \sum_{k=1}^{m} u_k(t)N_k(t)x(t) + B(t)u(t), \quad x(t_0) = x_0$$

auf ganz I. □

Zum Schluß soll der Zusammenhang unserer Definition eines bilinearen Systems mit dem axiomatischen Systembegriff aus Abschnitt 1.1 verdeutlicht werden.

Definition und Satz 2.12. Es sei $\mathcal{U} \subseteq \mathbb{R}^m$ eine nichtleere Menge. $\mathcal{X}$ sei entweder der gesamte $\mathbb{R}^n$ oder eine stark invariante Teilmenge des $\mathbb{R}^n$ gemäß Definition 3.11. Die Funktionen $A, N_1, \ldots, N_k : \mathbb{R} \to \mathbb{R}^{n\times n}$ und $B : \mathbb{R} \to \mathbb{R}^{n\times m}$ seien messbar und lokal (essentiell) beschränkt. Weiter sei eine Klasse zulässiger Kontrollfunktionen aus (I)-(III) vorgegeben.
Dann ist durch die Gleichung

$$\dot{x}(t) = A(t)x(t) + \sum_{k=1}^{m} u_k(t)N_k(t)x(t) + B(t)u(t), \quad x(t) \in \mathcal{X},\ u(t) \in \mathcal{U} \tag{2.8}$$

ein kontinuierliches bilineares System $\Sigma = (\mathbb{R}, \mathcal{X}, \mathcal{U}, \phi)$ gegeben mit einer Übergangsfunktion ϕ, die folgendermaßen definiert wird:
Definiere auf der Menge

$$D(\phi) := \big\{(t, t_0, x_0, u) \mid t_0 \le t, x_0 \in \mathcal{X}, u \in \mathcal{U}^{[t_0,t)} \text{ zulässig}\big\}$$

die Abbildung

$$\phi(t; t_0, x_0, u) := x(t; t_0, x_0, u)\ ,$$

wobei $x(\cdot\,; t_0, x_0, u)$ die (eindeutige) maximale Lösung von (AWP) bezüglich u auf dem Intervall $[t_0, t]$ sei.
In diesem Sinne darf man die Gleichung (2.8) als System bezeichnen.

Beweisskizze. Die Übergangsfunktion ist offensichtlich wohldefiniert.
Die Prüfung der Systemaxiome ist dem Leser überlassen.

2.2 Beispiele

Anhand einiger Beispiele soll die vielseitige Einsetzbarkeit von bilinearen Systemen zur Modellierung von Prozessen in der Physik, Biologie, Ökonomie, etc. verdeutlicht werden.

Beispiel 2.13 (Demographie)**.** Im Gegensatz zu linearen Systemen geht bei den bilinearen Systemen die Kontrollfunktion nicht nur additiv über den Term „Bu“, sondern auch multiplikativ über den Term „$\sum_{k=1}^{m} N_k u_k x$“ in die Systemgleichung ein. Das einfachste Beispiel ist gegeben durch

$$\dot{x}(t) = u(t)x(t)$$

mit skalaren Zustands- und Eingangsgrößen.
Wir können x als die Bevölkerungszahl und u als „Geburtenrate minus Sterberate plus Migrationsrate“ interpretieren, um die Demographie einer Region zu modellieren.

Die Bilineare Approximation ist eine naheliegende Verallgemeinerung der gewöhnlichen Linearisierung eines nichtlinearen Systems und basiert auf einer abgebrochenen Taylor-Entwicklung von der rechten Seite der Zustandsgleichung in einem Gleichgewicht.

Beispiel 2.14 (Bilineare Approximation)**.** In [25] wird die Bewegung eines geostationären Satelliten auf der Erdumlaufbahn anhand eines nichtlinearen Systems modelliert. Zur Vereinfachung wird dort vorausgesetzt, dass sich der Satellit stets direkt über dem Äquator befindet (konstanter Breitengrad 0°), d.h. er bewegt sich in der Äquatorebene, als deren Ursprung der Mittelpunkt der Erde gewählt wird. Die Position des Satelliten ist in den Polarkoordinaten (r, ρ) gegeben.

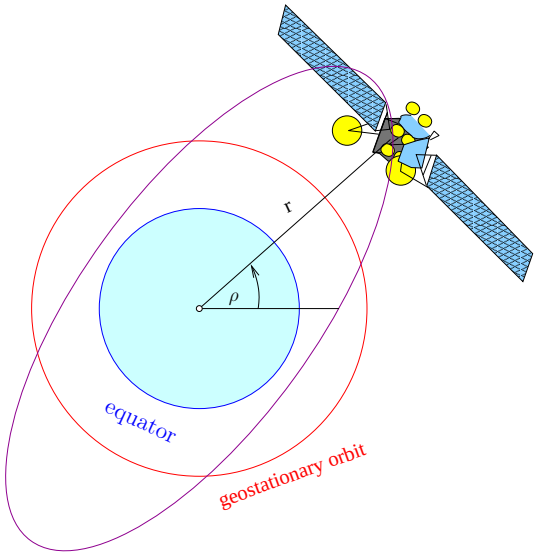

Abbildung 2.1: Satellit in Äquatorebene

Die folgenden Parameter werden berücksichtigt:

$$
\begin{aligned}
m_E &:= \text{Erdmasse}, \qquad m_S := \text{Masse des Satelliten},\\
g &:= \text{Gravitationskonstante der Erde},\\
\Omega &:= \text{Winkelgeschwindigkeit der Erde},\\
r_0 &:= \text{Radius des geostationären (kreisförmigen) Orbits},\\
\rho_0 + \Omega t &:= \text{Winkel, der von einer Antenne aus angepeilt wird.}
\end{aligned}
$$

Die Zustände des modellierten Satelliten sind

$$\begin{aligned} x_1(t) &:= r(t) - r_0 \ , \\ x_2(t) &:= \dot{r}(t) \ , \\ x_3(t) &:= \rho(t) - (\rho_0 + \Omega t) \ , \\ x_4(t) &:= \dot{\rho}(t) - \Omega \ . \end{aligned}$$

Man beachte, dass x_1 und x_3 die Abweichung der Position des Satelliten von der gewünschten Position im geostationären Orbits wiedergibt.
Weiter wird angenommen, dass am Satelliten zwei Düsenantriebe installiert sind, die den Satelliten in Richtung r bzw. ρ bewegen. Wir führen entsprechende Eingangsgrößen

$$\begin{aligned} u_1 &:= \text{Antriebskraft in Richtung } r \ , \\ u_2 &:= \text{Antriebskraft in Richtung } \rho \end{aligned}$$

ein.
Mit Hilfe der Newtonschen Gesetze kann man das dynamische Verhalten durch die nichtlineare Differentialgleichung

$$\left.\begin{array}{rc} \dot{x}_1(t) = & x_2(t) \\ \dot{x}_2(t) = & (x_1(t)+r_0)(x_4(t)+\Omega)^2 - \frac{g m_E}{(x_1(t)+r_0)^2} + \frac{u_1(t)}{m_S} \\ \dot{x}_3(t) = & x_4(t) \\ \dot{x}_4(t) = & -\frac{2x_2(t)(x_4(t)+\Omega)}{x_1(t)+r_0} + \frac{u_2(t)}{m_S(x_1(t)+r_0)} \end{array}\right\} =: f(x(t), u(t))$$

bestimmen.
Der Punkt $(x^*, u^*) = (0,0) \in \mathbb{R}^n \times \mathbb{R}^m$ ist ein Gleichgewicht, d.h. $f(x^*, u^*) = 0$. Solange sich das System in diesem Gleichgewicht befindet, kreist der Satellit auf dem geostationären Orbit genau in der Position, die von der Erde aus angepeilt wird. Eine optimale Kontrolle steuert daher den Satelliten in den Zustand $x^* = 0$, der ohne Kraftausübung ($u^* = 0$) „gehalten“ werden kann.
Die (komponentenweise) Taylorentwicklung der rechten Seite $f(x,u)$ im Gleichgewicht $(0,0)$ bis zur Ordnung 3 liefert für kleine x, u unter Berücksichtigung der Formel $\Omega^2 = \frac{m_E g}{{r_0}^3}$ das Ergebnis

$$f(x,u) \approx \begin{pmatrix} x_2 \\ 2\,\Omega x_1 x_4 + 2\,\Omega r_0\, x_4 + r_0 {x_4}^2 + 3\Omega^2 x_1 - 3\,\frac{m_E g {x_1}^2}{{r_0}^4} + \frac{u_1}{m_S} \\ x_4 \\ 2\,\frac{\Omega x_1 x_2}{{r_0}^2} - 2\,\frac{\Omega x_2}{r_0} - 2\,\frac{x_2 x_4}{r_0} - \frac{x_1 u_2}{{r_0}^2 m_s} + \frac{u_2}{r_0 m_s} \end{pmatrix} .$$

Wir streichen nun Summanden wie „$2\Omega x_1 x_4$“ oder „$r_0 {x_4}^2$“, die zu keinem bilinearen System gehören können. Auf diese Weise können wir für kleine x, u das Ausgangssystem durch das einfachere bilineare System

$$\dot{x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 3\Omega^2 & 0 & 0 & 2\Omega r_0 \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{2\Omega}{r_0} & 0 & 0 \end{pmatrix} x + u_2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{1}{{r_0}^2 m_S} & 0 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 & 0 \\ \frac{1}{m_S} & 0 \\ 0 & 0 \\ 0 & \frac{1}{m_S r_0} \end{pmatrix} u$$

„approximieren“.
Der Vorteil der bilinearen Approximation gegenüber der linearen ist, dass auch Terme der Form „$-\frac{1}{{r_0}^2 m_S} u_2 x_1$“, welche Produkte 1.Ordnung zwischen Zustands- und Eingangsgrößen enthalten, aus der Taylorentwicklung übernommen werden. Bei der linearen Approximation würden diese gestrichen werden. Die bilineare Approximation ist also näher am nichtlinearen Ausgangssystem.

Das nächste Beispiel wird auch in [9] und [22, p.96] studiert.

Beispiel 2.15 (Oszillator)**.** Ein harmonischer Oszillator ist ein physikalisches System, dessen Bestandteile harmonische Sinusschwingungen der Form

$$x(t) = \hat{x}\cos(\omega t + \rho), \qquad \hat{x}, \omega \geq 0, \quad \rho \in [0, 2\pi)$$

erzeugen. $\hat{x}$ nennt man Amplitude, ρ Nullphasenwinkel und ω Kreisfrequenz. $x(\cdot)$ ist die allgemeine Lösung der sog. Bewegungsgleichung

$$\frac{d^2 x(t)}{dt^2} + \omega^2 x(t) = 0 \ , \tag{2.9}$$

d.h. die Lösungen dieser DGL sind stets harmonische Sinusschwingungen, deren Parameter $\hat{x}$ und ρ durch Anfangswerte $(x(t_0), \dot{x}(t_0))$ eindeutig gegeben sind. Die Bewegungsgleichung ist nach Erweiterung des Zustandsraumes mittels $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} x \\ \dot{x} \end{smallmatrix}\right)$ durch die lineare Differentialgleichung

$$\begin{pmatrix} \dot{x_1} \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

ersetzbar. Auf diese Weise kann man einen harmonisches Oszillator als lineares System modellieren.
Ein typisches Beispiel für einen harmonischen Oszillator ist ein ungedämpfter LC-Schwingkreis. Das ist eine Baugruppe, bestehend aus einer Spule und einem Kondensator, die zueinander in Reihe geschaltet sind. Die Spule wird

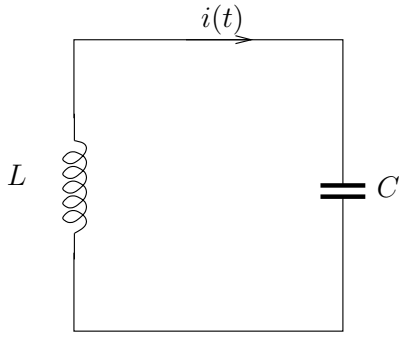

Abbildung 2.2: LC-Schwingkreis

durch die Induktivität L und der Kondensator durch die Kapazität C charakterisiert. Das Verhältnis zwischen L, C und der Stromstärke $i(t)$ läßt sich durch das physikalische Gesetz

$$L\frac{d}{dt}i(t) + \frac{1}{C}\int_0^t i(s)ds = 0$$

beschreiben. Differenzieren liefert die Bewegungsgleichung

$$\frac{d^2i(t)}{dt^2} + \underbrace{\frac{1}{LC}}_{:=\omega^2} i(t) = 0$$

für den offensichtlich sinusförmigen Wechselstrom $i(t)$.

Eine einfache Möglichkeit eine kontrollierte Oszillation zu erzeugen, ist es eine Schaltung zwischen zwei verschiedenen ungedämpften LC-Schwingkreisen zu installieren. Betrachte dazu das Netzwerk in Abbildung 2.3. Je nach Schalterstellung wählt man zwischen den Parametern L_1, C_1 und L_2, C_2. Bleibt der Schalter unberührt, so wird eine Sinuskurve mit Kreisfrequenz $\omega_k := \frac{1}{\sqrt{L_k C_k}}$ erzeugt $(k = 1, 2)$.
Dieses System kann durch das folgende bilineare Kontrollsystem mit Ausgang modelliert werden:

$$\begin{aligned} \begin{pmatrix} \dot{x_1} \\ x_2 \end{pmatrix} &= \underbrace{\begin{pmatrix} 0 & 1 \\ -\omega_1^2 & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + u \underbrace{\begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}}_{N} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} , \qquad (2.10) \\ y &= \begin{pmatrix} 1 & 0 \end{pmatrix} x \end{aligned}$$

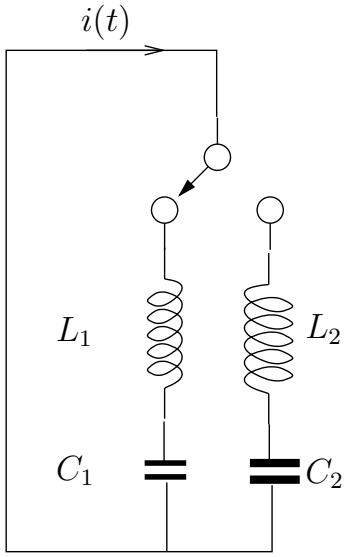

Abbildung 2.3: geschaltete LC-Schwingkreise

mit Zustandsraum $\mathcal{X} = \mathbb{R}^2$, Steuerbeschränkung $\mathcal{U} = \{\omega_2^2 - \omega_1^2, 0\}$ und Messbereich $\mathcal{Y} = \mathbb{R}$.

Wir werden in unserem letzten Beispiel ein bilineares System zur Modellierung eines Motors kennenlernen, das ausnahmsweise nicht autonom ist. Eine ausführlichere Beschreibung des Systems findet man in [22, p.98].

Beispiel 2.16. Wilson J. Rugh modelliert einen Gleichstrommotor, dessen Anker- und Erregerwicklung durch separable Kreisläufe mit Strom versorgt werden, anhand der folgenden Differentialgleichungen:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}} i_A(t) &= -\frac{R_A}{L_A} i_A(t) - \frac{K}{L_A} i_E(t)\omega(t) + \frac{1}{L_A} u_A(t) \ , \\ \frac{\mathrm{d}}{\mathrm{dt}} \omega(t) &= \frac{K}{J} i_E(t) i_A(t) - \frac{B}{J}\omega(t)\end{aligned}$$

Dabei ist

i_A, i_E der Anker- bzw. Erregerstrom,

u_A, u_E die Anker- bzw. Erregerspannung,

L_A, L_E die Induktivität der Anker- bzw. Erregerwicklung,

R_A, R_E der Widerstand der Anker- bzw. Erregerwicklung,

ω die Winkelgeschwindigkeit des Ankers,

J das Massenträgheitsmoment des Ankers.

K, B sind positive Konstanten, die in [22, p.98] charakterisiert werden.

Die Geschwindigkeit des Motors wird kontrolliert, indem man die Ankerspannung konstant hält (d.h. $u_A(t) \equiv U_A$) und den Erregerstrom i_E durch einen variablen Widerstand in der Erregerwicklung steuert.
Wir führen den Zustandsvektor $x(t) := \begin{pmatrix} i_A(t) \\ \omega(t) \end{pmatrix}$, die Eingangsgröße $u(t) := i_E(t)$ und die Ausgangsgröße $y(t) := \omega(t)$ ein, und erhalten basierend auf den Differentialgleichungen das bilineare System

$$\dot{x}(t) = \underbrace{\begin{pmatrix} -\frac{R_A}{L_A} & 0 \\ 0 & -\frac{B}{J} \end{pmatrix}}_{A} x(t) + u(t) \underbrace{\begin{pmatrix} 0 & -\frac{K}{L_A} \\ \frac{K}{J} & 0 \end{pmatrix}}_{N} x(t) + \underbrace{\begin{pmatrix} \frac{U_A}{L_A} \\ 0 \end{pmatrix}}_{d} ,$$
$$y(t) = \begin{pmatrix} 0 & 1 \end{pmatrix} x(t) .$$

Um die rechte Seite auf die Standardform (2.4) zu bringen, müssen wir den konstanten Term d beseitigen. Dies geschieht folgendermaßen:
Zunächst berechnen wir die Lösung $x_c(\cdot)$ zur Kontrolle $u \equiv 0$ mit $x_c(0) = 0$, d.h. x_c ist absolut stetig und

$$\dot{x}_c = Ax_c + d \quad \text{f.ü.}, \qquad x_c(0) = 0 .$$

In unserem Fall ergibt Variation der Konstanten B.27

$$\begin{aligned} x_c(t) &= \int_0^t e^{A(t-s)} d \, \mathrm{ds} = \int_0^t \begin{pmatrix} e^{-\frac{R_A}{L_A}(t-s)} & 0 \\ 0 & e^{-\frac{B}{J}(t-s)} \end{pmatrix} \cdot \begin{pmatrix} \frac{U_A}{L_A} \\ 0 \end{pmatrix} \mathrm{ds} \\ &= \int_0^t \begin{pmatrix} e^{-\frac{R_A}{L_A}(t-s)} \frac{U_A}{L_A} \\ 0 \end{pmatrix} \mathrm{ds} = \begin{pmatrix} e^{-\frac{R_A}{L_A}t} \frac{U_A}{L_A} \cdot \left(\frac{L_A}{R_A} e^{\frac{R_A}{L_A}t} - \frac{L_A}{R_A} \right) \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{U_A}{R_A} \left(1 - e^{-\frac{R_A}{L_A}t} \right) \\ 0 \end{pmatrix} . \end{aligned}$$

Als Nächstes setzen wir $z(t) := x(t) - x_c(t)$ und berechnen die Systemgleichung bezüglich $z(t)$:

$$\begin{aligned} \dot{z}(t) &= Ax + uNx + d - Ax_c - d = Az + uNz - uNx_c \\ &= \begin{pmatrix} -\frac{R_A}{L_A} & 0 \\ 0 & -\frac{B}{J} \end{pmatrix} z(t) + u(t) \begin{pmatrix} 0 & -\frac{K}{L_A} \\ \frac{K}{J} & 0 \end{pmatrix} z(t) + \begin{pmatrix} 0 \\ \frac{KU_A}{JR_A} \left(1 - e^{-\frac{R_A}{L_A}t} \right) \end{pmatrix} u(t) , \\ y(t) &= \begin{pmatrix} 0 & 1 \end{pmatrix} z(t), \qquad z(0) = \begin{pmatrix} i_A(0) \\ \omega(0) \end{pmatrix} \end{aligned}$$

Wir stellen fest, dass das umgeformte System zwar bilinear und in Standardform ist—aber nicht autonom.

Bemerkung 2.17. Eine weitere Anwendung aus der Physik ist die Modellierung der Kernspaltung und des Kühlsystems in einem Kernreaktor durch bilineare Systeme. In [19] wird das Verhalten dieser Systeme optimiert.
Das Massenwirkungsgesetz aus der Chemie ist die Grundlage für ein bilineares System in [18], das die Regulierung des Hormons T_4 im menschlichen Körper modelliert.
In [5, p.335] wird ein bilineares Modell zur Modellierung des Wachstums einer Volkswirtschaft vorgeschlagen. Dies zeigt, dass bilineare Systeme auch in der Ökonomie von Bedeutung sind.
Zahlreiche weitere bilineare Modelle sind in [18] und [5] aufgelistet.

2.3 Homogene bilineare Systeme

Betrachte das homogene System (2.5) im $\mathbb{R}^n$

$$\dot{x}(t) = \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) x(t), \quad t \in I, \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathcal{U} \tag{2.11}$$

und das „assoziierte Matrixsystem“ im $\mathbb{R}^{n\times n}$

$$\dot{X}(t) = \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) X(t), \quad t \in I, \quad X(t) \in \mathbb{R}^{n\times n}, \quad u(t) \in \mathcal{U} \tag{2.12}$$

mit messbaren, lokal (essentiell) beschränkten matrixwertigen Funktionen $A, N_1, \ldots, N_m : I \to \mathbb{R}^{n\times n}$ auf einem Intervall I und einer nichtleeren Teilmenge $\mathcal{U}$ des $\mathbb{R}^m$.

Sprechweise 2.18. Wir nennen (2.12) das „assoziierte Matrixsystem“ von (2.11), und (2.11) das „assoziierte Vektorsystem“ von (2.12).

Für ein vorgegebenes $u \in \mathcal{U}_u^I$ ist auch die Funktion

$$U(t) := A(t) + \sum_{k=1}^{m} u_k(t) N_k(t)$$

messbar und lokal (essentiell) beschränkt (siehe Beweis von 2.8). Folglich gehören die Systeme

$$\dot{x}(t) = U(t)x(t), \qquad t \in I, \qquad x(t) \in \mathbb{R}^n \tag{2.13}$$

und

$$\dot{X}(t) = U(t)X(t), \qquad t \in I, \qquad X(t) \in \mathbb{R}^{n\times n} \tag{2.14}$$

zur Klasse der linearen Systeme ohne Kontrolle, die ausführlich im Abschnitt B.2 studiert wird. Wir werden jetzt einige Aussagen von dort übernehmen.

Definition 2.19. Die Fundamentallösung von (2.14)—man könnte sich auch auf (2.13) beziehen—bezüglich einer vorgegebenen Kontrolle $u \in \mathcal{U}_u^I$ und der Anfangszeit $t_0 \in I$ bezeichnen wir mit $\Phi(\cdot\,; t_0, u)$. Sie ist die maximale (zulässige) Lösung des Anfangswertproblems

$$\dot{X}(t) = \underbrace{\left(A(t) + \sum_{k=1}^{m} u_k(t)N_k(t)\right)}_{=U(t)} X(t), \qquad t \in I, \qquad X(t_0) = \mathrm{I} \quad \text{(AWP)}$$

zur Kontrolle u, definiert auf ganz I.

Folgerung 2.20. Aus Satz B.18 erhalten wir die folgenden Aussagen:

(i) $X : t \mapsto \Phi(t; t_0, u)\cdot P$ ist die maximale Lösung von (2.12) bzgl. der Kontrolle $u \in \mathcal{U}_u^I$, welche die Anfangsbedingung $X(t_0) = P$ erfüllt.

(ii) $x : t \mapsto \Phi(t; t_0, u)\cdot p$ ist die maximale Lösung von (2.11) bzgl. der Kontrolle $u \in \mathcal{U}_u^I$, welche die Anfangsbedingung $x(t_0) = p$ erfüllt.

Für die nächste Folgerung sei o.B.d.A. $I = \mathbb{R}$ vorausgesetzt.

Folgerung 2.21. Es gilt für alle $t, t_0, t_1 \in \mathbb{R}$ und $u \in \mathcal{U}_u^{\mathbb{R}}$:

(i) $\Phi(t_0; t_0, u) = \mathrm{I}$.

(ii) Die Matrix $\Phi(t; t_0, u)$ hat vollen Rang.

(iii) $\Phi(t; t_0, u) = \Phi(t; t_1, u)\cdot \Phi(t_1; t_0, u)$.

(iv) $\Phi(t_1; t_0, u)^{-1} = \Phi(t_0; t_1, u)$.

(v) $\det \Phi(t; t_0, u) = \exp\left(\int_{t_0}^{t} \mathrm{tr}U(s)\mathrm{d}s\right)$.

(vi) $\Phi(t;t_0,u) = \mathrm{I} + \sum_{k=1}^{\infty} \int_{t_0}^{t} \int_{t_0}^{s_1} \dots \int_{t_0}^{s_{k-1}} U(s_1)\dots U(s_k)\mathrm{ds_k}\dots\mathrm{ds_2ds_1}$.

(vii) $\Phi(t;t_0,u) = \lim_{l\to\infty} Y_l(t)$, wobei $Y_0 \equiv I$ und
$Y_{l+1}(t) = I + \int_{t_0}^{t} U(s)Y_l(s)\mathrm{ds} \quad (l = 0,1,2,\dots)$.

(viii) $\frac{\mathrm{d}}{\mathrm{d}t_0}\Phi(t;t_0,u) = -\Phi(t;t_0,u)A(t_0)$ fast überall.

(Dabei sei $\int_{t_0}^{t} = -\int_{t}^{t_0}$ für $t_0 > t$.)

Beweis. Die aufgezählten Eigenschaften von $\Phi(t;t_0,u)$ werden in Lemma B.19 und Satz B.22 bewiesen.

Bemerkung 2.22. Um eine Lösungsdarstellung für inhomogene Systeme zu erhalten, verwenden wir Variation der Konstanten B.27. Die maximale Lösung $x : I \to \mathbb{R}^n$ vom Anfangswertproblem

$$\dot{x}(t) = A(t)x(t) + \sum_{k=1}^{m} u_k(t)N_k(t)x(t) + B(t)u(t), \quad t \in I, \quad x(t_0) = p$$

bzgl. $u \in \mathcal{U}_u^I$ ist gegeben[1] durch

$$x(t) = \Phi(t;t_0,u)p + \int_{t_0}^{t} \Phi(t;s,u)B(s)u(s)\mathrm{ds} \ . \tag{2.15}$$

Alternativ könnten wir $x(\cdot)$ über die Fundamentallösung $\tilde{\Phi}(\cdot;t_0,u)$ des homogenisierten Systems (2.6) bestimmen. Es ist nämlich

$$\begin{pmatrix} x(t) \\ 1 \end{pmatrix} = \tilde{\Phi}(t;t_0,u)\cdot \begin{pmatrix} p \\ 1 \end{pmatrix} \ .$$

Autonome Systeme

Falls sämtliche Matrixfunktionen in (2.5) konstant sind, so ergibt sich die Systemgleichung

$$\dot{x}(t) = \left(A + \sum_{k=1}^{m} u_k(t)N_k \right) x(t), \quad t \in I, \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathcal{U} \ . \tag{2.16}$$

O.B.d.A. sei $I = \mathbb{R}$.

Bezeichnung 2.23. In Anlehnung an Definition 1.9 setzen wir
$u^s(t) := u(t-s)$ für ein $s \in \mathbb{R}$ und eine zulässige Kontrollfunktion u. Da der Steuerbereich $\mathcal{U}$ zeitunabhängig ist, ist auch die die zeitlich verschobene Kontrolle $u^s(\cdot)$ zulässig.

[1] $B : I \to \mathbb{R}^{n\times m}$ messbar und lokal (essentiell) beschränkt

Lemma 2.24. Die Fundamentallösungen von (2.16) erfüllen:

$$\Phi(t; t_0, u) = \Phi(t + s; t_0 + s, u^s) \quad \forall t, s, t_0 \in \mathbb{R}, u \in \mathcal{U}_u^{\mathbb{R}}$$

Insbesondere ist durch (2.16) ein autonomes System gemäß Definition 1.9 bestimmt.

Beweis. Für vorgegebene $s, t_0 \in \mathbb{R}$ ist durch $Y(t) := \Phi(t + s; t_0 + s, u^s)$ eine Lösung von (AWP) gegeben. Denn es gilt

$$Y(t_0) = \Phi(t_0 + s; t_0 + s, u^s) = \mathrm{I}$$

und (mittels Kettenregel)

$$\begin{aligned}\dot{Y}(t) &= \frac{\mathrm{d}}{\mathrm{dt}}\Phi(t + s; t_0 + s, u^s) = \left(A + \sum_{k=1}^{m} u_k^s(t + s)N_k\right) \cdot \Phi(t + s; t_0 + s, u^s) \\ &= \left(A + \sum_{k=1}^{m} u_k(t)N_k\right) Y(t) \qquad \text{für fast alle } t \in \mathbb{R}.\end{aligned}$$

Es folgt $\Phi(t + s; t_0 + s, u^s) = Y(t) = \Phi(t; t_0, u) =$ für alle $t \in \mathbb{R}$. □

Bezeichnung 2.25. Die Autonomie des Systems erlaubt es ohne Einschränkung die Anfangszeit $t_0 := 0$ zu wählen. Wir schreiben daher kurz $\Phi(t; u)$ statt $\Phi(t; 0, u)$.
Mit $\Phi(t; c)$ für ein $c \in \mathbb{R}$ bezeichnen wir die Fundamentallösung bzgl. der konstanten Kontrollfunktion $u \equiv c$. (Bitte nicht mit der Notation $\Phi(t; t_0)$ aus Abschnitt B.2 verwechseln!)

Folgerung 2.26. Die Rechenregel (iii) aus Lemma 2.21 wird zu

$$\Phi(t; u) = \Phi(t - s; u^{-s}) \cdot \Phi(s; u) \qquad \forall t, s \in \mathbb{R}\ . \tag{2.17}$$

Beweis. Wir rechnen nach:

$$\Phi(t - s; u^{-s}) \cdot \Phi(s; u) = \underbrace{\Phi(t - s; 0, u^{-s})}_{\Phi(t;s,u)} \cdot \Phi(s; 0, u) = \Phi(t; 0, u) = \Phi(t; u) \quad \square$$

Die Formel (2.17) ist besonders hilfreich, wenn wir die Lösungen zu stückweise konstanten Kontrollen berechnen wollen, wie das folgende Beispiel zeigt.

Beispiel 2.27. Der kontrollierte Oszillator aus Beispiel 2.15 ist modelliert durch das autonome System

$$\dot{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} = \underbrace{\begin{pmatrix} 0 & 1 \\ -\omega_1^2 & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + u \underbrace{\begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}}_{N} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ,$$
$$y = \begin{pmatrix} 1 & 0 \end{pmatrix} x$$

mit $\mathcal{X} = \mathbb{R}^2$, $\mathcal{U} = \{\omega_2^2 - \omega_1^2, 0\}$ und $\mathcal{Y} = \mathbb{R}$. Die beiden Frequenzen seien $\omega_1 = 0.5$ und $\omega_2 = 1.5$.
Wir wollen heuristisch eine Bang-Bang Steuerung $u \in \mathcal{U}_b^I$ herleiten, welche den Ursprung des Zustandsraum ansteuert (d.h. $\|x\| \to 0$) und somit die Amplitude der erzeugten Sinuskurve auf dem Intervall $I = [0, 10]$ verringert. (In [9, p.4ff] findet der Leser eine ausführlichere Argumentation.) Dazu berechnen wir zunächst die Fundamentallösungen zu den trivialen Steuerungen $u \equiv 0$ bzw. $u \equiv \omega_2^2 - \omega_1^2 = 2$:

$$\Phi(t;0) = e^{At} = \begin{pmatrix} \cos(0.5t) & 2\sin(0.5t) \\ -0.5\sin(0.5t) & \cos(0.5t) \end{pmatrix} ,$$
$$\Phi(t;2) = e^{(A+2N)t} = \begin{pmatrix} \cos(1.5t) & \frac{2}{3}\sin(1.5t) \\ -1.5\sin(1.5t) & \cos(1.5t) \end{pmatrix} .$$

Die Multiplikation von rechts mit Anfangswerten $p = \binom{0}{1}, \binom{0}{2}, \binom{0}{3}$ ergibt die Lösungskurven in Abbildung 2.4. Dabei gehören die roten Kurven zu

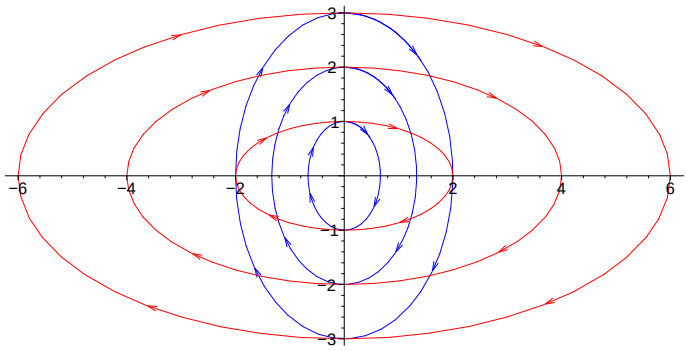

Abbildung 2.4: Phasenpotrait

$u = 0$ und die Blauen zu $u = 2$. Die Kurven bewegen sich mit konstanter Winkelgeschwindigkeit im Uhrzeigersinn und benötigen $t = 4\pi$ (rot) bzw. $t = \frac{4}{3}\pi$ (blau) für eine komplette Rotation. Anscheinend reduziert $u = 2$ im ersten und dritten Quadranten die Norm $\|x\|$ und $u = 0$ im zweiten und

vierten Quadranten. Dies motiviert die folgende Strategie:
Angenommen wir starten im Anfangszustand $p = \binom{0}{1}$ zum Zeitpunkt $t_0 = 0$. Dann wählen wir solange die Steuerung $u = 2$ bis wir die x-Achse schneiden—also bis zum Zeitpunkt $t_1 = \frac{\pi}{3}$. Danach schalten wir auf $u = 0$ um. Die Kurve bewegt sich nun im 2.Quadranten und schneidet die y-Achse zum Zeitpunkt $t_2 = t_1 + \pi = \frac{4}{3}\pi$. Dann schalten wir wieder auf $u = 2$ u.s.w. Die Schaltstellen sind genau die Koordinatenachsen. Wir erhalten schließlich die Steuerung

$$u(t) = \begin{cases} 2\ , & 0 \le t \le \frac{1}{3}\pi \\ 0\ , & \frac{1}{3}\pi < t \le \frac{4}{3}\pi \\ 2\ , & \frac{4}{3}\pi < t \le \frac{5}{3}\pi \\ 0\ , & \frac{5}{3}\pi < t \le \frac{8}{3}\pi \\ 2\ , & \frac{8}{3}\pi < t \le 3\pi \\ 0 & 3\pi < t \le 10 \end{cases} .$$

Die zugehörige Lösung $x : [0, 10] \to \mathbb{R}^n$ berechnen wir anhand (2.17). Es ist

$$x(t) = \begin{cases} \Phi(t;2)\cdot p\ , & 0 \le t \le \frac{1}{3}\pi \\ \Phi(t-\frac{\pi}{3};0)\cdot\Phi(\frac{\pi}{3};2)\cdot p\ , & \frac{1}{3}\pi < t \le \frac{4}{3}\pi \\ \Phi(t-\frac{4}{3}\pi;2)\cdot x(\frac{4}{3}\pi)\ , & \frac{4}{3}\pi < t \le \frac{5}{3}\pi \\ \Phi(t-\frac{5}{3}\pi;0)\cdot x(\frac{5}{3}\pi)\ , & \frac{5}{3}\pi < t \le \frac{8}{3}\pi \\ \Phi(t-\frac{8}{3}\pi;2)\cdot x(\frac{8}{3}\pi)\ , & \frac{8}{3}\pi < t \le 3\pi \\ \Phi(t-3\pi;0)\cdot x(3\pi)\ , & 3\pi < t \le 10 \end{cases} .$$

Die Abbildung 2.5 zeigt den Verlauf der Lösungskurve im Zustandsraum.

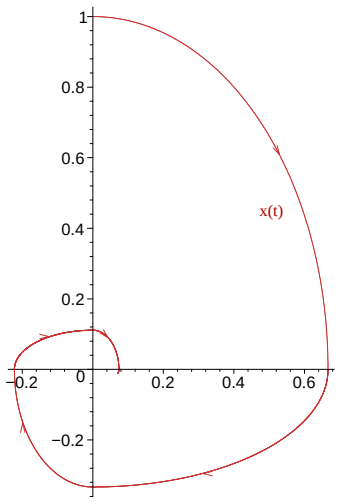

Abbildung 2.5: Lösungstrajektorie

Der Verlauf von $y(t) = x_1(t)$ ist in Abbildung 2.6 dargestellt. Offenbar haben

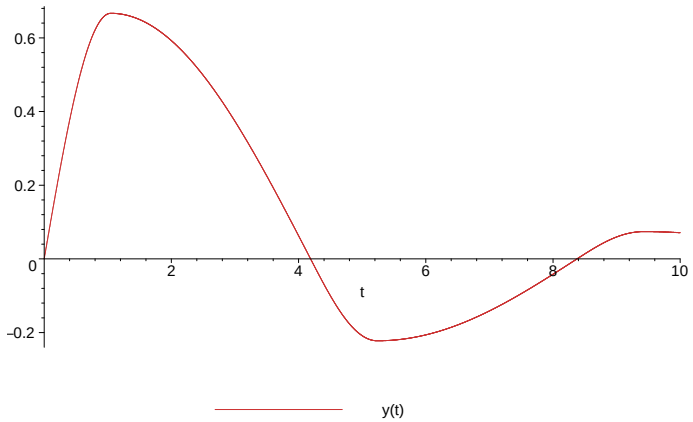

Abbildung 2.6: Ausgangsfunktion

wir uns wie geplant dem Ursprung angenähert. (Wie sich später zeigen wird, ist $(0,0)$ nicht erreichbar.)

Systeme mit quasikommutativen Matrizen

In diesem Teilabschnitt nehmen wir an, dass die Matrizen des autonomen homogenen Systems

$$\dot{x}(t) = \left(A + \sum_{k=1}^{m} u_k(t) N_k \right) x(t), \quad t \in I, \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathcal{U} \tag{2.18}$$

die Bedingung für Quasikommutativität[2] erfüllen, d.h.

Voraussetzung 2.28. Die Matrizen $A, N_1, \ldots, N_m \in \mathbb{R}^{n\times n}$ erfüllen

$$[\mathrm{ad}_A^k N_i, N_j] = 0 \qquad \forall i,j = 1,\ldots,m \; \forall k = 0,1,\ldots,n^2-1 \; .$$

Dabei verwenden wir die Notation

$$\mathrm{ad}_A^0 N = N, \mathrm{ad}_A^1 N = [A,N], \mathrm{ad}_A^2 N = [A,[A,N]], \mathrm{ad}_A^k N = [A, \mathrm{ad}_A^{k-1} N] \; .$$

(Mit $[\cdot,\cdot]$ bezeichnen wir die Lie-Klammer des $\mathbb{R}^{n\times n}$, d.h. $[A,N] = AN - NA$.)

Satz 2.29. Wir setzen $Q_k(t) := e^{-At} N_k e^{At}$ für alle $k \in \{1,\ldots,m\}$. Dann ist

$$\Phi(t;u) = e^{At} \cdot \exp\left(\int_0^t \sum_{k=1}^{m} u_k(s) Q_k(s) \mathrm{ds} \right) \tag{2.19}$$

die Fundamentallösung von (2.18) zu der Kontrolle $u \in \mathcal{U}_u^I$.

[2] Begriff stammt aus [14]

Beweis. 1. Die Baker-Hausdorff Gleichung [20, p.14] liefert die Potenzreihen-Entwicklung

$$e^{-At}Ne^{At} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{ad}_A^k N \ . \tag{2.20}$$

(Insbesondere lösen die beiden Seiten der Gleichung dieselbe DGL.)

2. Da $(A, N) \mapsto [A, N]$ eine beschränkte Bilinearform ist, gilt

$$[e^{-At}N_ie^{At}, N_j] \overset{1.}{=} \sum_{k=0}^{\infty} \frac{t^k}{k!} [\mathrm{ad}_A^k N_i, N_j] \ . \tag{2.21}$$

Nach Voraussetzung ist $[\mathrm{ad}_A^k N_i, N_j] = 0$ für alle $i, j \in \{1, \ldots, m\}$ und $k \in \{0, 1 \ldots n^2 - 1\}$. Allerdings ist $\mathrm{ad}_A : N \mapsto [A, N]$ ein linearer Operator von einem n^2-dimensionalen Vektorraum in sich selbst, und wir dürfen aus dem Satz von Cayley-Hamilton folgern, dass jede Potenz $\mathrm{ad}_A^k = \mathrm{ad}_A \circ \mathrm{ad}_A \circ \ldots \mathrm{ad}_A$ mit k größer $n^2 - 1$ eine $\mathbb{R}$-Linearkombination der ersten $n^2 - 1$ Potenzen ist. Daher ist $[\mathrm{ad}_A^k N_i, N_j] = 0$ für alle $k \geq 0$, und $[e^{-At}N_ie^{At}, N_j]$ ist identisch Null wegen (2.21).

3. Es ist

$$0 \overset{2.}{=} [e^{-At}N_ie^{At}, N_j] = e^{-At}N_ie^{At}N_j - N_je^{-At}N_ie^{At}$$

und folglich

$$0 = e^{-A\sigma}e^{-At}N_ie^{At}N_je^{A\sigma} - e^{-A\sigma}N_je^{-At}N_ie^{At}e^{A\sigma} \qquad \forall \sigma, t \in \mathbb{R} \ .$$

Es sei $t_1 = t + \sigma$ und $t_2 = \sigma$. Dann erhalten wir

$$\begin{aligned} 0 &= e^{-At_1}N_ie^{A(t_1-t_2)}N_je^{At_2} - e^{-At_2}N_je^{A(t_2-t_1)}N_ie^{At_1} \\ &= [e^{-At_1}N_ie^{At_1}, e^{-At_2}N_je^{At_2}] = [Q_i(t_1), Q_j(t_2)] \qquad \forall t_1, t_2 \in \mathbb{R} \ . \end{aligned}$$

4. Für ein vorgegebenes $u \in \mathcal{U}_u^I$ sei $\Phi_u(\cdot\,) := \Phi(\cdot\,; u)$ die Fundamentallösung von (2.18). Dann ist $Z(t) := e^{-At}\Phi_u(t)$ absolut stetig und erfüllt

$$\begin{aligned} \dot{Z}(t) &= -Ae^{-At}\Phi_u(t) + e^{-At}\dot{\Phi}_u(t) \\ &= -Ae^{-At}\Phi_u(t) + \left(e^{-At}A + \sum_{k=1}^{m} u_k(t)e^{-At}N_k \right) \Phi_u(t) \\ &= \underbrace{\left(\sum_{k=1}^{m} u_k(t)e^{-At}N_ke^{At} \right)}_{:=A(t)} Z(t) = A(t)Z(t) \end{aligned} \tag{2.22}$$

fast überall. Weiter ist

$$\begin{aligned}[A(t_1), A(t_2)] &= \Big[\sum_{i=1}^{m} u_i(t_1)Q_i(t_1), \sum_{j=1}^{m} u_j(t_2)Q_j(t_2)\Big] \\ &= \sum_{i,j=1}^{m} u_i(t_1)u_j(t_2)[Q_i(t_1), Q_j(t_2)] \overset{3.}{=} 0 \qquad \forall t_1, t_2 \in [0,T] \ .\end{aligned}$$

Schließlich wenden wir den Satz B.20 an und sehen, dass

$$Z(t) = \exp \int_{t_0}^{t} A(s)\mathrm{ds} = \exp \int_{t_0}^{t} \left(\sum_{k=1}^{m} u_k(s)Q_k(s) \right) \mathrm{ds} \ .$$

Aus $\Phi_u(t) = e^{At}Z(t)$ gewinnen wir die erhoffte Darstellung von $\Phi(\cdot\,; u)$. □

Beispiel 2.30. 1. Es sei $m = 1$ in Satz 2.29. Falls die Matrizen A, N kommutieren ($AN = NA$), so kommutiert auch die Exponentialmatrix e^{At} mit N. Also ist $Q(t) \equiv N$ und (2.19) wird zu

$$\Phi(t; u) = e^{At} \cdot e^{N \int_0^t u(s)\mathrm{ds}} = e^{At + N \int_0^t u(s)\mathrm{ds}} \ . \tag{2.23}$$

2. Für inhomogene Single-Input Systeme im $\mathbb{R}^n$

$$\dot{x}(t) = Ax(t) + u(t)Nx(t) + cu(t), \qquad t \in [0,T] \ ,$$

mit kommutativen Matrizen $A, N \in \mathbb{R}^{n\times n}$ und einem Vektor $c \in \mathbb{R}^n$ erhalten wir durch Variation der Konstanten eine explizite Lösungsdarstellung. Aus der Formel (2.15) folgt, dass

$$x(t) = e^{At + N \int_0^t u(s)\mathrm{ds}} x(0) + \int_0^t e^{A(t-s) + N \int_s^t u(\tau)\mathrm{d}\tau} cu(s)\mathrm{ds}$$

die Lösung zur Kontrolle u auf dem Intervall $[0,T]$ ist.

Bemerkung 2.31. In den Artikeln [14] und [15] werden bilineare Systeme mit quasikommutativen Matrizen, ausgehend von den eben berechneten Lösungsdarstellungen, analysiert.

Bilineare Systeme von Rang 1

Der Mathematiker Otomar Hàjek hat sich in seinen Arbeiten [9, p.146-167] und [10] mit speziellen Typen von kontinuierlichen bilinearen Systemen

beschäftigt, die zusammen die Klasse der „Bilinearen Systeme von Rang 1“ bilden. Er unterscheidet zwischen Systeme im $\mathbb{R}^n$ der Form

$$\dot{x} = (A + uc^*)x, \qquad u(t) \in \mathcal{U} \tag{2.24}$$

und

$$\dot{x} = (A + bu^*)x, \qquad u(t) \in \mathcal{U} \tag{2.25}$$

mit Parametern $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}^n$, und einem Steuerbereich $\mathcal{U} \subseteq \mathbb{R}^n$. Dem Aussehen der Kontrolle u entsprechend, bezeichnet er (2.24) als „Spalten-Kontrollsystem“ und (2.25) als „Zeilen-Kontrollsystem“. Single-Input Systeme der Form

$$\dot{x} = (A + ubc^*)x, \qquad -1 \leq u(t) \leq 1 \tag{2.26}$$

können sowohl den Spalten- als auch den Zeilen-Kontrollsystemen zugeordnet werden. Insbesondere sind die Matrizen uc^*, bu^* und bc^* höchstens vom Rang 1.
Die spezielle Struktur der Systeme von Rang 1 erlaubt es, theoretische Aussagen herzuleiten, die für allgemeinere bilineare Systeme in der Regel nicht gelten (Stichworte: Bang-Bang Prinzip, Konvexität der erreichbaren Menge). Andererseits ist die Unterklasse groß genug, um interessante und relevante Beispiele zu liefern:

Beispiel 2.32 (Parametersteuerung)**.** Angenommen, ein dynamisches System ist durch eine lineare skalare Differentialgleichung n-ter Ordnung

$$x^{(n)} = \sum_{k=0}^{n-1} \alpha_k x^{(k)} \tag{2.27}$$

gegeben, wobei die Parameter $\alpha_0, \ldots, \alpha_{n-1}$ Funktionen der Zeit t seien. (Hier bezeichnet $x^{(k)}$ die k-te Ableitung der Funktion x.) Durch Matrixschreibweise

können wir das System in ein lineares System 1.Ordnung im $\mathbb{R}^n$ überführen:

$$\begin{pmatrix} \dot{x}^{(0)} \\ \dot{x}^{(1)} \\ \vdots \\ \dot{x}^{(n-1)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & 1 \\ \alpha_0 & \alpha_1 & \dots & \alpha_{n-1} \end{pmatrix} \cdot \underbrace{\begin{pmatrix} x^{(0)} \\ x^{(1)} \\ \vdots \\ x^{(n-1)} \end{pmatrix}}_{x}$$

$$= \underbrace{\begin{pmatrix} 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & 1 \\ 0 & 0 & \dots & 0 \end{pmatrix}}_{A} x + \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_{b} \cdot \underbrace{\begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{n-1} \end{pmatrix}}_{u^*} x \qquad (2.28)$$

Können die Parameter u gemäß einer Kontrollfunktion gesteuert werden, so erhalten wir ein bilineares Zeilen-Kontrollsystem.

Beispiel 2.33 (Switching)**.** Gegeben seien zwei dynamische Systeme durch lineare Differentialgleichungen n-ter Ordnung

$$v^{(n)} + \sum_{k=0}^{n-1} \alpha_k v^{(k)} = 0 \qquad \text{System 1}$$

$$w^{(n)} + \sum_{k=0}^{n-1} \beta_k w^{(k)} = 0 \qquad \text{System 2}$$

mit konstanten Parametern α_k, β_k $(k = 1, \dots, n-1)$. Diese lassen sich in Matrixschreibweise (2.28) in der Form $\dot{v} = Av + \mathrm{e}_n c_1^* v$ bzw. $\dot{w} = Aw + \mathrm{e}_n c_2^* w$ darstellen[3], wobei die Vektoren c_1 und c_2 die jeweiligen Systemparameter als Komponenten besitzen. Wir setzen $C := A + \mathrm{e}_n c_1^*$ und $D := A + \mathrm{e}_n c_2^*$. Unter „Schalten“ zwischen den Systemen $\dot{v} = Cv$ und $\dot{w} = Dw$ verstehen wir eine Bang-Bang-Steuerung $u : I \to \{-1, +1\}$ innerhalb des Systems

$$\dot{x}(t) = \left(\frac{C+D}{2} + u(t) \frac{C-D}{2} \right) x(t), \qquad t \in I, \;\; -1 \le u(t) \le 1, \;\; x(t) \in \mathbb{R}^n \;,$$

wobei

$$\frac{C-D}{2} = \tfrac{1}{2} \mathrm{e}_n (c_1 - c_2)^* \;.$$

[3] e_n bezeichnet den n-ten Einheitsvektor.

(Man beachte, dass die Werte $u = \pm 1$ genau den beiden Ausgangs-Systemen entsprechen.) Dies ist ein Single-Input System von Rang 1. Zum Beispiel, die Schaltung zwischen zwei LC-Schaltkreisen aus Beispiel 2.15 könnte so modelliert werden.

Beispiel 2.34 (Lineare Systeme)**.** Betrachte das autonome lineare System im $\mathbb{R}^n$

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad u(t) \in \mathbb{R}^m \ .$$

Wir setzen $v(t) := Bu(t)$ und erhalten so das äquivalente System

$$\dot{x}(t) = Ax(t) + v(t), \qquad v(t) \in \mathcal{U}$$

mit Steuerbereich $\mathcal{U} := \mathrm{Im}B = \{Bu \mid u \in \mathbb{R}^m\}$. Dieses bringen wir—wie in (2.6)—auf die homogene Form im erweiterten Zustandsraum $\mathbb{R}^{n+1}$

$$\dot{x}(t) = Ax(t) + \varphi(t)v(t), \qquad \dot{\varphi}(t) = 0, \qquad v(t) \in \mathcal{U} \ .$$

Umformen liefert ein bilineares System von Rang 1 (genauer: Spalten-Kontrollsystem)

$$\begin{pmatrix} \dot{x}(t) \\ \dot{\varphi}(t) \end{pmatrix} = \left(\begin{pmatrix} A & 0 \\ 0^* & 0 \end{pmatrix} + \begin{pmatrix} v(t) \\ 0 \end{pmatrix} \begin{pmatrix} 0^* & 1 \end{pmatrix} \right) \cdot \begin{pmatrix} x(t) \\ \varphi(t) \end{pmatrix}, \quad v(t) \in \mathrm{Im}B \ . \quad (2.29)$$

Auf diese Weise können lineare Kontrollsysteme als bilineare Systeme von Rang 1 interpretiert werden.

Kapitel 3

Die erreichbaren Mengen bilinearer Systeme

3.1 Einführung

Bevor wir uns mit den Eigenschaften der erreichbaren Mengen von verschiedenen bilinearen Systemen auseinandersetzen, werden nochmals einige Begriffe zur Erreichbarkeit ins Gedächtnis gerufen, die leicht modifiziert aus Abschnitt 1.3 übernommen sind.

Definition 3.1. Gegeben sei ein kontinuierliches System Σ in $\mathbb{R}^n$ mit einem nichtleeren Steuerbereich $\mathcal{U} \subseteq \mathbb{R}^m$ und eine Klasse zulässiger Kontrollfunktionen.
Angenommen, es existiert eine (Caratheodory-) Lösung $x : [t_0, T] \to \mathbb{R}^n$ bzgl. einer zulässigen Kontrolle $u : [t_0, T] \to \mathcal{U}$, welche die Anfangsbedingung $x(t_0) = p$ erfüllt. Falls $x(t) = q$ für ein $t \in [t_0, T]$, so sagen wir, dass q *erreichbar von* (p, t_0) *zur Zeit* $t - t_0$ ist.

Die erreichbare Menge von $p \in \mathbb{R}^n$ (zur Zeit $t \geq 0$) ist

$$\mathcal{A}_t(p) = \{q \in \mathbb{R}^n \mid q \text{ erreichbar von } (p, 0) \text{ zur Zeit } t\}$$

bzw.

$$\mathcal{A}(p) = \bigcup_{t \geq 0} \mathcal{A}_t(p) \ .$$

Bemerkung 3.2. In der Definition von $\mathcal{A}_t(p)$ gehen wir davon aus, dass der Anfangszustand p zur Anfangszeit $t_0 = 0$ angenommen wird. Bei autonomen Systemen ist dies ohne Einschränkung möglich. Wir sagen dann „erreichbar

von p“ anstatt „erreichbar von $(p, 0)$“. Doch auch bei nichtautonomen Systemen würden die Resultate aus den nächsten Abschnitten für eine Anfangszeit $t_0 \neq 0$ ihre Gültigkeit behalten.

Die erreichbaren Mengen des zeitumgekehrten Systems

Manchmal ist es hilfreich, die zeitliche Entwicklung eines kontinuierlichen Systems umzukehren. Wir definieren dazu das sogenannte “zeitumgekehrte System“ innerhalb der Klasse der bilinearen Systeme.

Definition 3.3. Es $\Sigma = (\mathbb{R}, \mathcal{X}, \mathcal{U}, \phi)$ das kontinuierliche bilineare System aus Definition 2.12. Das *zeitumgekehrte System* von Σ (zum Zeitpunkt 0) ist gegeben durch

$$\dot{x} = -A(-t)x - \sum_{k=1}^{m} u_k N_k(-t) - B(-t)u \ , \quad t \in \mathbb{R},\ x(t) \in \mathcal{X},\ u(t) \in \mathcal{U} \ .$$

Es wird mit Σ^{-1} bezeichnet.

Bemerkung 3.4. Offenbar ist Σ^{-1} ein wohldefiniertes bilineares System.

Definition 3.5. Wir bezeichnen mit $\mathcal{C}_t(p)$ und $\mathcal{C}(p)$ die erreichbaren Mengen $\mathcal{A}_t(p)$ bzw. $\mathcal{A}(p)$ des zeitumgekehrten Systems Σ^{-1}. (Bei autonomen linearen Systemen ist $\mathcal{C}_t(0)$ auch als „Kontrollierbarkeitsmenge“ bekannt.)

Die Elemente von $\mathcal{C}_t(p)$ werden im folgenden Lemma charakterisiert.

Lemma 3.6. Ein Zustand p ist genau dann erreichbar von $(q, -t)$ zur Zeit $t \geq 0$ bzgl. Σ, wenn q erreichbar ist von $(p, 0)$ zur Zeit t bzgl. des zeitumgekehrten Systems Σ^{-1}.
(Für ein autonomes System Σ bedeutet dies: $p \in \mathcal{A}_t(q) \iff q \in \mathcal{C}_t(p)$.)

Beweis. Wir setzen

$$f(x, u, t) := A(t)x + \sum_{k=1}^{m} u_k N_k(t) + B(t)u \ .$$

Dann ist Σ^{-1} durch die rechte Seite $-f(x, u, -t)$ gegeben.
Es sei $x : [-t, 0] \to \mathbb{R}^n$ eine Lösung von $\dot{x} = f(x, u, t)$ zu einer zulässigen Kontrolle u mit $x(-t) = q$ und $x(0) = p$. Dann ist $\tilde{x}(s) := x(-s)$ eine Lösung von $\dot{x} = -f(x, v, -t)$ auf $[0, t]$ zur zulässigen Kontrolle $v(s) := u(-s)$ mit $\tilde{x}(0) = p$ und $\tilde{x}(t) = q$. Denn es gilt fast überall

$$\frac{\mathrm{d}}{\mathrm{dt}}\tilde{x}(t) = \dot{x}(-t) \cdot (-1) = -f(x(-t), u(-t), -t) = -f(\tilde{x}(t), v(t), -t) \ .$$

Die Rückrichtung ist ebenso gültig. □

Die erreichbaren Mengen homogener Systeme

Bei homogenen Systemen der Form (2.11)

$$\dot{x}(t) = \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) x(t), \quad t \in I, \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathcal{U}$$

können wir uns auf das Studium einer einzigen erreichbaren Menge des assoziierten Matrixsystems (2.12) beschränken.

Definition 3.7. Völlig analog zu Definition 3.1 definieren wir die erreichbaren Mengen von Matrixsystemen in $\mathbb{R}^{n\times n}$. Ist $P \in \mathbb{R}^{n\times n}$, so bezeichnen wir mit $\mathcal{A}_t(P)$ bzw. $\mathcal{A}(P)$ die erreichbare Menge von $(P, 0)$ (zur Zeit $t \geq 0$). Wie in 3.3 können wir ein zeitumgekehrtes (Matrix-) System zu (2.12) bestimmen, dessen erreichbaren Mengen durch $\mathcal{C}_t(P)$ bzw. $\mathcal{C}(P)$ beschrieben werden sollen. Von besonderer Bedeutung ist die erreichbare Menge $\mathcal{A}_t(\mathrm{I})$ bzw. $\mathcal{A}(\mathrm{I})$ von der Einheitsmatrix I. Wir führen daher die Kurzformen $\mathcal{A}_t := \mathcal{A}_t(\mathrm{I})$ und $\mathcal{C}_t := \mathcal{C}_t(\mathrm{I})$ ein.

Bezeichnung 3.8 (Mengenwertige Operationen)**.** Für nichtleere Mengen $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathbb{R}^{n\times m}$, $\mathcal{N} \subseteq \mathbb{R}^{m\times s}$ und $\alpha \in \mathbb{R}$ setzen wir

$$\begin{aligned}
\mathcal{M}_1 + \mathcal{M}_2 &:= \{M_1 + M_2 |\ M_1 \in \mathcal{M}_1, M_2 \in \mathcal{M}_2\}\ , \\
\mathcal{M}_1 - \mathcal{M}_2 &:= \{M_1 - M_2 |\ M_1 \in \mathcal{M}_1, M_2 \in \mathcal{M}_2\}\ , \\
\mathcal{M}_1 \cdot \mathcal{M}_2 &:= \{M_1 \cdot M_2 |\ M_1 \in \mathcal{M}_1, M_2 \in \mathcal{M}_2\}\ , \\
\alpha \cdot \mathcal{N} &:= \{\alpha \cdot N |\ N \in \mathcal{N}\}\ .
\end{aligned}$$

Falls $\mathcal{N} \subseteq \mathrm{GL}(n;\mathbb{R})$, so ist auch

$$\mathcal{N}^{-1} := \{N^{-1} |\ N \in \mathcal{N}\}$$

wohldefiniert.

Folgerung 3.9. Für alle $P \in \mathbb{R}^{n\times n}$, $p \in \mathbb{R}^n$ und $t, s \geq 0$ gilt:

(i) $\mathcal{A}_0 = \{\mathrm{I}\}$.

(ii) $\mathcal{A}_t = \{\Phi(t; 0, u) \mid u \text{ zulässig}\}$, wobei Φ die Fundamentallösung aus Definition 2.19 ist.

(iii) Die Elemente aus $\mathcal{A}_t$ sind invertierbar.

(iv) $\mathcal{A}_t(p) = \mathcal{A}_t \cdot p$ sind die erreichbaren Mengen zu (2.11).

(v) $\mathcal{A}_t(P) = \mathcal{A}_t \cdot P$ sind die erreichbaren Mengen zu (2.12).

Falls das Matrixsystem (2.12) autonom ist (d.h. alle matrixwertigen Funktionen konstant), gilt zusätzlich für $t, s \geq 0$:

(vi) $\mathcal{A}_{t+s} = \mathcal{A}_t \cdot \mathcal{A}_s$ (Additionstheorem).

(vii) $\mathcal{C}_t = \mathcal{A}_t^{-1} \overset{\text{(ii)}}{=} \{\Phi(0; t, u) \mid u \text{ zulässig}\}$. (Insbesondere ist $\mathcal{C}_t(p) = \mathcal{A}_t^{-1} \cdot p$.)

(viii) $\mathcal{A} = \mathcal{A}(\mathrm{I})$ ist eine multiplikative Halbgruppe mit Einselement.

Beweis. (i) und (ii) sind trivial. (iv) und (v) folgen aus Folgerung 2.20. (iii) wird in Folgerung 2.21 gezeigt. (vi) folgt aus (2.17). Die Eigenschaft (vii) rechnen wir nach:

$$P \in \mathcal{C}_t \overset{3.6}{\iff} \mathrm{I} \in \mathcal{A}_t(P) \overset{\text{(v)}}{\iff} \mathrm{I} \in \mathcal{A}_t \cdot P \iff P^{-1} \in \mathcal{A}_t \iff P \in \mathcal{A}_t^{-1}$$

Schließlich implizieren (i) und (vi) zusammen mit der Assoziativität von $(\mathbb{R}^{n\times n}, \cdot\,)$ die Eigenschaft (viii). □

Bemerkung 3.10. Das Lemma zeigt, dass man aus $\mathcal{A}_t$ problemlos per Multiplikation mit dem Anfangswert alle weiteren erreichbaren Mengen von (2.11) und (2.12) ableiten kann.
Die erreichbare Menge von $p \in \mathbb{R}^n$ zur Zeit $t \geq 0$ bezüglich des inhomogenen bilinearen Systems (2.4) ist gemäß Bemerkung 2.22 gleich der erreichbaren Menge $\mathcal{A}_t \cdot \binom{p}{1}$ des homogenisierten Systems (2.6), eingebettet in die Hyperebene $\{x \in \mathbb{R}^{n+1} \mid x_{n+1} = 1\}$ des $\mathbb{R}^{n+1}$.

3.2 Stark invariante Mengen

Bisher hatten wir meist den gesamten $\mathbb{R}^n$ als Zustandsraum gewählt. Leider ist dieser Zustandsraum bei der Modellierung realer Systeme—etwa aus der Biologie—oft ungeeignet. Soll z.B. in einer Zustandsgröße der Bestand einer Spezies aufgenommen werden, so währen negative Bestände biologisch nicht interpretierbar.
Invariante Mengen erlauben es nun den Zustandsraum zu beschränken—es ist leicht zu sehen, dass sämtliche Eindeutigkeits- und Existenzsätze ihre Gültigkeit bewahren—und ermöglichen somit eine bessere Modellierung. Sie sind insbesondere auch ein Maß für die Güte der Modellierung. Wäre z.B. in unserem Bestandsmodell die Menge der nichtnegativen reellen Zahlen keine stark invariante Menge, so ist vermutlich bei der Konstruktion der Zustandsübergangsfunktion ein Fehler aufgetreten.

Besonders hilfreich sind kompakte invariante Mengen bei der Optimierung. Bei Matrixsystemen besitzen invariante Mengen häufig eine Gruppenstruktur, wie wir in diesem Abschnitt sehen werden.

Definition 3.11. Eine Teilmenge $M \subseteq \mathbb{R}^n$ heißt *stark invariant* oder *positiv invariant* (bzgl. eines Systems $\Sigma = (\mathcal{T}, \mathcal{X}, \mathcal{U}, \phi)$ im $\mathbb{R}^n$), falls

$$\mathcal{A}(p) \subseteq M \qquad \forall p \in M \ .$$

M heißt *negativ invariant*, falls

$$\mathcal{C}(p) \subseteq M \qquad \forall p \in M \ .$$

Ist M positiv und negativ invariant, so nennen wir M *beidseitig invariant*.

Bemerkung 3.12. Natürlich ist der Durchschnitt oder die Vereinigung stark invarianter Mengen stark invariant.

Lemma 3.13. Es sei M eine beidseitig invariante Menge bzgl. eines autonomen bilinearen Systems im $\mathbb{R}^n$. Dann ist auch das Komplement $M^c = \mathbb{R}^n \backslash M$ beidseitig invariant. Anders gesagt, jede zulässige Lösung $x : I \to \mathbb{R}^n$ besitzt die Eigenschaft

$$x(t_0) \in M \text{ für ein } t_0 \in I \Longrightarrow x(t) \in M \text{ für alle } t \in I \ .$$

(Bei nichtautonomen Systemen gilt dies nur für $t_0 = 0$.)

Beweis. Es wird die zweite Aussage gezeigt. Dazu sei $x : I \to \mathbb{R}^n$ eine zulässige Lösung mit $p := x(t_0) \in M$. Die Autonomie von Σ erlaubt es die Lösungstrajektorie zeitlich zu „verschieben“. Wir können also ohne Einschränkung $t_0 := 0$ setzen. Aus $p \in M$ und der positiven Invarianz von M folgt $x(t) \in M$ für alle $t \in I$ mit $t \geq 0$. Für irgendein $t \in I$ mit $t < 0$ betrachten wir $q := x(t)$. Nach Konstruktion ist p erreichbar von (q, t) zur Zeit $-t > 0$. Wegen Satz 3.6 ist dann q erreichbar von p zur Zeit $-t$ bzgl. des zeitumgekehrten Systems, d.h. $q \in \mathcal{C}_{-t}(p) \subseteq C(p) \subseteq M$. □

Beispiel 3.14. Zur Modellierung des Wachstums der Weltbevölkerung verwenden wir das skalare bilineare System

$$\dot{x}(t) = u(t)x(t), \qquad u(t) \in \mathcal{U}, \qquad x(t) > 0, \qquad t \geq 0$$

mit Eingabegröße $u(t) := \log(1 + \frac{r(t)}{100})$, wobei $r(t)$ die Wachstumsrate zum Zeitpunkt t (in Prozent) bezeichne. Wir gehen davon aus, dass die Wachstumsrate ungefähr zwischen 0.5 und 2.5 Prozent liegt, und wählen demzufolge

den Steuerbereich $\mathcal{U} := [0.005, 0.025]$. Die zulässigen Kontrollfunktion seien die stetigen Funktionen mit Werten in $\mathcal{U}$. Die zugehörigen Lösungen sind gegeben durch

$$x(t) = x(t_0) \cdot e^{\int_{t_0}^{t} u(s)\mathrm{ds}} \ .$$

Indem wir die Werte von $u(\cdot)$ variieren, erhalten wir für eine Anfangspopulation $p > 0$ die erreichbare Menge

$$\mathcal{A}_t(p) = \left\{ p \cdot e^{\lambda t} \;\middle|\; 0.005 \leq \lambda \leq 0.025 \right\}$$

und folglich

$$\mathcal{A}(p) = [p, \infty) \qquad \forall p \geq 0 \ .$$

Die entsprechende erreichbare Menge des zeitumgekehrten Systems „$\dot{x} = -ux$“ ist

$$\mathcal{C}(p) = (0, p] \qquad \forall p > 0 \ .$$

Die Menge $[p, \infty)$ ist also positiv invariant für alle $p > 0$—aber nicht negativ invariant. $(0, \infty)$ ist sogar beidseitig invariant.

Bilineare Systeme, definiert auf Lie-Gruppen

Wir wollen zunächst den Satz von Liouville B.25 nutzen, um invariante Mengen zum bilinearen Matrixsystem

$$\dot{X}(t) = \underbrace{\left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right)}_{:= U(t)} X(t), \quad X(t) \in \mathbb{R}^{n \times n}, \quad u(t) \in \mathcal{U} \tag{3.1}$$

aus Abschnitt 2.3 herzuleiten.

Folgerung 3.15. Die allgemeine lineare Gruppe $\mathrm{GL}(n, \mathbb{R})$ und deren Untergruppe $\mathrm{GL}^+(n, \mathbb{R}) = \{A \in \mathrm{GL}(n, \mathbb{R}) \mid \det A > 0\}$ bilden zusammen mit ihren Komplementen $\mathbb{R}^{n \times n} \backslash \mathrm{GL}(n, \mathbb{R})$ bzw. $\mathrm{GL}(n, \mathbb{R}) \backslash \mathrm{GL}^+(n, \mathbb{R})$ beidseitig invariante Mengen bezüglich des Systems (3.1).

Beweis. Ist $X(\cdot)$ eine zulässige Lösung von (3.1) auf einem Intervall $I \ni 0$ zu einer Kontrolle $u \in \mathcal{U}_u^I$, so gilt gemäß Eigenschaft (v) aus Folgerung 2.21

$$\det X(t) = \det \Phi(t; 0, u) \cdot \det X(0) = \exp\left(\int_0^t \mathrm{tr} U(s) \mathrm{ds} \right) \cdot \det X(0) \quad \forall t \in I \ ,$$

wobei $\int_0^t = -\int_t^0$ für $0 > t$. Wegen $\exp > 0$ folgt sofort die Behauptung. □

Beispiel 3.16. In Beispiel 3.14 ist $\mathrm{GL}^+(1,\mathbb{R}) = (0,\infty)$ beidseitig invariant.

Eine besonders interessante Klasse autonomer bilinearer Systeme erhält man, indem man alle matrixwertigen Funktionen aus (3.1) als konstant und schiefsymmetrisch (z.B. $A = -A^*$) voraussetzt, d.h.

$$\dot{X}(t) = \left(A + \sum_{k=1}^{m} u_k(t) N_k\right) X(t), \quad X(t) \in \mathbb{R}^{n\times n}, \quad u(t) \in \mathcal{U} \tag{3.2}$$

mit schiefsymmetrischen Matrizen $A, N_1, \ldots, N_m \in \mathbb{R}^{n\times n}$.

Satz 3.17. Die orthogonale Gruppe $\mathrm{O}(n) = \{T \in \mathbb{R}^{n\times n} \mid T^*T = \mathrm{I}\}$ und die spezielle orthogonale Gruppe $\mathrm{SO}(n) = \mathrm{GL}^+(n,\mathbb{R}) \cap \mathrm{O}(n)$ bilden beidseitig invariante Mengen von (3.2).

Beweis. Ist $X(\cdot)$ eine Lösung von (3.2) zu einer zulässigen Kontrolle u auf $I \ni 0$, welche der Bedingung $X(0) \in \mathrm{O}(n)$ genügt, so gilt fast überall

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dt}} X^*X =& \dot{X}^*X + X^*\dot{X} \\
=& X^*\left(A^* + \sum_{k=1}^{m} u_k N_k^*\right) X + X^*\left(A + \sum_{k=1}^{m} u_k N_k\right) X \\
=& X^*\left(-A - \sum_{k=1}^{m} u_k N_k\right) X + X^*\left(A + \sum_{k=1}^{m} u_k N_k\right) X = 0\ .
\end{aligned}$$

Dies impliziert $X(t)^*X(t) = X(0)^*X(0) = \mathrm{I}$ für alle $t \in I$. Die orthogonale Gruppe $\mathrm{O}(n)$ bildet also eine beidseitig invariante Menge. Da nach Lemma 3.15 auch $\mathrm{GL}^+(n,\mathbb{R})$ beidseitig invariant ist, gilt dies sogar für den Durchschnitt $\mathrm{SO}(n)$. □

Diese Klasse bilinearer Systeme ist besonders relevant in der Physik [2], da sich die Zustände des assoziierten Vektorsystems auf der Sphäre $\|x(t)\| = \|x(0)\|$ bewegen. Denn aus $\mathrm{I} \in \mathrm{O}(n)$ folgt

$$\|x(t)\|^2 = (\Phi(t;u)\cdot x(0))^*(\Phi(t;u)\cdot x(0)) = x(0)^* \underbrace{\Phi(t;u)^*\Phi(t;u)}_{\mathrm{I}} x(0) = \|x(0)\|^2.$$

Man benötigt diese Eigenschaft, um beispielsweise einen starren Körper modellieren zu können.

Beispiel 3.18 (Starrer Körper)**.** Die Bewegungsgleichung eines starren Körper im $\mathbb{R}^3$, der um eine Menge fester Rotationsachsen rotiert und den Ursprung als Schwerpunkt besitzt, lautet

$$r(t) = A(t)r(0)\ ,$$

wobei $r(t) \in \mathbb{R}^3$ die Position eines Punktes auf dem Körper ist, und $A(t) \in \mathrm{SO}(3)$ die Orientierung des Körpers (zum Zeitpunkt t). Die Drehmatrix $A(t)$ erfüllt die Differentialgleichung

$$\dot{A}(t) = \begin{pmatrix} 0 & \omega_z(t) & \omega_y(t) \\ -\omega_z(t) & 0 & \omega_x(t) \\ \omega_y(t) & -\omega_x(t) & 0 \end{pmatrix} A(t) \ ,$$

wobei $\omega = (\omega_x, \omega_y, \omega_z)$ die Winkelgeschwindigkeit zum Zeitpunkt t bezeichnet. Indem wir w als Eingabegröße interpretieren, erhalten wir ein bilineares Matrixsystem mit Zustandsraum $\mathcal{X} = \mathrm{SO}(3)$, welches die Orientierung eines starren Körpers modelliert.

Bemerkung 3.19. Ohne auf Details eingehen zu wollen, sei bemerkt, dass $\mathrm{GL}(n, \mathbb{R})$ und $\mathrm{SO}(n)$ zusammen mit der Verknüpfung $[A, B] := AB - BA$ („Lie-Klammer“) sogenannte „Lie-Gruppen“ bilden. Jeder Lie-Gruppe G kann man eine „Lie-Algebra“ $L(G)$ zuordnen, deren Elemente analytische Vektorfelder auf G sind. Die Lie-Algebra von $\mathrm{GL}(n, \mathbb{R})$ ist $\mathbb{R}^{n \times n}$; die Lie-Algebra von $\mathrm{SO}(n)$ ist der Vektorraum $\mathrm{Alt}(n, \mathbb{R})$ aller reellen schiefsymmetrischen Matrizen.
Wie wir für $\mathrm{GL}(n, \mathbb{R})$ und $\mathrm{SO}(n)$ gesehen haben, bildet eine Lie-Gruppe G aus $\mathbb{R}^{n \times n}$ eine beidseitig invariante Menge von unserem homogenen System (3.2), wenn die Matrizen $A, N_1, \ldots, N_p$ aus $L(G)$ stammen. Wir können dann den Zustandsraum auf $\mathcal{X} = G$ beschränken und erhalten so ein System, das auf einer Lie-Gruppe definiert ist. Man spricht auch von einem „System auf G“.
In [13], [2] und [3] werden derartige Kontrollsysteme ausführlich untersucht. Für das Studium dieser Arbeiten sind Kenntnisse auf dem Gebiet der Lie-Theorie dringend erforderlich.

3.3 Die erreichbaren Mengen autonomer Systeme

3.3.1 Systeme mit kompaktem Steuerbereich

In Abschnitt B.4 werden die erreichbaren Mengen von autonomen Kontrollsystemen mit kompaktem Steuerbereich und stetiger rechter Seite untersucht. Es ist daher naheliegend die dortigen Resultate auf das autonome bilineare System

$$\dot{x}(t) = Ax(t) + \sum_{k=1}^{m} u_k(t) N_k x(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathcal{U} \qquad (3.3)$$

mit kompaktem Steuerbereich $\mathcal{U} \subset \mathbb{R}^m$ und Matrizen $A, N_1, \ldots, N_m, B$ von angemessener Form zu übertragen.
Die zulässigen Funktionen seien die messbaren Funktionen mit Werten in $\mathcal{U}$ (kurz: $u \in \mathcal{U}_r^I$). Da die rechte Seite

$$f(x,u) := Ax + \sum_{k=1}^{m} u_k N_k x + Bu$$

stetig auf $\mathbb{R}^n \times \mathbb{R}^m$ ist, dürfen wir jetzt die Theorie aus B.4 auf das System (3.3) anwenden, um die erreichbaren Mengen $\mathcal{A}_t(p)$ von (3.3) zu charakterisieren.

Folgerung 3.20 (Monotonie)**.** In Satz B.35 wird festgestellt, dass die erreichbare Menge eines kritischen Punktes $p \in \mathbb{R}^n$ monoton ist, d.h. aus

$$0 = Ap + \sum_{k=1}^{m} u_k N_k p + Bu$$

für ein $u \in \mathcal{U}$ folgt

$$\mathcal{A}_s(p) \subseteq \mathcal{A}_t(p) \qquad \forall 0 \le s \le t \ .$$

Folgerung 3.21 (Beschränktheit)**.** Aus Folgerung B.38 erhalten wir eine kompakte Obermenge der erreichbaren Menge $\mathcal{A}_t(p)$ von $p \in \mathbb{R}^n$ zur Zeit $t \ge 0$. Es ist

$$\mathcal{A}_t(p) \subseteq \overline{B\big(0, \max\{1, \|p\|\} \cdot \exp(\mu t)\big)} \qquad \forall t \ge 0$$

für

$$\mu := \max_{u \in \mathcal{U}} \left\{ \left\| A + \sum_{k=1}^{m} u_k N_k \right\| + \|Bu\| \right\} \ .$$

Die Menge $\mathcal{A}_t(p)$ ist also beschränkt.

Folgerung 3.22 (Abgeschlossenheit und Stetigkeit)**.** Der Satz von Fillipov B.44 besagt, dass für alle $p \in \mathbb{R}^n$ und $t \ge 0$ die erreichbare Menge $\mathcal{A}_t(p)$ eine kompakte und nichtleere Teilmenge des $\mathbb{R}^n$ ist, falls der Steuerbereich $\mathcal{U}$ konvex ist. Außerdem ist dann die Abbildung $t \mapsto \mathcal{A}_t(p)$ stetig bezüglich der Hausdorffmetrik.

Beweis. Um den Satz von Fillipov anwenden zu können, müssen wir zeigen, dass für alle $x \in \mathbb{R}^n$ die Menge $F(x) := \bigcup_{u\in\mathcal{U}} f(x,u)$ konvex ist.
Da die rechte Seite eines bilinearen Systems definitionsgemäß affin-linear in u ist für ein vorgegebenes $x \in \mathbb{R}^n$, existiert eine Matrix $H(x) \in \mathbb{R}^{n\times m}$ und ein Vektor $h(x) \in \mathbb{R}^n$, so dass

$$f(x,u) = H(x)u + h(x) \ .$$

Es folgt für alle $\lambda \in [0,1]$ und $u, v \in \mathcal{U}$:

$$\begin{aligned}\lambda f(x,u) + (1-\lambda) f(x,v) &= H(x)\cdot(\lambda u + (1-\lambda)v) + h(x) \\ &= f(x, \underbrace{\lambda u + (1-\lambda)v}_{\in\mathcal{U},\text{ da }\mathcal{U}\text{ konvex}}) \in F(x) \ .\end{aligned}$$

Wie gefordert ist $F(x)$ konvex. Der Rest folgt aus Satz B.44. □

Beispiel 3.23. Betrachte das initialisierte Single-Input System

$$\dot{x}(t) = (A + u(t)N)\,x(t), \quad x(t) \in \mathbb{R}^n, \quad -1 \le u(t) \le 1, \quad x(0) = p$$

mit kommutierenden Matrizen A und N. Aus der Lösungsdarstellung (2.23) folgt direkt

$$\mathcal{A}_t(p) = \mathcal{A}_t\cdot p = \left\{ e^{At}\cdot e^{N\int_0^t u(s)\mathrm{d}s}\cdot p \;\middle|\; u:[0,t]\to[-1,1] \text{ messbar} \right\} \ .$$

Indem wir $u(\cdot)$ über die zulässigen Kontrollen variieren, gewinnen wir die erreichbare Menge

$$\mathcal{A}_t(p) = \left\{ e^{At}\cdot e^{Ns}\cdot p \mid -t \le s \le t \right\} \ .$$

Dies ist der Graph einer Kurve im $\mathbb{R}^n$ über einem abgeschlossenen Intervall, woraus die Kompaktheit von $\mathcal{A}_t(p)$ resultiert.

3.3.2 Die affine Hülle erreichbarer Mengen

Wir wenden uns nun dem initialisierten homogenen System

$$\dot{x}(t) = Ax(t) + \sum_{k=1}^{m} u_k(t) N_k x(t), \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathcal{U}, \quad x(0) = p \quad (3.4)$$

zu. Diesmal ist $\mathcal{U}$ eine (nichtleere) Teilmenge des $\mathbb{R}^m$, und $\mathcal{U}_u^I$ ist die Klasse zulässiger Kontrollen. Weiter ist $A, N_1, \ldots, N_m \in \mathbb{R}^{n\times n}$ und $p \in \mathbb{R}^n$. Zur Vereinfachung setzen wir

$$U_u := \sum_{k=1}^{m} u_k N_k$$

für ein $u \in \mathcal{U}$. Die rechte Seite wird dann zu $f(x,u) = (A + U_u)x$.

Das Ziel dieses Abschnittes ist das Auffinden affiner Unterräume, welche die erreichbare Menge $\mathcal{A}_t(p)$ zu (3.4) enthalten.

Definition 3.24. Ein *affiner Unterraum* des $\mathbb{R}^n$ ist eine Teilmenge $M \subseteq \mathbb{R}^n$ von der Form

$$M = p + L \tag{3.5}$$

für ein $p \in M$ und einem linearen Unterraum L des $\mathbb{R}^n$. Während der Punkt $p \in M$ beliebig in (3.5) wählbar ist, ist der lineare Unterraum L eindeutig bestimmt durch

$$L = M - M = \{r - s \mid r, s \in M\} \ .$$

L wird auch die *lineare Komponente* von M genannt.
Hat L die Dimension $n-1$, so nennt man M eine *affine Hyperebene*.

Anhand des nächsten Satzes kann man affine Hyperebenen konstruieren, in denen $\mathcal{A}_t(p)$ enthalten ist.

Satz 3.25 (affine Hyperebenen)**.** Ist $q \in \mathbb{R}^n$ ein Vektor, der die Bedingung

$$q^* A^k U_u = 0 \qquad \forall\ 0 \le k \le n-1,\ u \in \mathcal{U} \tag{3.6}$$

erfüllt, so ist

$$q^* e^{-At} \mathcal{A}_t = q^* \qquad \text{und} \qquad q^* \mathcal{A}_t = q^* e^{At}$$

für alle $t \ge 0$. (Dabei ist $\mathcal{A}_t$ die erreichbare Menge des assoziierten Matrixsystems von (3.4) zum Anfangswert I.)
Multiplikation von rechts mit $p \in \mathbb{R}^n$ zeigt, dass die erreichbare Menge $\mathcal{A}_t(p)$ in den Hyperebenen

$$\{x \in \mathbb{R}^n \mid q^* e^{-At} x = q^* p\} \text{ und } \{x \in \mathbb{R}^n \mid q^* x = q^* e^{At} p\}$$

enthalten ist.

Beweis. 1. Wegen des Satzes von Cayley-Hamilton ist jede Potenz A^j ($j \in \mathbb{N}$) als Linearkombination $A^j = \sum_{k=0}^{n-1} \lambda_k A^k$ im $\mathbb{R}$-Vektorraum $\mathbb{R}^{n\times n}$ darstellbar. Also wird die Bedingung (3.6) zu

$$q^* A^j U_u = \sum_{k=0}^{n-1} \lambda_k \underbrace{q^* A^k U_u}_{=0} = 0 \qquad \forall\ j \in \mathbb{N},\ u \in \mathcal{U} \ ,$$

und folglich

$$q^* e^{-At} U_u = \sum_{k=0}^{\infty} q^* A^k U_u \frac{(-t)^k}{k!} = 0 \qquad \forall\, t \in \mathbb{R},\ u \in \mathcal{U}\ . \tag{3.7}$$

Sei $X(\cdot\,) := \Phi(\cdot\,; u)$ die Fundamentallösung des assoziierten Matrixsystems von (3.4) zu einer zulässigen Kontrolle $u(\cdot\,)$. Dann gilt fast überall

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}} q^* e^{-At} X(t) &= q^* \left(-A e^{-At} X(t) + e^{-At} \left(A + U_u(t)\right) X(t)\right) \\ &= \underbrace{q^* e^{-At} U_u(t)}_{=0} X(t) = 0\ .\end{aligned}$$

Folglich ist $q^* e^{-At} X(t)$ konstant gleich $q^* e^{0\cdot A} X(0) = q^*$, und die erste Aussage ist gezeigt.

2. Wir stellen fest, dass mit q auch $q_1 := e^{A^* t} q$ $(t \geq 0)$ die Voraussetzung (3.7) erfüllt. Denn $q_1^* e^{-At} U_u = q^* e^{At} e^{-At} U_u = q^* U_u \overset{(3.6)}{=} 0$. Wir dürfen also die erste Aussage auf q_1 anwenden und bekommen $q_1^* e^{-At} \mathcal{A}_t = q_1^*$. Substituieren von q_1 liefert die zweite Aussage. □

Es ist bekannt, dass bei linearen Kontrollsystemen der Form

$$\dot{x} = Ax + Bu\ , \qquad x(t) \in \mathbb{R}^n, \qquad \mathcal{U} = \mathbb{R}^m$$

die erreichbare Menge $\mathcal{A}_t(p)$ einen affinen Unterraum bildet, wenn man jede messbare und lokal (essentiell) beschränkte Kontrollfunktion zulässt. Es ist nämlich

$$\mathcal{A}_t(p) = e^{At} p + \underbrace{\mathrm{Im}\big(B|AB|\ldots|A^{n-1}B\big)}_{L}\ .$$

(Insbesondere ist die lineare Komponente von $\mathcal{A}_t(p)$ unabhängig von t.)

Leider gilt dies i.A. nicht für bilineare Systeme; eine Gerade, die durch zwei erreichbare Zustände verläuft, muß nicht in der erreichbaren Menge liegen.

Beispiel 3.26. Wir wählen für das System (3.4) die Daten

$$\mathcal{U} = \mathbb{R}, \quad \mathcal{X} = \mathbb{R}^3, \quad A = 0, \quad N = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad p = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\ .$$

Offenbar sind die Matrizen A, N schiefsymmetrisch, weshalb sich die erreichbaren Punkte von p zur Zeit $t > 0$ auf der Kugel $\overline{B_1(0)}$ mit Radius $\|p\| = 1$ befinden.
Außerdem können wir mittels Satz 3.25 eine affine Hyperebene bestimmen, die $\mathcal{A}_t(p)$ enthält. Denn für $q := \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ folgt aus

$$q^* N = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix} \ ,$$

dass

$$q^* \mathcal{A}_t(p) = q^* p = 1 \ . \tag{3.8}$$

Wegen $AN = NA$ läßt sich die erreichbare Menge $\mathcal{A}_t(p)$ aus der Lösungsdarstellung (2.23) ableiten. Mit MAPLE berechnen wir

$$\mathcal{A}_t(p) = \left\{ e^{Ns} \cdot p \mid s \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 0.5 \\ 0 \\ 0.5 \end{pmatrix} + 0.5 \begin{pmatrix} \cos(\sqrt{2}s) \\ -\sin(\sqrt{2}s) \\ -\cos(\sqrt{2}s) \end{pmatrix} \ \middle| \ s \in \mathbb{R} \right\}$$

für alle $t > 0$. Wie man in Abbildung 3.1 sieht, ist $\mathcal{A}_t(p)$ ein Kreis, der

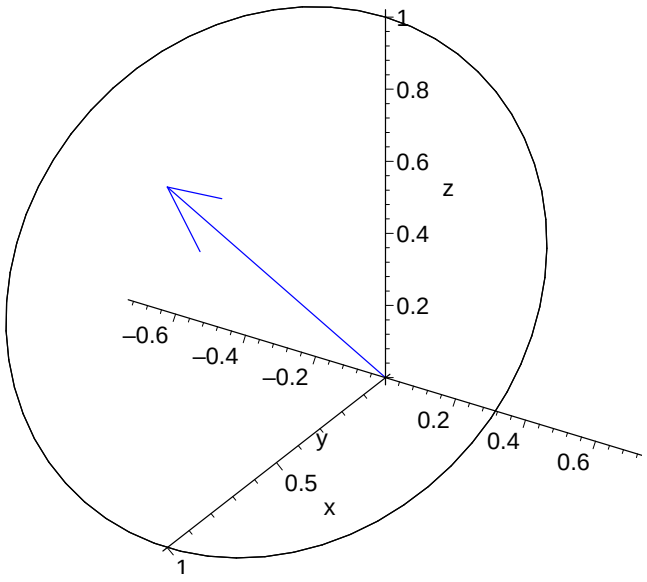

Abbildung 3.1: Die erreichbare Menge aus Bsp. 3.26.

senkrecht auf dem blauen Vektor q steht, und kein affiner Unterraum des $\mathbb{R}^3$.

Falls $\mathcal{A}_t(p)$ kein affiner Unterraum ist, so können wir zumindest auf die affine Hülle von $\mathcal{A}_t(p)$ und deren linearen Komponente Techniken aus der linearen Algebra anwenden.

Definition 3.27. Die *affine Hülle* von einer nichtleeren Teilmenge $N \subseteq \mathbb{R}^n$ ist der kleinste affine Unterraum des $\mathbb{R}^n$, der N enthält. Wir bezeichnen sie mit $\mathrm{aff}(N)$.

Hilfssatz 3.28. Indem man in den Definitionen 3.24 und 3.27 den Vektorraum $\mathbb{R}^n$ durch die $\mathbb{R}$-Algebra $\mathbb{R}^{n\times n}$ ersetzt, kann man ebenso die affine Hülle $\mathrm{aff}(M)$ einer nichtleeren Teilmenge M des $\mathbb{R}^{n\times n}$ definieren. Es gilt dann

$$\mathrm{aff}(M_1 \cdot M_2) \supseteq \mathrm{aff}(M_1) \cdot \mathrm{aff}(M_2) \quad \text{und} \quad \mathrm{aff}(A \cdot M_1) = A \cdot \mathrm{aff}(M_1)$$

für alle $A \in \mathbb{R}^{n\times n}$ und alle nichtleeren Teilmengen M_1, M_2 des $\mathbb{R}^{n\times n}$.

Beweis. Die affine Hülle $\mathrm{aff}(M_1)$ ist die Menge aller affinen Kombinationen von Elementen aus M_1, d.h.

$$\mathrm{aff}(M_1) = \left\{ \sum_{i=1}^{k} \alpha_i m_i \;\middle|\; m_i \in M_1,\ \alpha_i \in \mathbb{R},\ \sum_{i=1}^{k} \alpha_i = 1,\ k \in \mathbb{N} \right\} ,$$

woraus sofort $\mathrm{aff}(A \cdot M_1) = A \cdot \mathrm{aff}(M_1)$ folgt. Außerdem ist es leicht zu sehen, dass das Produkt einer affinen Kombination von Elementen aus $\mathrm{aff}(M_1)$ mit einer affinen Kombination von Elementen aus $\mathrm{aff}(M_2)$ eine affine Kombination der Produkte dieser Elemente ist. Dies beweist die Behauptung. □

Ein Begriff aus der Algebra soll helfen die affine Hülle der erreichbaren Menge $\mathcal{A}_t$ zu charakterisieren.

Bezeichnung 3.29. $S \subseteq \mathbb{R}^{n\times n}$ sei eine Menge reeller $n \times n$-Matrizen. Dann bezeichne $\{S\}_{\mathrm{AA}}$ die kleinste (assoziative) Unteralgebra von $\mathbb{R}^{n\times n}$, welche S enthält. Das bedeutet insbesondere, dass $\{S\}_{\mathrm{AA}}$ ein Untervektorraum von $\mathbb{R}^{n\times n}$ ist (mit Dimension kleiner gleich n^2), der abgeschlossen unter der Matrixmultiplikation ist.

Um den nächsten Satz beweisen zu können, benötigen wir noch zwei elementare Resultate aus der Funktionalanalysis [1, Anhang 1]. Wir verwenden stillschweigend von dort die Definition des Lebesgue-Integrals—in diesem Zusammenhang auch Bochner-Integral genannt—für Funktionen mit Werten in einem Banach-Raum Y. Da in dieser Arbeit für Y nur lineare Unterräume von $\mathbb{R}^n$ und $\mathbb{R}^{n\times n}$ in Frage kommen, bleibt dem Leser eine Einführung in die zugehörige erweiterte Lebesgue-Theorie erspart.

Hilfssatz 3.30. Es sei $\Omega \subseteq \mathbb{R}^n$ eine Lebesgue-messbare Menge, und $f : \Omega \to Y$ sei eine Funktion mit Werten in einem Banachraum Y.
Ist f in einem Punkt $s \in \Omega$ partiell differenzierbar bezüglich der i-ten Koordinatenrichtung, so folgt

$$\partial_i f(s) := \lim_{h \to 0} \frac{f(s + h\mathrm{e}_i) - f(s)}{h} \in Y \ .$$

Ist f Lebesgue-integrierbar, so gilt für das Bochner-Integral

$$\int_\Omega f(s)\mathrm{ds} \in Y \ .$$

Satz 3.31 (Affine Hülle)**.** Es gelte $0 \in \mathcal{U}$ im assoziierten Matrixsystem zu (3.4). (Dies kann man z.B. gewährleisten, indem man ein festes $v \in \mathcal{U}$ wählt und dann $\mathcal{U} := \mathcal{U} - v$, $A := A + \sum N_k v_k$ setzt.)
Dann ist für alle $t > 0$ die affine Hülle von $\mathcal{A}_t$ gegeben durch $e^{At}(\mathrm{I} + \mathcal{L})$, wobei

$$\mathcal{L} = \left\{ \sum_{k=1}^{m} u_k e^{-As} N_k e^{As} \;\middle|\; u \in \mathcal{U}, s \in \mathbb{R} \right\}_{\mathrm{AA}} \ .$$

Entsprechend ist $e^{At}(\mathrm{I} + \mathcal{L})p$ die affine Hülle von $\mathcal{A}_t(p) = \mathcal{A}_t \cdot p$.

Insbesondere ist $\mathcal{L}$ eine Unteralgebra von $\mathbb{R}^{n\times n}$, welche die Menge

$$\left\{ \sum_{k=1}^{m} u_k N_k \;\middle|\; u \in \mathcal{U} \right\} \tag{3.9}$$

enthält, unabhängig von t ist, und invariant unter der Abbildung $\mathrm{ad}_A : X \mapsto [A, X] = AX - XA$ ist.

Beweis. 1. Anstelle von $\mathcal{A}_t$ betrachten wir die sogenannten „reduzierten“ erreichbaren Mengen $\mathcal{B}_t := e^{-At} \cdot \mathcal{A}_t$ für $t \geq 0$.
Wegen $0 \in \mathcal{U}$ ist $\Phi(t; 0) = e^{At} \in \mathcal{A}_t$ und $\mathrm{I} \in \mathcal{B}_t$. Weiter folgt aus dem Additionstheorem in Folgerung 3.9, dass

$$\mathcal{B}_{t+s} := e^{-A(t+s)} \underbrace{\mathcal{A}_{t+s}}_{=\mathcal{A}_t \cdot \mathcal{A}_s} = e^{-As} e^{-At} \mathcal{A}_t e^{As} e^{-As} \mathcal{A}_s = e^{-As} \underbrace{\mathcal{B}_t}_{\ni \mathrm{I}} e^{As} \mathcal{B}_s \tag{3.10}$$

und somit

$$\mathcal{B}_{t+s} \supseteq \mathcal{B}_s \qquad \forall t, s \geq 0 \ .$$

Die reduzierten erreichbaren Mengen sind also monoton.

2. Mit $\mathcal{M}_t$ bezeichnen wir die affine Hülle von $\mathcal{B}_t$. Sie ist von der Form $\mathrm{I} + \mathcal{L}$ mit einer linearen Komponente $\mathcal{L}$, die von t abhängen könnte. Hilfssatz 3.28 und (3.10) implizieren

$$\mathcal{M}_{t+s} \supseteq e^{-As}\mathcal{M}_t e^{As}\mathcal{M}_s \qquad \forall t, s \geq 0 \; . \tag{3.11}$$

Da

$$\mathrm{aff}(\mathcal{A}_t) = \mathrm{aff}(e^{At}\mathcal{B}_t) = e^{At} \cdot \mathrm{aff}(\mathcal{B}_t) = e^{At} \cdot \mathcal{M}_t = e^{At}(\mathrm{I} + \mathcal{L}) \; ,$$

hat die affine Hülle von $\mathcal{A}_t$ die gewünschte Form.

3. Wir wollen nun beweisen, dass $\mathcal{L}$ eine Algebra bildet. Dazu zeigen wir zunächst, dass $\mathcal{L}$ unabhängig von t ist.
Aus 1. wissen wir, dass $\mathcal{M}_t \supseteq \mathcal{M}_s$ für alle $t \geq s \geq 0$. Der affine Raum $\mathcal{M}_t$ kann nur dann „größer“ werden, wenn sich die Dimension der linearen Komponente erhöht. Diese Dimension ist kleiner gleich n^2. Wir folgern, dass $\mathcal{M}_t$ bis auf endlich viele Ausnahmen konstant in t sein muss.
Es gibt daher ein $\delta > 0$ und einen affinen Raum $\mathcal{M}$, so dass

$$\mathcal{M}_t = \mathcal{M} \qquad \forall 0 < t < \delta \; . \tag{3.12}$$

Aus $\mathrm{I} \in \mathcal{B}_s \subseteq \mathcal{M}_s$ und (3.11) folgt dann

$$e^{-As}\mathcal{M}e^{As} \subseteq \mathcal{M} \qquad \forall 0 < s < \delta \; .$$

(Wähle t genügend klein in (3.11), so dass $t + s < \delta$.) Man sieht leicht, dass diese Inklusion auch für jedes $s \geq \delta$ gilt. Denn aus $s = k\delta + r$ für ein $k \in \mathbb{N}$ und $r < \delta$ (Division mit Rest) folgt rekursiv

$$e^{-As}\mathcal{M}e^{As} = \underbrace{e^{-A\frac{\delta}{2}} \ldots e^{-A\frac{\delta}{2}} \underbrace{e^{-A\frac{\delta}{2}} \underbrace{\underbrace{e^{-Ar}\mathcal{M}e^{Ar}}_{\subseteq\mathcal{M}} e^{A\frac{\delta}{2}}}_{\subseteq\mathcal{M}} e^{A\frac{\delta}{2}} \ldots e^{A\frac{\delta}{2}}}_{\subseteq\mathcal{M}}}_{\subseteq\mathcal{M}} \subseteq \mathcal{M} \; .$$

Andererseits ist die Dimension von $\mathcal{M}$ (d.h. die Dimension der linearen Komponente als Unterraum des $\mathbb{R}^{n\times n}$) gleich der Dimension von $e^{-As}\mathcal{M}e^{As}$, da e^{As} und e^{-As} vollen Rang haben. Dies impliziert letztendlich

$$e^{-As}\mathcal{M}e^{As} = \mathcal{M} \qquad \forall s \geq 0$$

und nach Umformung

$$e^{-As}\mathcal{M}e^{As} = \mathcal{M} \qquad \forall s \in \mathbb{R} \; . \tag{3.13}$$

Insbesondere erhalten wir

$$\mathcal{M} \overset{(3.11)}{\supseteq} e^{-As}\mathcal{M}e^{As}\mathcal{M} \overset{(3.13)}{=} \mathcal{M}\cdot\mathcal{M}$$

und daher (wegen $\mathrm{I} \in \mathcal{M}$ gilt auch „$\subseteq$“)

$$\mathcal{M} = e^{-As}\mathcal{M}e^{As}\mathcal{M} = \mathcal{M}\cdot\mathcal{M} \qquad \forall s \in \mathbb{R} \; . \tag{3.14}$$

Die letzte Gleichung zeigt, dass für $0 < t < \delta$ das mengenwertige Produkt $e^{-At}\mathcal{M}_t e^{At}\mathcal{M}_t$ den affinen Raum $\mathcal{M}$ bildet, der somit die Menge

$$e^{-At}\mathcal{B}_t e^{At}\mathcal{B}_t \overset{(3.10)}{=} \mathcal{B}_{2t}$$

enthalten muß. Es folgt aus der Minimalität der affinen Hülle $\mathcal{M}_{2t}$, dass

$$\mathcal{M} \supseteq \mathcal{M}_{2t}$$

und somit (Monotonie)

$$\mathcal{M} = \mathcal{M}_{2t} \qquad \forall 0 < t < \delta \; .$$

Durch wiederholtes Anwenden dieser Argumentation können wir jetzt die δ-Restriktion in (3.12) fallenlassen. Für $0 < t < \delta$ ergibt sich Schritt für Schritt

$$\begin{aligned}
&\mathcal{M} = e^{-At}\mathcal{M}e^{At}\mathcal{M} = e^{-At}\mathcal{M}_{2t}e^{At}\mathcal{M}_t = \mathcal{M}_{3t} \; , \\
&\mathcal{M} = e^{-At}\mathcal{M}e^{At}\mathcal{M} = e^{-At}\mathcal{M}_{3t}e^{At}\mathcal{M}_t = \mathcal{M}_{4t} \; , \\
&\vdots
\end{aligned}$$

Auf diese Weise wird (3.12) zu

$$\mathcal{M}_t = \mathcal{M} \qquad \forall t > 0 \; .$$

Zusammen mit $\mathcal{M}_t$ ist die zugehörige lineare Komponente $\mathcal{L}$ unabhängig von t. Wir prüfen nun, ob $\mathcal{L}$ abgeschlossen unter der Matrixmultiplation ist:

$$\mathcal{L}\cdot\mathcal{L} = (\mathcal{M} - \mathrm{I})\cdot(\mathcal{M} - \mathrm{I}) = \mathcal{M}\cdot\mathcal{M} - 2\mathcal{M} + \mathrm{I} \overset{(3.14)}{=} \mathrm{I} - \mathcal{M} = -\mathcal{L} = \mathcal{L}$$

$\mathcal{L}$ ist also eine Algebra.

4. Wir zeigen, dass $\mathcal{L}$ invariant unter der Abbildung $\mathrm{ad}_A : X \mapsto [A, X] = AX - XA$ ist.
Wegen (3.13) gilt für alle $X \in \mathcal{L}$

$$e^{-As} X e^{As} \in \mathcal{L} \qquad \forall s \in \mathbb{R} \,. \tag{3.15}$$

Die Funktion $s \mapsto e^{-As} X e^{As}$ ist differenzierbar in s mit Werten im Banachraum $\mathcal{L}$. Wie es im Hilfsatz 3.30 bemerkt wird, ist dann die Ableitung in einem Punkt s ein Element von $\mathcal{L}$. Speziell

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{ds}} e^{-As} X e^{As}\Big|_{s=0} = \left(-Ae^{-As} X e^{As} + e^{-As} X A e^{As}\right)\Big|_{s=0} &= -AX + XA \\ &= -[A, X] \in \mathcal{L} \,. \end{aligned}$$

Also $[A, X] \in \mathcal{L}$ für alle $X \in \mathcal{L}$.

5. Um nachzuweisen, dass $\mathcal{L}$ die Menge (3.9) enthält, setzen wir für ein beliebiges $u \in \mathcal{U}$

$$U_u := \sum_{k=1}^{m} u_k N_k$$

und zeigen $U_u \in \mathcal{L}$. Wählt man $u \in \mathcal{U}$ als konstante Konstrolle über $[0, t]$, so folgt $\Phi(t; u) = e^{(A+U_u)t} \in \mathcal{A}_t$. Wir sehen

$$e^{-At} e^{(A+U_u)t} - \mathrm{I} \;\in\; \mathcal{B}_t - \mathrm{I} \;\subseteq\; \mathcal{M} - \mathrm{I} \;=\; \mathcal{L} \,.$$

Analog zu 4. ergibt Ableiten in $t = 0$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{dt}} \left(e^{-At} e^{(A+U_u)t} - \mathrm{I}\right)\Big|_{t=0} &= \left(-Ae^{-At} e^{(A+U_u)t} + e^{-At}(A + U_u) e^{(A+U_u)t}\right)\Big|_{t=0} \\ &= (-A + (A + U_u)) = U_u \in \mathcal{L} \,. \end{aligned}$$

Der zweite Teil des Satzes ist somit bewiesen.

6. Wir wissen, dass $\mathrm{aff}(\mathcal{A}_t)$ von der Form $e^{At}(\mathrm{I} + \mathcal{L})$ ist, wobei $\mathcal{L}$ eine Algebra ist, für die wegen (3.15) und 5.

$$e^{-As} \left(\sum_{k=1}^{m} u_k N_k\right) e^{As} = \sum_{k=1}^{m} u_k e^{-As} N_k e^{As} \in \mathcal{L} \qquad \forall s \in \mathbb{R}, \forall u \in \mathcal{U} \tag{3.16}$$

gilt. Das bedeutet

$$\mathcal{L}^\star := \left\{ \sum_{k=1}^{m} u_k e^{-As} N_k e^{As} \;\middle|\; u \in \mathcal{U}, s \in \mathbb{R} \right\}_{\mathrm{AA}} \subseteq \mathcal{L} \,.$$

7. Es fehlt noch die Inklusion $\mathcal{L}^\star \supseteq \mathcal{L}$. Wegen 2. ist $\mathcal{L} = \mathcal{M} - \mathrm{I}$, wobei $\mathcal{M} = \mathrm{aff}(\mathcal{B}_t)$ und $\mathcal{B}_t = e^{-At}\mathcal{A}_t$. Es genügt daher zu zeigen, dass $\mathcal{M} \subseteq \mathrm{I} + \mathcal{L}^\star$ gilt. Aufgrund der Minimalität der affinen Hülle $\mathcal{M}$ wird dies bereits durch

$$\mathcal{B}_t \subseteq \mathrm{I} + \mathcal{L}^\star \tag{3.17}$$

impliziert für ein $t > 0$. Wir werden jetzt die Inklusion (3.17) beweisen und somit den Beweis beenden.
Ein beliebiges Element $Y(t)$ aus $\mathcal{B}_t$ ist von der Form $Y(t) = e^{-At}\Phi(t;u)$, wobei $\Phi(\cdot\,;u)$ eine Fundamentallösung von (3.4) zu einer zulässigen Kontrolle u ist. Wie in (2.22) vorgerechnet, ist $t \mapsto Y(t)$ selbst eine Fundamentallösung (bzgl. u) des bilinearen Systems

$$\dot{Y}(t) = \underbrace{\left(\sum_{k=1}^{m} u_k(t) e^{-At} N_k e^{At}\right)}_{:=U(t)} Y(t) = U(t)Y(t)\ , \qquad Y(0) = \mathrm{I}\ .$$

Nach Lemma 2.21 (vii) ist folglich $Y(t)$ der Grenzwert der Picard-Iterierten $Y_0(t) := \mathrm{I}$ und $Y_{l+1}(t) := \mathrm{I} + \int_0^t U(s)Y_l(s)ds$:

$$Y(t) = \lim_{l\to\infty} Y_l(t)$$

Per Induktion wird nachgewiesen, dass $Y_l(s) \in \mathrm{I} + \mathcal{L}^\star$ für alle $l \in \mathbb{N}_0$ und $s \in [0,t]$ gilt.
Für $l = 0$ ist $Y_0 \equiv \mathrm{I} \in \mathrm{I} + \mathcal{L}^\star$ (Induktionsstart).
Für ein $k \in \mathbb{N}$ sei $Y_k(s) \in \mathrm{I} + \mathcal{L}^\star$ für alle $s \in [0,t]$. Dann ist

$$U(s)Y_k(s) \in \mathcal{L}^\star \cdot (\mathrm{I} + \mathcal{L}^\star) = \mathcal{L}^\star + \mathcal{L}^\star \cdot \mathcal{L}^\star \subseteq \mathcal{L}^\star$$

und folglich (Hilfssatz 3.30)

$$Y_{k+1}(t) := \mathrm{I} + \int_0^t U(s)Y_k(s)\mathrm{ds} \in \mathrm{I} + \mathcal{L}^\star \qquad \forall k \in \mathbb{N}\ .$$

Natürlich sind $\mathcal{L}^\star$ und somit $\mathrm{I} + \mathcal{L}^\star$ abgeschlossen, da $\mathcal{L}^\star$ ein endlich-dimensionaler linearer Unterraum ist. Der Grenzwert $Y(t)$ liegt also ebenfalls in $\mathrm{I} + \mathcal{L}^\star$, was zu zeigen war. □

Folgerung 3.32. Der Steuerbereich $\mathcal{U}$ des Systems (3.4) enthalte den Einheitswürfel $E_m := \{u \in \mathbb{R}^m \mid -1 \le u_k \le 1,\ k = 1, \dots m\}$. Dann ist der lineare Raum $\mathcal{L}$ aus Satz 3.31 gleich

$$\{\mathrm{ad}_A^k N_j \mid 0 \le k \le n^2 - 1,\ 1 \le j \le m\}_{\mathrm{AA}}\ .$$

Zur Erinnerung:

$$\mathrm{ad}_A^0 N = N,\ \mathrm{ad}_A^1 N = [A, N] \quad \text{und} \quad \mathrm{ad}_A^k N = [A, \mathrm{ad}_A^{k-1} N]\ , k \geq 1$$

Insbesondere ist hier $\mathcal{L}$ die kleinste Unteralgebra von $\mathbb{R}^{n\times n}$, welche die Matrizen $N_1, \ldots, N_m$ enthält und invariant unter der Abbildung $\mathrm{ad}_A : X \mapsto [A, X]$ ist.

Beweis. Es ist

$$\mathcal{L} = \left\{ \sum_{k=1}^{m} u_k e^{-As} N_k e^{As} \;\middle|\; u \in \mathcal{U}, s \in \mathbb{R} \right\}_{\mathrm{AA}} . \tag{3.18}$$

Wegen $E_m \subseteq \mathcal{U}$ ist der j-te kanonische Einheitsvektor des $\mathbb{R}^m$ eine Eingangsgröße aus $\mathcal{U}$. Es folgt direkt $e^{-As} N_k e^{As} \in \mathcal{L}$ für alle $s \in \mathbb{R}$ und $1 \leq k \leq m$. Da diese Matrizen ganz $\mathcal{L}$ erzeugen, erhalten wir

$$\mathcal{L} = \left\{ e^{-As} N_j e^{As} \;\middle|\; s \in \mathbb{R},\ 1 \leq j \leq m \right\}_{\mathrm{AA}} .$$

Die Baker-Hausdorff-Formel (2.20) besagt

$$e^{-As} N_j e^{As} = \sum_{k=0}^{\infty} \frac{s^k}{k!} \mathrm{ad}_A^k N_j \qquad \forall s \in \mathbb{R},\ j = 1, \ldots, m\ . \tag{3.19}$$

Wie im Beweis von Satz 2.29 festgestellt, ist jede Potenz $\mathrm{ad}_A^k N_j$ mit $k \geq n^2$ im $\mathbb{R}$-Vektorraum $\{\mathrm{ad}_A^k N_j \mid 0 \leq k \leq n^2 - 1,\ 1 \leq j \leq m\}_{\mathrm{AA}}$ enthalten. Wir können daher die Reihe aus (3.19) als eine konvergente Folge von Partialsummen in diesem Vektorraum ansehen (bzgl. der induzierten Operatornorm). Da der Vektorraum endlich-dimensional und somit abgeschlossen ist, gilt für den Grenzwert

$$e^{-As} N_j e^{As} \in \{\mathrm{ad}_A^k N_j \mid 0 \leq k \leq n^2 - 1,\ 1 \leq j \leq m\}_{\mathrm{AA}}$$

und folglich

$$\mathcal{L} \subseteq \{\mathrm{ad}_A^k N_j \mid 0 \leq k \leq n^2 - 1,\ 1 \leq j \leq m\}_{\mathrm{AA}}\ .$$

Andererseits ist $\mathcal{L}$ invariant unter der Abbildung $\mathrm{ad}_A : X \mapsto [A, X]$, und enthält die Matrizen $N_1, \ldots, N_m$. Es folgt

$$\mathrm{ad}_A^0 N_j = N_j \in \mathcal{L}\ ,\ \mathrm{ad}_A^1 N_j \in \mathcal{L}\ ,\ \mathrm{ad}_A^2 N_j = [A, \mathrm{ad}_A^1 N_j] \in \mathcal{L}\ , \quad \text{u.s.w.}$$

Dies ergibt schließlich die Behauptung. □

Beispiel 3.33. In Beispiel 3.26 ist die affine Hülle der erreichbaren Menge von $p = \left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)$ zur Zeit $t > 0$ gleich $p + \mathcal{L}\cdot p$, wobei $\mathcal{L}$ die kleinste Unteralgebra ist, erzeugt von der Matrix N (beachte $A = 0$).
Wir rechnen nach:

$$N = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad N^2 = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \quad N^3 = -2N \ .$$

Folglich ist $\mathcal{L} = \mathbb{R}\cdot N + \mathbb{R}\cdot N^2$. Wir erhalten daher die affine Hülle

$$\mathrm{aff}(\mathcal{A}_t(p)) = \left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right) + \mathbb{R}\cdot \left(\begin{smallmatrix}0\\-1\\0\end{smallmatrix}\right) + \mathbb{R}\cdot \left(\begin{smallmatrix}-1\\0\\1\end{smallmatrix}\right) \ .$$

Wie die Abbildung 3.1 vermuten lässt, ist dies genau die affine Hyperebene aus (3.8).

Beispiel 3.34 (Sussmann)**.** Betrachte das bilineare Single-Input System in $\mathrm{GL}(4, \mathbb{R})$

$$\dot{X}(t) = \left(\begin{smallmatrix}0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{smallmatrix}\right) X(t) + u(t) \left(\begin{smallmatrix}0&1&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{smallmatrix}\right) X(t),\ u(t) \in [-1, 1],\ X(0) = \mathrm{I} \ .$$

Wir vereinfachen die Notation für dünn besetzte Matrizen, indem wir durch E_{ij} diejenige 4×4 Matrix bezeichen, bei der eine 1 an der Stelle (i, j) und sonst überall Nullen stehen. Dann ist unser System durch die Matrizen $A = E_{34}$ und $N = E_{12} + E_{23}$ gegeben.
Mit Folgerung 3.32 wollen wir die affine Hülle von $\mathcal{A}_t$ berechnen.
Die Erzeuger der Algebra $\mathcal{L}$ sind

$$\mathrm{ad}_A^0 N = N = E_{12} + E_{23},\ \mathrm{ad}_A^1 N = [A, N] = -E_{24},\ \mathrm{ad}_A^2 N = [A, \mathrm{ad}_A^1 N] = 0 \ldots$$

Durch Matrixmultiplikation erhalten wir aus diesen weitere Matrizen ungleich Null:

$$N^2 = E_{13}, \qquad \mathrm{ad}_A^1 N \cdot N = -E_{14}$$

Es folgt, dass $\mathcal{L}$ als linearer Teilraum des $\mathbb{R}^{n\times n}$ von den Matrizen $E_{12} + E_{23}$, E_{24}, E_{13} und E_{14} aufgespannt wird.
Die lineare Komponente von $\mathrm{aff}(\mathcal{A}_t)$ ist

$$e^{At}\mathcal{L} = (\mathrm{I} + tE_{34})\mathcal{L} = \mathcal{L} \quad \text{(unabhängig von } t) \ .$$

Dies ergibt die affine Hülle

$$\mathrm{aff}(\mathcal{A}_t) = e^{At} + \mathcal{L} = \mathrm{I} + tE_{34} + \mathbb{R}(E_{12} + E_{23}) + \mathbb{R}\cdot E_{13} + \mathbb{R}\cdot E_{14} + \mathbb{R}\cdot E_{24} \ .$$

Insbesondere ist die affine Hülle der erreichbaren Menge $\mathcal{A}_t(p) = \mathcal{A}_t \cdot p$ des assoziierten Vektorsystems höchstens von Dimension 2:

$$\operatorname{aff}(\mathcal{A}_t(p)) = \operatorname{aff}(\mathcal{A}_t)\cdot p \subseteq \begin{pmatrix} p_1 \\ p_2 \\ p_3+tp_4 \\ p_4 \end{pmatrix} + \mathbb{R}\cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mathbb{R}\cdot \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

3.4 Topologische Eigenschaften

3.4.1 Kompaktheit und Zusammenhang

In diesem Abschnitt werden elementare topologische Eigenschaften, wie Zusammenhang oder Kompaktheit, für bilineare Systeme hergeleitet.

In Folgerung 3.22 haben wir mit Hilfe des Satzes von Fillipov B.44 gezeigt, dass die erreichbaren Mengen autonomer bilinearer Systeme mit kompaktem konvexen Steuerbereich kompakt sind. Diese Eigenschaft ist sehr wichtig in der Theorie optimaler Steuerungen, da sie die Existenz zeitoptimaler Steuerungen garantiert (Folgerung B.45).
Leider ist der Satz von Fillipov in seiner klassischen Fassung [11, p.107] nur für Kontrollsysteme der Form „$\dot{x} = f(x,u,t)$“ mit stetiger rechter Seite $f(x,u,t)$ anwendbar. Deshalb verfolgen wir eine andere Strategie, um die Kompaktheit erreichbarer Mengen auch für nichtautonome bilineare Systeme nachzuweisen, deren rechte Seiten nicht notwendigerweise stetig in t sein müssen.
Die Strategie besteht darin, zu zeigen, dass

1. die Menge aller zulässigen Kontrollen $\mathcal{U}_r^{[0,T]}$ schwach kompakt ist, falls der Steuerbereich $\mathcal{U}$ kompakt und konvex ist;
2. es eine stetige surjektive Abbildung von $\mathcal{U}_r^{[0,T]}$ nach $\mathcal{A}_T(p)$ gibt, wenn $\mathcal{U}_r^{[0,T]}$ mit der schwachen Topologie versehen ist.

Dann wäre die erreichbare Menge $\mathcal{A}_T(p)$—als Bild einer kompakten Menge unter einer stetigen Abbildung—selbst kompakt.
Wie üblich beschränken wir uns zunächst auf das Studium homogener Matrixsysteme und übertragen später die Resultate auf inhomogene Systeme in $\mathbb{R}^n$.

Betrachte das initialisierte homogene bilineare System in $\mathrm{GL}(n,\mathbb{R})$

$$\dot{X}(t) = \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) X(t),\ u(t) \in \mathcal{U},\ X(0) = \mathrm{I},\ t \in [0,T] \tag{BS1}$$

mit kompaktem Steuerbereich $\mathcal{U} \subseteq \mathbb{R}^m$ und messbaren, lokal (essentiell) beschränkten Funktionen $A(\cdot), N_1(\cdot), \ldots, N_m(\cdot)$ mit Werten in $\mathbb{R}^{n\times n}$.
Als Klasse zulässiger Kontrollfunktionen kommt anfangs nur der Funktionenraum $\mathcal{U}_r^{[0,T]}$ in Frage, der alle messbaren Funktionen der Form $u : [0,T] \to \mathcal{U}$ enthält. Offenbar ist dies eine Teilmenge des Banachraums $L^\infty([0,T],\mathbb{R}^m)$, wenn man Funktionen, die fast überall übereinstimmen, zu Äquivalenzklassen zusammenfasst.
Wir sehen leicht, dass jedes $f \in L^\infty([0,T],\mathbb{R}^m)$ bezüglich der kanonischen Norm von $L_2([0,t],\mathbb{R}^m)$ beschränkt ist. Denn

$$\|f\|_{L_2}^2 = \int_0^T \|f(s)\|^2 \mathrm{ds} \leq T\|f\|_\infty^2 \ .$$

Unser Funktionenraum $\mathcal{U}_r^{[0,T]}$ ist daher ein metrischer Teilraum des Hilbertraums $L_2([0,T],\mathbb{R}^m)$, für welchen auch die kürzere Notation $L_2[0,T]$ verwendet wird.
Wir können $L_2[0,T]$ und dessen Teilräume sowohl mit der gewöhnlichen Normtopologie als auch mit der schwachen Topologie versehen. Eine in der schwachen Topologie konvergente Folge wird als „schwach konvergent“ bezeichnet. In Anhang A.3 werden wichtige Aussagen zu schwacher Konvergenz und schwacher Topologie aufgelistet. Da $L_2[0,T]$ ein separabler Hilbertraum ist, sind alle Sätze und Lemmas anwendbar.
Insbesondere wird dort festgestellt, dass $L_2[0,T]$ zusammen mit der schwachen Topologie einen Hausdorff-Raum bildet, und dass abgeschlossene, beschänkte, konvexe Teilmengen von $L_2[0,T]$ schwach kompakt sind (Folgerung A.28).
Das heißt, falls $\mathcal{U}$ konvex ist, so würde die Abgeschlossenheit und Beschränktheit von $\mathcal{U}_r^{[0,T]}$ dessen Kompaktheit bezüglich der schwachen Topologie implizieren. (Man beachte, dass in ∞-dimensionalen Funktionenräumen der Satz von Heine-Borel nicht anwendbar ist.)

Definition 3.35. Wir sagen, eine beschränkte Folge (u_k) konvergiert in $L_2[0,T]$ schwach gegen u, falls die Bedingung

$$\int_0^T \big\langle u_k(s), \phi(s)\big\rangle_{\mathbb{R}^m} \mathrm{ds} \to \int_0^T \big\langle u(s), \phi(s)\big\rangle_{\mathbb{R}^m} \mathrm{ds} \qquad \text{für } k \to \infty \tag{3.20}$$

für alle quadratintegrablen Funktionen $\phi \in L_2[0,T]$ erfüllt ist.

Da die Treppenfunktionen dicht in $L_2[0,T]$ liegen, ist die Bedingung (3.20) bereits dann hinreichend für schwache Konvergenz, wenn sie für alle Treppenfunktionen $\phi : [0,T] \to \mathbb{R}^m$ erfüllt ist.

Erinnerung 3.36. $\Phi(\cdot\,;u)$ bezeichnet die (Fundamental-) Lösung von (BS1) zur Kontrolle $u \in \mathcal{U}_r^{[0,T]}$. Die erreichbare Menge zum Zeitpunkt $T \geq 0$ ist gegeben durch

$$\mathcal{A}_T = \left\{ \Phi(T;u) \;\middle|\; u \in \mathcal{U}_r^{[0,T]} \right\} \ .$$

Bemerkung 3.37. Wir können (BS1) als System mit Ausgang (Definition 1.5) interpretieren, indem wir als Messbereich den Zustandsraum $\mathrm{GL}(n,\mathbb{R})$ zusammen mit der trivialen Ausgangsfunktion „$y(t) = x(t)$“ wählen. Halten wir die Anfangszeit 0 und die Endzeit T fest, so ist die Ein-Ausgangsfunktion aus Definition 1.19 gegeben durch

$$\begin{aligned} \Psi: \quad \mathcal{U}_r^{[0,T]} &\longrightarrow C^0([0,T],\mathbb{R}^{n\times n}) \\ \Psi(u)(t) &:= \ \Phi(t;u) \ . \end{aligned}$$

Die Responsefunktion λ aus Gleichung (1.3) ist die punktweise Auswertung der Ein-Ausgangsfunktion, d.h.

$$\begin{aligned} \lambda: \quad [0,T] \times \mathcal{U}_r^{[0,T]} &\longrightarrow \mathbb{R}^{n\times n} \\ \lambda(t;u) &:= \ \Phi(t;u) \ . \end{aligned}$$

Satz 3.38 (Stetigkeit der Ein-Ausgangsfunktion)**.** Die Abbildung

$$u(\cdot\,) \mapsto \ \Phi(\cdot\,;u)$$

ist eine stetige Abbildung vom Hausdorffraum $\mathcal{U}_r^{[0,T]}$, versehen mit der schwachen Topologie, in den Banachraum $C^0([0,T],\mathbb{R}^{n\times n})$, versehen mit der Normtopologie der Supremumsnorm.
Dies bedeutet, dass wenn u_k schwach gegen u konvergiert, so konvergiert $\Phi(t;u_k)$ gleichmäßig gegen $\Phi(t;u)$ auf $[0,T]$ für $k \to \infty$.

Beweis. Für ein vorgegebenes $u \in \mathcal{U}_r^{[0,T]}$ gilt

$$\Phi(t;u) = \mathrm{I} + \int_0^t \left(A(s) + \sum_{i=1}^m u_i(s) N_i(s) \right) \Phi(s;u)\mathrm{ds} \ .$$

Da die Matrixfunktionen (bzgl. der Operatornorm) essentiell beschränkt sind und $\mathcal{U}$ kompakt ist, gibt es ein $C > 0$, so dass

$$\Bigg\| \underbrace{A(t) + \sum_{i=1}^m u_i(t) N_i(t)}_{:=U(t)} \Bigg\| \leq C \qquad \forall' t \in [0,T], \forall u \in \mathcal{U}_r^{[0,T]} \ .$$

Mit Hilfe der absolut konvergenten Neumannreihe aus Folgerung 2.21 gewinnen wir die Abschätzung

$$\begin{aligned}\|\Phi(t;u)\| &\leq 1 + \sum_{k=1}^{\infty} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \underbrace{\|U(s_1)\| \cdot \|U(s_2)\| \dots \|U(s_k)\|}_{\leq C^k} \mathrm{ds_k} \dots \mathrm{ds_1} \\ &\leq e^{Ct} \qquad \forall t \in [0,T], \forall u \in \mathcal{U}_r^{[0,T]} \ .\end{aligned}$$

Insbesondere ist die Menge aller zulässigen Lösungen

$$\left\{ \Phi(\cdot\,;u) \;\middle|\; u \in \mathcal{U}_r^{[0,T]} \right\} \tag{3.21}$$

gleichmässig beschränkt. Aus der Systemgleichung (BS1) folgt, dass auch die Ableitungen gleichmässig (essentiell) beschränkt sind. Es existiert also eine obere Schranke $M > 0$, so dass alle zulässigen Lösungen Lipschitz-stetig sind mit gemeinsamer Lipschitz-Konstante M:

$$\|\Phi(t;u) - \Phi(s;u)\| \leq \int_s^t \underbrace{\left\|\dot{\Phi}(\tau;u)\right\|}_{\leq M} \mathrm{d}\tau \leq M \cdot |t-s| \ \ \forall t,s \in [0,T], u \in \mathcal{U}_r^{[0,T]}$$

Wir dürfen daher den Satz von Arzelà-Ascoli A.9 auf die Menge (3.21) anwenden. Es gelte nun $u_k \overset{w}{\to} u$ für $k \to \infty$. Der Folgerung A.10 zufolge existiert zu jeder Teilfolge von $(\Phi(\cdot\,;u_k))_{k\in\mathbb{N}}$ eine Teilfolge $(v_k) \subseteq (u_k)$, so dass $\Phi(\cdot\,;v_k)$ gleichmässig gegen eine stetige Funktion $X(\cdot)$ konvergiert für $k \to \infty$. Wir zeigen nun, dass $X(\cdot) = \Phi(\cdot\,;u)$ gelten muß, d.h. jede Teilfolge von $(\Phi(\cdot\,;u_k))_{k\in\mathbb{N}}$ besitzt eine Teilfolge, die gleichmässig gegen $\Phi(\cdot\,;u)$ konvergiert. Offenbar muß dann jede Teilfolge von $(\Phi(\cdot\,;u_k))_{k\in\mathbb{N}}$ zumindest den Häufungspunkt $\Phi(\cdot\,;u)$ besitzen. Es ist

$$\begin{aligned}\Phi(t;v_k) = \mathrm{I} &+ \int_0^t \left(A(s) + \sum_{i=1}^m (v_k)_i(s) N_i(s) \right) [\Phi(s;v_k) - X(s)] \mathrm{ds} \\ &+ \int_0^t \left(A(s) + \sum_{i=1}^m (v_k)_i(s) N_i(s) \right) X(s) \mathrm{ds} \ .\end{aligned}$$

Für $k \to \infty$ konvergiert $\Phi(\cdot\,;v_k)$ gleichmässig gegen $X(\cdot)$, weshalb das obere Integral gegen Null konvergiert, und v_k konvergiert schwach gegen u, weshalb die Komponenten $(v_k)_i$ von v_k im unteren Integral schwach gegen die Komponenten u_i von u konvergieren. Es bleibt

$$X(t) = \mathrm{I} + \int_0^t \left(A(s) + \sum_{i=1}^m u_i(s) N_i(s) \right) X(s) \mathrm{ds} \qquad \forall t \in [0,T] \ .$$

Wir folgern $X(\cdot) = \Phi(\cdot\,; u)$.
Angenommen, $(\Phi(\cdot\,; u_k))_{k\in\mathbb{N}}$ konvergiert nicht gleichmässig gegen $\Phi(\cdot\,; u)$. Dann existiert eine Teilfolge mit höchstens einem Häufungspunkt ungleich $\Phi(\cdot\,; u)$. Dies ist nicht möglich, wie wir bereits festgestellt haben. □

Folgerung 3.39 (Stetigkeit der Responsefunktion)**.** Die Responsefunktion

$$\begin{aligned} \lambda : \qquad [0,T] \times \mathcal{U}_r^{[0,T]} &\longrightarrow \mathbb{R}^{n\times n} \\ (t,u) &\longmapsto \Phi(t;u) \end{aligned}$$

ist stetig, falls $\mathcal{U}_r^{[0,T]}$ mit der schwachen Topologie versehen ist.

Beweis. λ ist die Komposition stetiger Funktionen:

$$(t,u) \longmapsto \underbrace{(t, \Phi(\cdot\,; u))}_{\in [0,T]\times C^0[0,T]} \longmapsto \Phi(t;u)$$

□

Satz 3.40. Angenommen, der Steuerbereich $\mathcal{U}$ ist konvex und kompakt. Dann ist $\mathcal{U}_r^{[0,T]}$ eine schwach abgeschlossene Teilmenge von $L_2[0,T]$. Es folgt, dass $\mathcal{U}_r^{[0,T]}$ schwach kompakt ist.

Beweis. Da die messbaren Funktionen $u : [0,T] \to \mathcal{U}$ eine konvexe und beschränkte Menge in $L_2[0,T]$ bilden ($\mathcal{U}$ ist konvex und beschränkt), genügt es wegen Folgerung A.26 zu zeigen, dass $\mathcal{U}_r^{[0,T]}$ dort abgeschlossen ist.
Dazu sei $(f_k) \subseteq \mathcal{U}_r^{[0,T]}$ irgendeine konvergente Folge in $L_2[0,T]$ mit Grenzwert f. Wie jede konvergente Folge in $L_2[0,T]$ besitzt (f_k) eine Teilfolge (f_{k_i}), die punktweise fast überall gegen den Grenzwert f konvergiert [7, p.96]. Natürlich ist f messbar und

$$f(t) = \lim_{i\to\infty} \underbrace{f_{k_i}(t)}_{\in\mathcal{U}} \in \mathcal{U}$$

fast überall ($\mathcal{U}$ ist abgeschlossen). Es folgt $f \in \mathcal{U}_r^{[0,T]}$ (bis auf Äquivalenz in $L_2[0,T]$). Also ist $\mathcal{U}_r^{[0,T]}$ eine beschränkte, konvexe und schwach abgeschlossene Teilmenge des Hilbertraums $L_2[0,T]$. Die Folgerung $A.28$ besagt, dass dann $\mathcal{U}_r^{[0,T]}$ schwach kompakt ist. □

Folgerung 3.41. Es sei $\mathcal{U}$ kompakt und konvex. (Man beachte, dass der Steuerbereich eines Systems nichtleer ist.)
Dann sind die erreichbare Menge $\mathcal{A}_T$ und die Vereinigung $\mathcal{A}_{\leq T} := \bigcup_{0\leq t\leq T} \mathcal{A}_t$ nichtleer, kompakt und wegzusammenhängend für alle $T \geq 0$.

Beweis. Natürlich ist $\mathcal{U}_r^{[0,T]}$ eine konvexe und daher wegzusammenhängende Menge, falls $\mathcal{U}$ konvex ist. Wie wir in Folgerung 3.39 festgestellt haben, sind die Mengen

$$\mathcal{A}_T = \left\{\Phi(T;u) \;\middle|\; u \in \mathcal{U}_r^{[0,T]}\right\} \quad \text{und} \quad \mathcal{A}_{\leq T} = \left\{\Phi(t;u) \;\middle|\; u \in \mathcal{U}_r^{[0,T]}, t \in [0,T]\right\}$$

Bilder von (schwach) kompakten (Satz 3.40) und wegzusammenhängenden Mengen unter einer stetigen Funktion. Es folgt die Behauptung. □

Folgerung 3.42. Betrachte das assoziierte Vektorsystem von (BS1)

$$\dot{x}(t) = \left(A(t) + \sum_{k=1}^{m} v_k(t) N_k(t)\right) x(t),\ u(t) \in \mathcal{U},\ x(t) \in \mathbb{R}^n,\ t \in [0,T] \tag{3.22}$$

und das inhomogene System

$$\dot{x}(t) = A(t)x(t) + \sum_{k=1}^{m} u_k(t) N_k(t) x(t) + B(t)u(t),\quad u(t) \in \mathcal{U},\ t \in [0,T] \tag{3.23}$$

mit zusätzlichem additiven Term „$B(t)u(t)$“, wobei $B : [0,T] \to \mathbb{R}^{n\times m}$ messbar und lokal (essentiell) beschränkt ist.

Falls der Steuerbereich $\mathcal{U}$ konvex und kompakt ist, so sind die erreichbaren Mengen $\mathcal{A}_T(p)$ dieser Systeme kompakt und wegzusammenhängend für alle $T \geq 0$ und $p \in \mathbb{R}^n$.

Beweis. 1. Da die Abbildung $p \mapsto \Phi(T;u)\cdot p$ stetig ist, sind die Eigenschaften aus Folgerung 3.41 auf die erreichbare Menge $\mathcal{A}_T(p) = \mathcal{A}_T \cdot p$ des assoziierten Vektorsystems (3.22) übertragbar.

2. Das inhomogene System (3.23) können wir homogenisieren. Auf diese Weise erhalten wir ein homogenes System der Form (3.22) im $\mathbb{R}^{n+1}$.
Laut Bemerkung 3.10 sind die erreichbaren Mengen $\mathcal{A}_T(p)$ von (3.23) (bis auf Einbettung) gegeben durch

$$\mathcal{A}_T^{n+1} \cdot \binom{p}{1} \ ,$$

wobei $\mathcal{A}_T^{n+1} = \mathcal{A}_T^{n+1}(\mathrm{I})$ eine erreichbare Menge vom assoziierten Matrixsystem des homogenisierten Systems ist.
Wir folgern wie in 1., dass auch die erreichbaren Mengen von inhomogenen Systemen kompakt und wegzusammenhängend sind. □

3.4.2 Das Bang-Bang Prinzip

Der Steuerbereich $\mathcal{U}$ des homogenen Systems (BS1) sei kompakt und konvex. Zusätzlich zu (BS1) berücksichtigen wir ein weiteres System in $\mathrm{GL}(n,\mathbb{R})$

$$\dot{X}(t) = \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) X(t),\ v(t) \in \mathcal{V},\ X(0) = \mathrm{I},\ t \in [0,T]\ , \tag{BS2}$$

welches einen kleineren Steuerbereich $\mathcal{V} \subseteq \mathcal{U}$ besitzt—aber sonst mit (BS1) übereinstimmt. Genauer, $\mathcal{V}$ sei der Abschluß der extremen Punkte aus der konvexen kompakten Menge $\mathcal{U}$.
Dies sind genau die Punkte $v \in \mathcal{U}$, für die aus $u_1, u_2 \in \mathcal{U}$, $\lambda \in (0,1)$ und $v = \lambda u_1 + (1-\lambda)u_2$ stets $v = u_1 = u_2$ folgt. Notwendigerweise liegen extreme Punkte auf dem Rand von $\mathcal{U}$.

Bezeichnung 3.43. Wir bezeichnen die Menge aller extremen Punkte—manchmal sagt man auch „extremal“ statt „extrem“—von $\mathcal{U}$ mit $\operatorname{ext}\mathcal{U}$. In dieser Schreibweise ist dann $\mathcal{V} = \overline{\operatorname{ext}\mathcal{U}}$.
Die konvexe Hülle einer nichtleeren Menge $\mathcal{N} \subseteq \mathbb{R}^m$ wird mit $\operatorname{cvx}\mathcal{N}$ bezeichnet. Sie ist der Durchschnitt aller $\mathcal{N}$ enthaltenden konvexen Teilmengen des $\mathbb{R}^m$.

Aus der konvexen Analysis ist bekannt, dass $\mathcal{V}$ eine nichtleere kompakte Menge ist [17, p.40]. Der Satz von Krein-Milmann [17, p.41,202] besagt, dass $\mathcal{U}$ die konvexe Hülle von $\mathcal{V}$ ist, d.h.

$$\mathcal{U} = \operatorname{cvx}\overline{\operatorname{ext}\mathcal{U}} = \operatorname{cvx}\mathcal{V}\ .$$

Wir unterscheiden zwei Klassen von zulässigen Steuerfunktionen:

(i) $\mathcal{U}_r^{[0,T]}$ ist die Klasse der messbaren Funktionen, definiert auf $[0,T]$, mit Werten in $\mathcal{U}$.

(ii) $\mathcal{V}_b^{[0,T]}$ ist die Klasse der stückweise konstanten Funktionen, definiert auf $[0,T]$, mit Werten in $\mathcal{V}$.

Die Klasse (i) gehört zum Ausgangssystem (BS1), und (ii) zu System (BS2).

Beispiel 3.44. Ist $\mathcal{U}$ der Einheitswürfel

$$E^m = \{u \in \mathbb{R}^m \mid -1 \le u_i \le 1, i = 1,\dots,m\}\ ,$$

so ist $\mathcal{V}$ die Punktmenge $\{u \in \mathbb{R}^m \mid u_i = \pm 1, i = 1,\dots,m\}$. Die Funktionen $v \in \mathcal{V}_b^{[0,T]}$ sind stückweise konstante Bang-Bang Funktionen.

Bezeichnung 3.45. Für die erreichbaren Mengen von (BS2) führen wir eine neue Notation ein. Es sei

$$\mathcal{A}_T^b := \left\{ \Phi(T;u) \;\middle|\; u \in \mathcal{V}_b^{[0,T]} \right\}$$

die erreichbare Menge von (BS2) zum Zeitpunkt $T \geq 0$, wobei $\Phi(\cdot\,;u)$ die Fundamentallösung von (BS2) zur Kontrolle $u(\cdot\,)$ bezeichnet.

Ähnlich definieren wir durch

$$\mathcal{A}_T^b(p) := \mathcal{A}_T^b \cdot p$$

die erreichbaren Mengen des assoziierten Vektorsystems in $\mathbb{R}^n$

$$\dot{x}(t) = \left(A(t) + \sum_{k=1}^{m} v_k(t) N_k(t) \right) x(t),\ u(t) \in \mathcal{V},\ x(0) = p,\ t \in [0,T] \quad (3.24)$$

mit festem Anfangswert $p \in \mathbb{R}^n$.

Folgerung 3.46. Die erreichbare Menge $\mathcal{A}_T^b$ ist wegzusammenhängend für alle $T \geq 0$. (Wegen Folgerung 3.42 gilt dies entsprechend auch für $\mathcal{A}_T^b(p)$.)

Beweis. Es genügt zu zeigen, dass $\mathcal{V}_b^{[0,T]}$ schwach wegzusammenhängend ist. Dann können wir wie im Beweis von 3.41 diese Eigenschaft durch die stetige Ein-Ausgangsfunktion auf $\mathcal{A}_T^b$ übertragen.
Es seien u und v Elemente aus $\mathcal{V}_b^{[0,T]}$. Für alle $t \in [0,T]$ setzen wir

$$w_t(s) := \begin{cases} u(s) & \text{falls } 0 \leq s \leq t \\ v(s) & \text{falls } t < s \leq T \end{cases} .$$

Offenbar ist $w_t(\cdot\,) \in \mathcal{V}_b^{[0,T]}$ für alle $t \in [0,T]$. Weiter ist $w_0 = u$ und $w_T = v$. Falls $t \mapsto w_t$ eine stetige Abbildung vom Intervall $[0,T]$ nach $\mathcal{V}_b^{[0,T]}$—versehen mit der schwachen Topologie—ist, so haben wir einen „Weg“ gefunden, der die Punkte u und v verbindet, und das Lemma wäre bewiesen.
In der Tat folgt aus $t_k \overset{<}{\to} t$ für $k \to \infty$, dass

$$\|w_{t_k} - w_t\|_{L_2}^2 = \int_0^T \|w_{t_k}(s) - w_t(s)\|_{\mathbb{R}^m}^2 \mathrm{ds} = \int_{t_k}^t \|u(s) - v(s)\|_{\mathbb{R}^m}^2 \mathrm{ds} \longrightarrow 0$$

für $k \to \infty$. (Rechtseitige Stetigkeit lässt sich analog beweisen.)
Weil $w_{t_k} \overset{s}{\to} w_t$ die schwache Konvergenz $w_{t_k} \overset{w}{\to} w_t$ impliziert, ist das Folgenkriterium für Stetigkeit erfüllt. Dies endet den Beweis. □

Definition 3.47. Es sei Σ ein kontinuierliches Kontrollsystem mit kompaktem konvexen Steuerbereich $\mathcal{U}$. Weiter sei $\mathcal{V} = \overline{\text{ext}\,\mathcal{U}}$.
$\mathcal{A}_T(p)$ und $\mathcal{A}_T^b(p)$ bezeichnen die erreichbaren Mengen von einem festen Anfangszustand p zur Zeit $T \geq 0$ bezüglich Kontrollfunktionen aus $\mathcal{U}_r^{[0,T]}$ bzw. $\mathcal{V}_b^{[0,T]}$.
Wir sagen, dass

- das *approximative Bang-Bang Prinzip* für das System Σ gültig ist, falls $\mathcal{A}_T^b(p)$ dicht in $\mathcal{A}_T(p)$ liegt für alle $T \geq 0$.

- das *schwache Bang-Bang Prinzip* gültig ist, falls es zu jedem erreichbaren Punkt $q \in \mathcal{A}(p)$ eine zeitoptimale stückweise konstante Kontrolle $u \in \mathcal{V}_b^{[0,T]}$ gibt, die p nach q in minimaler Zeit steuert.

- das *starke Bang-Bang Prinzip* gültig ist, falls $\mathcal{A}_T^b(p) = \mathcal{A}_T(p)$ für alle $T \geq 0$.

In vielen Optimierungsproblemen ist das Bang-Bang Prinzip ein wichtiges Hilfsmittel zur Berechnung optimaler Steuerungen. Ist etwa das schwache oder starke Bang-Bang Prinzip erfüllt und der Steuerbereich $\mathcal{V}$ gleich dem Einheitswürfel E^m, so können wir aus dem Wissen, dass ein Zustand q zu einer Zeit T erreichbar ist, folgern, dass es eine Bang-Bang Funktion gibt, die den Zielpunkt $q \in \mathbb{R}^n$ in minimaler Zeit $t_{\min} \leq T$ erreicht. Wir können dann versuchen die Umschaltzeitpunkte dieser Bang-Bang Funktion zu bestimmen. Eine genauere Analyse ergibt eventuell eine obere Schranke für die Anzahl der benötigten Schaltungen. Auf diese Weise könnte die zeitoptimale Bang-Bang Steuerung durch einen endlich-dimensionalen Parameter parametrisiert werden, der über eine kompakte Menge variiert. Ein solches Vorgehen hat sich bei vielen Problemstellungen bewährt.
Falls das approximative Bang-Bang Prinzip beispielsweise für das bilineare System (BS1) gültig ist, so verhält sich das System (BS2) mit reduziertem Steuerbereich $\mathcal{V}$ sehr ähnlich zum Ausgangssystem (BS1). Wir können uns daher häufig auf das Studium des „Einfacheren“ dieser beiden Systeme beschränken.
Es ist wohlbekannt [9, p.119], dass für gewöhnliche lineare Kontrollsysteme mit kompaktem und konvexen Steuerbereich das schwache Bang-Bang Prinzip gültig ist. Falls dieser Steuerbereich sogar ein Polytop ist, d.h. die konvexe Hülle endlich vieler Punkte aus $\mathbb{R}^m$, so gilt sogar das starke Bang-Bang Prinzip [9, p.116].
Es stellt sich die Frage, ob sich für bilineare Systeme vergleichbare Resultate herleiten lassen. Im Verlauf dieser Arbeit erarbeiten wir die folgenden:

- Das approximative Bang-Bang Prinzip ist für alle bilinearen Systeme mit kompaktem konvexen Steuerbereich gültig.
- Mit Hilfe des Maximumprinzips können wir nachweisen, dass das schwache Bang-Bang Prinzip für gewisse bilineare Systeme von Rang 1 erfüllt ist (siehe Satz 3.88).
- Das starke Bang-Bang Prinzip gilt nur unter sehr restriktiven Voraussetzungen für bilineare Systeme (siehe Satz 3.52).

Das approximative Bang-Bang Prinzip

Wie angekündigt, wollen wir beweisen, dass das approximative Bang-Bang Prinzip für bilineare Systeme mit kompaktem konvexen Steuerbereich gültig ist.
Das zentrale Hilfsmittel wird wie im letzten Teilabschnitt die Stetigkeit der Ein-Ausgangsfunktion gemäß Satz 3.38 sein. Doch zunächst brauchen wir eine Aussage aus der Funktionalanalysis [1, p.236] zur schwachen Konvergenz oszillierender Funktionen.

Hilfssatz 3.48. Es sei $g \in L^\infty(\mathbb{R}, \mathbb{R}^m)$ eine oszillierende Funktion mit Periode $\kappa > 0$, d.h. $g(t+\kappa) = g(t)$ für fast alle t, und

$$\frac{1}{\kappa}\int_0^\kappa g(s)\mathrm{ds} = C$$

für ein $C \in \mathbb{R}^m$. Dann konvergieren die Funktionen $f_k(t) := g(kt)$ für $k \to \infty$ schwach in $L_2[0,T]$ gegen die konstante Funktion C.

Beweis. Ohne Einschränkung dürfen wir $C = 0$ voraussetzen (sonst verwende $g - C$ statt g). Da $g \in L^\infty(\mathbb{R}, \mathbb{R}^m)$ und $C = 0$ folgt, dass

$$h(t) := \int_0^t g(s)\mathrm{ds}$$

eine auf ganz $\mathbb{R}$ beschränkte stetige Funktion ist. Denn ist $t \in [\lambda\kappa, (\lambda+1)\kappa]$ für irgendein $\lambda \in \mathbb{N}_0$, so erhalten wir aufgrund der Periodizät von g eine obere Schranke von h, die unabhängig von t ist:

$$h(t) = \int_0^t g(s)\mathrm{ds} = \sum_{i=0}^{\lambda-1} \underbrace{\int_{i\kappa}^{(i+1)\kappa} g(s)\mathrm{ds}}_{=\int_0^\kappa g(s)\mathrm{ds}=\kappa C=0} + \int_{\lambda\kappa}^t g(s)\mathrm{ds} = \int_0^{t-\lambda\kappa} g(s)\mathrm{ds} \le \|g\|_\infty \kappa$$

Auf jedem Teilintervall $[a,b] \subseteq [0,T]$ gilt

$$\int_a^b f_k(s)\mathrm{d}s = \int_a^b g(ks)\mathrm{d}s = \frac{1}{k}\int_{ka}^{kb} g(s)\mathrm{d}s = \frac{1}{k}\underbrace{(h(kb)-h(ka))}_{\text{beschänkt}} \longrightarrow 0 \quad (3.25)$$

für $k \to \infty$. Es sei nun $\phi : [0,T] \to \mathbb{R}^m$ eine Treppenfunktion der Form

$$\phi(t) = \sum_{j=1}^{r} \alpha_j \chi_{I_j}(t) \ ,$$

wobei die paarweise disjunkten Intervalle $I_1, \ldots, I_r$ das Intervall $[0,T]$ zerlegen und $\alpha_j \in \mathbb{R}^m$. Mit χ_I wird die charakteristische Funktion von einem Intervall I bezeichnet, definiert durch

$$\chi_I(t) := \begin{cases} 1 \ , & \text{falls } t \in I \\ 0 \ , & \text{falls } t \in \mathbb{R}\backslash I \end{cases} \ .$$

Aus (3.25) ergibt sich

$$\int_0^T \left\langle f_k(s), \phi(s) \right\rangle_{\mathbb{R}^m} \mathrm{d}s = \sum_{j=1}^{r} \alpha_j^* \int_{I_j} f_k(s)\mathrm{d}s \longrightarrow 0 \qquad \text{für } k \to \infty \ .$$

Weil dies für jede Treppenfunktion gilt und weil (f_k) beschränkt ist, konvergiert f_k schwach gegen C. □

Satz 3.49. $\mathcal{V}_b^{[0,T]}$ ist schwach dicht in $\mathcal{U}_r^{[0,T]}$, d.h. der schwache Abschluss von $\mathcal{V}_b^{[0,T]}$ ist ganz $\mathcal{U}_r^{[0,T]}$. (Zur Erinnerung: $\mathcal{U} = \mathrm{cvx}\ \mathcal{V}$.)

Beweis. 1. Ohne Einschränkung sei $0 \in \mathcal{V}$ (sonst argumentiere mit $\mathcal{V} - v_0$ und $\mathcal{U} - v_0$ für ein festes $v_0 \in \mathcal{V}$). Es bezeichne W den schwachen Abschluss von $\mathcal{V}_b^{[0,T]}$, d.h. die Menge aller Grenzwerte von schwach konvergenten Folgen aus $\mathcal{V}_b^{[0,T]}$.

2. Zunächst wird gezeigt, dass W alle Funktionen der Form $u_0 \cdot \chi_{[a,b]}$ mit $[a,b] \subseteq [0,T]$ und $u_0 \in \mathcal{U} = \mathrm{cvx}\ \mathcal{V}$ enthält.
Da u_0 in der konvexen Hülle von $\mathcal{V}$ liegt, lässt sich u_0 als Konvexkombination

$$u_0 = \sum_{i=1}^{N} \lambda_i v_i \qquad \text{mit } \lambda_i > 0, \sum_{i=1}^{N} \lambda_i = 1$$

von Elementen $v_1, \ldots, v_N$ aus $\mathcal{V}$ darstellen. Wir konstruieren eine periodische stückweise konstante Funktion mit Werten in $\mathcal{V}$, die schwach gegen u_0 konvergiert. Dazu setze

$$g(t) := \begin{cases} v_1 & \text{für } i \le t < i + \lambda_1, i \in \mathbb{Z}, \\ v_2 & \text{für } i + \lambda_1 \le t < i + \lambda_1 + \lambda_2, i \in \mathbb{Z}, \\ \vdots & \vdots \\ v_N & \text{für } i + \lambda_1 + \lambda_2 + \cdots \lambda_{N-1} \le t < i + 1, i \in \mathbb{Z} \ . \end{cases}$$

Die Funktion hat die Periode $\kappa = 1$, und es ist

$$\int_0^1 g(s)\mathrm{ds} = \sum_{j=1}^{N} \lambda_j v_j = u_0 \ .$$

Nach Hilfssatz (3.48) konvergiert $f_k(t) := g(kt)$ schwach in $L_2[0,T]$ gegen u_0 für $k \to \infty$. Wegen $0 \in \mathcal{V}$ ist auch $f_k \cdot \chi_{[a,b]}$ ein Element aus $\mathcal{V}_b^{[0,T]}$. Weiter gilt für jede Funktion $\phi \in L_2[0,T]$

$$\langle f_k \chi_{[a,b]}, \phi \rangle_{L_2} = \langle f_k, \phi \chi_{[a,b]} \rangle_{L_2} \longrightarrow \langle u_0, \phi \chi_{[a,b]} \rangle_{L_2} = \langle u_0 \chi_{[a,b]}, \phi \rangle_{L_2} \ \text{ für } k \to \infty \ ,$$

da (f_k) schwach gegen u_0 konvergiert. Elemente der Form $u_0 \chi_{[a,b]}$ liegen daher in W.

3. Weil endliche Summen von Elementen aus W in W liegen, enthält W auch alle Treppenfunktionen der Form $\sum_{j=1}^{N} u_j \chi_{I_j}$ mit $u_j \in \mathcal{U}$ und disjunkten Intervallen $I_j \subseteq [0,T]$. Solche Treppenfunktionen liegen (schwach) dicht in $\mathcal{U}_r^{[0,T]}$, d.h. $\mathcal{U}_r^{[0,T]}$ liegt in W. Da $\mathcal{U}_r^{[0,T]}$ schwach abgeschlossen ist (Satz 3.40), ist der schwache Abschluss von $\mathcal{V}_b^{[0,T]}$ gleich $\mathcal{U}_r^{[0,T]}$. □

Folgerung 3.50 (approximatives Bang-Bang Prinzip)**.** Die erreichbare Menge $\mathcal{A}_T^b$ von (BS2) liegt dicht in der erreichbaren Menge $\mathcal{A}_T$ von (BS1).

Beweis. Es sei $\Phi(T;u) \in \mathcal{A}_T$, d.h. $u \in \mathcal{U}_r^{[0,T]}$.
Nach Satz 3.49 existiert eine Folge $(u_k) \subset \mathcal{V}_b^{[0,T]}$, die schwach gegen u konvergiert. Aus Satz 3.38 folgt, dass $\Phi(T;u_k) \in \mathcal{A}_T^b$ stark gegen $\Phi(T;u)$ konvergiert für $k \to \infty$. Die Punkte von $\mathcal{A}_T$ liegen also im Abschluß von $\mathcal{A}_T^b$. Da $\mathcal{A}_T$ abgeschlossen ist, ist $\mathcal{A}_T$ gleich dem Abschluss von $\mathcal{A}_T^b$. □

Bemerkung 3.51. Wie in Folgerung 3.42 können wir nachweisen, dass das approximative Bang-Bang Prinzip auch für allgemeinere bilineare Systeme der Form (3.23) mit kompaktem konvexen Steuerbereich gültig ist.

Das starke Bang-Bang Prinzip

Wir werden gleich sehen, dass das starke Bang-Bang Prinzip im Allgemeinen für bilineare Systeme nicht erfüllt ist. Dennoch können wir ein Theorem [24, p.473] zitieren, das für eine sehr kleine Klasse bilinearer Systeme der Form (BS1) die Gleichheit $\mathcal{A}_T^b = \mathcal{A}_T$ für alle $T \geq 0$ sicherstellt.
Wegen Folgerung 3.50 gilt $\overline{\mathcal{A}_T^b} = \mathcal{A}_T$ für solche Systeme. Es ist daher klar, dass die Abgeschlossenheit von $\mathcal{A}_T^b$ hinreichend und notwendig für die Gültigkeit des starken Bang-Bang Prinzips ist.

Satz 3.52 (starkes Bang-Bang Prinzip)**.** Betrachte das System in $\mathrm{GL}(n, \mathbb{R})$

$$\dot{X}(t) = \left(A(t) + \sum_{k=1}^{m} u_k(t) N_k(t) \right) X(t),\ u(t) \in E^m,\ X(0) = \mathrm{I},\ t \in [0, T] \tag{3.26}$$

mit dem Einheitswürfel als Steuerbereich und beschränkten, stückweise analytischen Funktionen $A, N_1, \ldots N_m : [0, T] \to \mathbb{R}^{n \times n}$.
Angenommen, es gilt

$$[A(t), N_i(s)] = 0 \quad \text{und} \quad [N_j(t), N_i(s)] = 0 \quad \forall t, s \in [0, T]\ \forall i, j = 1, \ldots m\ .$$

Dann ist die erreichbare Menge $\mathcal{A}_T^b = \left\{ \Phi(T; 0, u) \;\middle|\; u \in \mathcal{V}_b^{[0,T]} \right\}$ abgeschlossen, wobei $\mathcal{V}$ die Menge der Randpunkte von E^m ist.
(Folglich ist das starke Bang-Bang Prinzip gültig für (3.26), d.h. für alle $P \in \mathcal{A}_T$ existiert eine stückweise konstante Bang-Bang Funktion, die I nach P zur Zeit T steuert.)

Beweisskizze. Es werden nur die beiden Beweisschritte aus dem Beweis in [24, p.473] angegeben.

1. Nach Voraussetzung ist für ein vorgegebenes $u \in \mathcal{U}_r^{[0,T]}$ (hier $\mathcal{U} = E^m$)

$$\left[A(t) + \sum_{k=1}^{m} u_k(t) N_k(t), A(s) + \sum_{k=1}^{m} u_k(s) N_k(s) \right] = 0 \qquad \forall t, s \in [0, T]\ .$$

Aus Satz B.20 erhalten wir

$$\Phi(T; 0, u) = \exp \int_0^T \left(A(s) + \sum_{k=1}^{m} u_k(s) N_k(s) \right) \mathrm{ds}\ . \tag{3.27}$$

Es ist wohlbekannt, dass für zwei kommutierende Matrizen A und B stets $e^{A+B} = e^A \cdot e^B$ gilt. (3.27) wird auf diese Weise zu

$$\Phi(T;0,u) = \exp\left(\int_0^T A(s)\mathrm{ds}\right) \cdot \prod_{k=1}^{m} \exp\left(\int_0^T u_k(s)N_k(s)\mathrm{ds}\right) \ . \qquad (3.28)$$

2. Der Satz von Ljapunov-Halkin [9, p.112] besagt, dass für jede analytische Funktion $F : [0,T] \to \mathbb{R}^{n\times n}$ das Folgende gilt:
Jedes Element aus der Menge

$$\left\{ \int_0^T u_0(s)F(s)\mathrm{ds} \;\middle|\; u_0 : [0,T] \to [-1,1] \text{ messbar} \right\}$$

ist von der Form $\int_0^T v_0(s)F(s)ds$ für eine stückweise konstante Funktion $v_0 : [0,T] \to \{-1,1\}$.

Die Schritte 1. und 2. führen zum Beweis. □

Das nächste Beispiel stammt von Héctor J. Sussmann [24]. Es soll zeigen, dass die Kommutations-Bedingung aus Satz 3.52 nicht abgeschwächt werden kann. Dazu wählt Sussmann ein autonomes bilineares Single-Input System der Form

$$\dot{X}(t) = (A + u(t)N)X(t), \qquad -1 \le u(t) \le 1, \qquad X(t) \in \mathbb{R}^{n\times n} \ , \quad (3.29)$$

bei dem die Matrizen A und N nicht kommutieren, und weist für dieses System nach, dass $\mathcal{A}_T^b \subsetneq \mathcal{A}_T$ gilt. Das starke Bang-Bang Prinzip ist also für bilineare Systeme i.A. ungültig.

Beispiel 3.53 (Sussmanns Gegenbeispiel)**.** Es sei das System aus Beispiel 3.34 vorgegeben, d.h. $n = 4$, $A = E_{34}$ und $N = E_{12} + E_{23}$ in (3.29).
Die Fundamentallösung $\Phi(\cdot\,;u)$ läßt sich als endliche Volterra-Reihe darstellen. In Beispiel 4.24 berechnen wir

$$\Phi(T;u) = \begin{pmatrix} 1 & \int_0^T u & \int_0^T \int_0^{s_1} u(s_1)u(s_2)\ \mathrm{ds}_2\mathrm{ds}_1 & \int_0^T \int_0^{s_1} s_2u(s_1)u(s_2)\ \mathrm{ds}_2\mathrm{ds}_1 \\ 0 & 1 & \int_0^T u & \int_0^T su \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

Wir setzen $v(T) := \int_0^T u(s)ds$ und beseitigen durch partielle Integrationen

die Kontrolle u aus der Lösungsdarstellung. Es ist

$$\begin{aligned}
&\int_0^T su(s)\mathrm{ds} = Tv(T) - \int_0^T v(s)\mathrm{ds}\ ,\\
&\int_0^T \int_0^{s_1} u(s_1)u(s_2)\ \mathrm{ds}_2\mathrm{ds}_1 = \int_0^T u(s)v(s)\ \mathrm{ds} = \frac{1}{2}\int_0^T \frac{\mathrm{d}v^2}{\mathrm{ds}}(s)\ \mathrm{ds} = \frac{1}{2}v(T)^2\ ,\\
&\int_0^T \int_0^{s_1} s_2u(s_1)u(s_2)\ \mathrm{ds}_2\mathrm{ds}_1 = \int_0^T u(s_1)\int_0^{s_1} s_2u(s_2)\ \mathrm{ds}_2\mathrm{ds}_1\\
&\quad = v(T)\int_0^T su(s)\mathrm{ds} - \int_0^T v(s)u(s)s\mathrm{ds}\\
&\quad = v(T)\int_0^T su(s)\mathrm{ds} - \frac{1}{2}\int_0^T \frac{\mathrm{d}v^2}{\mathrm{ds}}(s)s\mathrm{ds}\\
&\quad = v(T)\int_0^T su(s)\mathrm{ds} - \frac{1}{2}Tv(T)^2 + \frac{1}{2}\int_0^T v(s)^2\mathrm{ds}\\
&\quad = \frac{1}{2}Tv(T)^2 - v(T)\int_0^T v(s)\mathrm{ds} + \frac{1}{2}\int_0^T v(s)^2\mathrm{ds}\ .
\end{aligned}$$

Auf diese Weise erhalten wir

$$\Phi(T;u) = \begin{pmatrix} 1 & v(T) & \frac{1}{2}v(T)^2 & \frac{1}{2}Tv(T)^2 - v(T)\int_0^T v + \frac{1}{2}\int_0^T v^2\\ 0 & 1 & v(T) & Tv(T) - \int_0^T v\\ 0 & 0 & 1 & T\\ 0 & 0 & 0 & 1 \end{pmatrix}\ .$$

Mit $\Phi(\cdot;0)$ und $\Phi(\cdot;1)$ bezeichnen wir die Fundamentallösungen zu den konstanten Kontrollen $u \equiv 0$ bzw. $u \equiv 1$.
Offenbar ist $\Phi(T;0) = \mathrm{I} + TE_{34} \in \mathcal{A}_T$. Angenommen, es gibt eine weitere zulässige Kontrolle u, welche $\Phi(T;0)$ erreicht zur Zeit $T > 0$. Dann gilt notwendigerweise $\Phi(T;u) = \mathrm{I}+TE_{34}$. Es bezeichne Φ_{ij} die (i,j)-te Komponente von $\Phi(T;u)$. Dann folgt schrittweise:

$$0 = \Phi_{12} = v(T) \Longrightarrow 0 = \Phi_{24} = -\int_0^T v(s)\mathrm{ds} \Longrightarrow 0 = \Phi_{14} = \frac{1}{2}\underbrace{\int_0^T v(s)^2\mathrm{ds}}_{\|v\|^2_{L_2[0,T]}}$$

Dies impliziert, dass $v(t) = 0$ fast überall gilt. Da u fast überall die Ableitung von v ist, muß auch $u(t) = 0$ fast überall gelten. Wir sehen somit, dass es keine Bang-Bang Steuerung $u : [0,T] \to \pm 1$ gibt, die I nach $\mathrm{I} + TE_{34} \in \mathcal{A}_T$ steuert. Es folgt $\mathcal{A}_T^b \subsetneq \mathcal{A}_T$.
Da $\mathrm{I}+TE_{34}$ höchstens zum Zeitpunkt T erreichbar gewesen wäre, sind sowohl

das schwache als auch das starke Bang-Bang Prinzip ungültig.

Wir nutzen die Gelegenheit, um noch zu zeigen, dass $\mathcal{A}_T$ zu keiner Zeit $T > 0$ konvex ist. Dazu betrachte die Konvexkombination

$$P = \frac{1}{2}\Phi(T;0) + \frac{1}{2}\Phi(T;1) = \begin{pmatrix} 1 & \frac{T}{2} & \frac{1}{4}T^2 & * \\ 0 & 1 & * & * \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

Es ist $\frac{1}{2}P_{12}^2 \neq P_{13}$, weshalb es keine zulässige Kontrolle u mit $P = \Phi(T;u)$ geben kann. $\mathcal{A}_T$ kann folglich nicht konvex sein für $T > 0$.
In Abbildung 3.2 ist die erreichbare Menge

$$\mathcal{A}_{0.5} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

des assoziierten Vektorsystems in der (x, y) Ebene dargestellt. Sie ist beschränkt und zusammenhängend, aber nicht konvex.

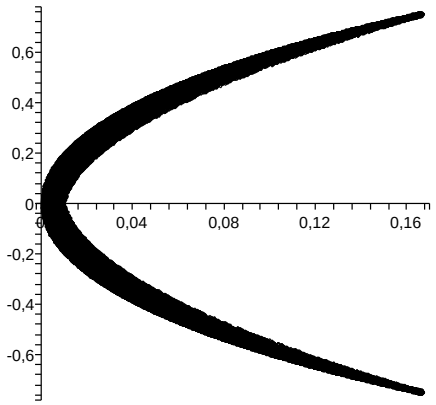

Abbildung 3.2: Erreichbare Menge in (x,y) Ebene

3.5 Bilineare Systeme von Rang 1

Otomar Hàjek stellt in [9] fest, dass sich ein allgemeines bilineares System bezüglich seiner Analyse fast so schlecht verhält wie ein x-beliebiges nichtlineares System. Als Beispiel für „extremes Fehlverhalten“ führt er Sussmanns

Beispiel auf, das wir in Beispiel 3.53 kennengelernt haben. Denn hier scheitert das schwache Bang-Bang Prinzip bereits an einem einfachen bilinearen Single-Input System.
Er konzentriert sich daher auf eine Unterklasse bilinearer Systeme—nämlich auf bilineare Systeme von Rang 1, für die man bedeutsame Resultate (z.B. Konvexität erreichbarer Mengen, schwaches Bang-Bang Prinzip) herleiten kann, die genauso leicht verständlich sind wie vergleichbare Resultate aus der linearen Kontrolltheorie. In Beispiel 2.34 haben wir gesehen, dass diese Klasse eine bilineare Interpretation von linearen Systemen enthält. Die gesamte Theorie linearer Kontrollsysteme wird somit zu einem Spezialfall der Theorie dieses Abschnitts.
Weitere wichtige Anwendungen von bilinearen Systemen von Rang 1 sind die Modellierung des Schaltens zwischen mehreren dynamischen Systemen (Beispiel 2.33) und die parametrische Kontrolle von Systemen, die durch eine lineare Differentialgleichung n-ter Ordnung gegeben sind (Beispiel 2.32).

Es gibt zwei Typen von bilinearen Systemen von Rang 1: Spalten- und Zeilen-Kontrollsysteme (siehe (2.24) und (2.25)). Unsere Aufmerksamkeit wird sich allerdings nur auf die erstgenannten richten.

3.5.1 Spalten-Kontrollsysteme mit kompaktem Steuerbereich

Das sind Systeme der Form

$$\dot{x}(t) = (A + u(t)c^*)x(t), \qquad u(t) \in \mathcal{U}, \qquad x(t) \in \mathbb{R}^n \ , \tag{3.30}$$

mit Parametern $A \in \mathbb{R}^{n\times n}$ und $c \in \mathbb{R}^n$. Dabei ist $\mathcal{U}$ eine kompakte nichtleere Teilmenge des $\mathbb{R}^n$. Eine Steuerung $u : I \to \mathcal{U}$ ist genau dann zulässig, wenn sie messbar ist. Außerdem fordern wir o.B.d.A., dass $0 \in \mathcal{U}$ gilt.

Voraussetzung 3.54. $0 \in \mathcal{U}$

Dies ist keine Einschränkung, da man anstelle von (3.30) für ein festes $u_0 \in \mathcal{U}$ das äquivalente Rang-1 System

$$\dot{x}(t) = ((A + u_0c^*) + u(t)c^*)\, x(t), \qquad u(t) \in \mathcal{U} - u_0$$

untersuchen könnte.

Erinnerung 3.55. Rang-1 Systeme sind homogen und autonom. Die erreichbare Menge von $p \in \mathbb{R}^n$ zur Zeit $t > 0$ ist

$$\mathcal{A}_t(p) = \mathcal{A}_t \cdot p = \{\Phi(t; u)\cdot p \mid u : [0, t] \to \mathcal{U} \text{ messbar } \} \ ,$$

wobei $\Phi(\cdot\,;u)$ die Fundamentallösung des assoziierten Matrixsystems ist. Für $\Phi(\cdot\,;u)$ gelten sämtliche Eigenschaften aus Folgerung 2.21. Insbesondere haben wir die Darstellung als Neumann-Reihe

$$\Phi(t;u) = \mathrm{I} + \sum_{k=1}^{\infty} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} U(s_1)\dots U(s_k)\mathrm{ds_k}\dots\mathrm{ds_2}\mathrm{ds_1} \ ,$$

wobei

$$U(t) := A + u(t)c^* \ ,$$

und als Grenzwert der Picard-Iterierten

$$\Phi(t;u) = \lim_{l\to\infty} Y_l(t) \ ,$$

wobei

$$Y_0 \equiv \mathrm{I} \quad \text{und} \quad Y_{l+1}(t) = \mathrm{I} + \int_0^t U(s)Y_l(s)\mathrm{ds} \quad (l = 0,1,2,\dots) \ .$$

Satz 3.56 (Beschränktheit)**.** Es sei $k \in \mathbb{N}_0$ und $\alpha := \max_{u\in\mathcal{U}} \|A + uc^*\|$. Erfüllt $p \in \mathbb{R}^n$ die Bedingung

$$c^*A^j p = 0 \qquad \forall\, 0 \le j \le k-1 \ , \tag{3.31}$$

so erhält man die Abschätzung

$$\mathcal{A}_t(p) \subseteq e^{At}p + 2\rho_k(\alpha t)\|p\|K \ , \tag{3.32}$$

wobei $\rho_k(s) := \sum_{j=k+1}^{\infty} \frac{s^j}{j!}$ die Restfunktion der Exponentialreihe und K die abgeschlossene Einheitskugel bezeichnet.

Beweis. Wir betrachten die Neumann-Reihe und die Picard-Iterierten in Erinnerung 3.55, und übernehmen die Notation.

1. Wir zeigen, dass wegen (3.31) gilt

$$Y_j(t)p = \sum_{i=0}^{j} \frac{1}{i!} t^i A^i p \qquad \forall\, 0 \le j \le k \ . \tag{3.33}$$

Dies läßt sich per Induktion nach $j = 0, 1, \dots k$ nachweisen:
Induktionsanfang $j = 0$.

$$Y_0(t)p = \mathrm{I}p = p$$

Induktionsschritt $j \to j+1,\ j < k$.

$$\begin{aligned} Y_{j+1}(t)p &= \left(\mathrm{I} + \int_0^t \underbrace{(A + u(s)c^*)}_{U(s)} Y_j(s)\mathrm{ds} \right) \cdot p \\ &\overset{\text{I.V}}{=} p + \int_0^t (A + u(s)c^*) \cdot \left(\sum_{i=0}^{j} \frac{1}{i!} s^i A^i p \right) \mathrm{ds} \\ &= p + \sum_{i=0}^{j} \frac{1}{i!} A^{i+1} p \int_0^t s^i \mathrm{ds} + \sum_{i=0}^{j} \frac{1}{i!} \int_0^t s^i u(s) \underbrace{c^* A^i p}_{=0} \mathrm{ds} \\ &= p + \sum_{i=0}^{j} \frac{1}{i+1!} t^{i+1} A^{i+1} p = \sum_{i=0}^{j+1} \frac{1}{i!} t^i A^i p \end{aligned}$$

2. Insbesondere folgt, dass

$$Y_k(t)p = \sum_{i=0}^{k} \frac{1}{i!} t^i A^i p = e^{At} p - \sum_{i=k+1}^{\infty} \frac{1}{i!} t^i A^i p$$

unabhängig von der Kontrolle u ist. In (B.19) sehen wir, dass die Picard-Iterierte $Y_k(t)$ genau der k-ten Partialsumme der Neumann-Reihe entspricht.

$$\begin{aligned} R_k(t) :=& \Phi(t;u) - Y_k(t) \\ &= \sum_{i=k+1}^{\infty} \int_0^t \int_0^{s_1} \int_0^{s_2} \cdots \int_0^{s_{i-1}} U(s_1)U(s_2)\ldots U(s_i)\mathrm{ds_i}\ldots \mathrm{ds_2}\mathrm{ds_1} \end{aligned}$$

bezeichne den Restterm. Die Elemente der erreichbaren Menge lassen sich nun zerlegen in

$$\underbrace{\Phi(t;u)p}_{\in \mathcal{A}_t(p)} = Y_k(t)p + (\Phi(t;u) - Y_k(t))\, p = e^{At}p - \sum_{i=k+1}^{\infty} \frac{1}{i!} t^i A^i p + R_k(t)p\ . \tag{3.34}$$

3. Da $0 \in \mathcal{U}$ ist $\|A\| \le \alpha$, und es gilt

$$\left\| - \sum_{i=k+1}^{\infty} \frac{1}{i!} t^i A^i p \right\| \le \left(\sum_{i=k+1}^{\infty} \frac{1}{i!} t^i \|A\|^i \right) \|p\| \le \rho_k(\alpha t)\|p\|\ .$$

4. Im Beweis vom Satz B.22 (Teil 2) steht die Abschätzung

$$\|Y_i(t) - Y_{i-1}(t)\| \le \alpha^i \frac{t^i}{i!} \qquad \forall\, i \ge 1\ .$$

Daraus folgern wir

$$\|R_k(t)\| = \|\Phi(t;u) - Y_k(t)\| \overset{(B.19)}{=} \left\| \sum_{i=k+1}^{\infty} (Y_i(t) - Y_{i-1}(t)) \right\| \tag{3.35}$$
$$\leq \sum_{i=k+1}^{\infty} \|Y_i(t) - Y_{i-1}(t)\| \leq \sum_{i=k+1}^{\infty} \alpha^i \frac{t^i}{i!} = \rho_k(\alpha t) \ .$$

Man beachte, dass diese Rechnung für jede zulässige Kontrolle $u(\cdot)$ korrekt ist.

5. Schließlich kombinieren wir (3.34) mit 3. und 4., und sehen, dass für jedes Element $\Phi(t;u)p \in \mathcal{A}_t(p)$

$$\|\Phi(t;u)p - e^{At}p\| \leq 2\rho_k(\alpha t)\|p\|$$

folgt. □

Beispiel 3.57. Wir haben innerhalb dieses Kapitels bereits drei verschiedene Obermengen für die erreichbaren Mengen von (3.30) hergeleitet, die wir jetzt vergleichen wollen. Mit den Bezeichnungen aus Satz 3.56 erhalten wir

(i) $\mathcal{A}_t(p) \subseteq e^{\alpha t} \cdot \max\{1, \|p\|\} \cdot K$ (Folgerung 3.21)

(ii) $\mathcal{A}_t(p) \subseteq e^{\alpha t}\|p\|K$ (Beweis von Satz 3.38)

(iii) $\mathcal{A}_t(p) \subseteq e^{At}p + 2\rho_k(\alpha t)\|p\|K$ (Satz 3.56)

Offenbar ist (ii) eine strikt bessere Abschätzung als (i). Für kleine t ist die letzte Obermenge am kleinsten (denn $\rho_k(\alpha t) \to 0$ für $t \to 0$), und bildet in diesem Sinne die lokal beste Abschätzung.
Dies können wir anhand eines Zahlenbeispiels veranschaulichen. Betrachte dazu das initialisierte Rang-1 Single-Input System

$$\begin{pmatrix} \dot{x}_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + u \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad 0 \leq u \leq 1, \quad x(0) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

mit den Parametern

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad p = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \ . \tag{3.36}$$

(Durch eine Bang-Bang Kontrolle $u : I \to \{0, 1\}$ schaltet man zwischen dem Doppelintegrator $\ddot{x} = 0$ und dem linearen Oszillator $\ddot{x} + x = 0$ hin und her.)

Wir rechnen nach:

$$\|p\| = 1\ , \quad e^{At}p = \left(\begin{smallmatrix}1 & t\\ 0 & 1\end{smallmatrix}\right)\cdot\left(\begin{smallmatrix}-1\\ 0\end{smallmatrix}\right) = \left(\begin{smallmatrix}-1\\ 0\end{smallmatrix}\right)\ ,$$

$$\alpha = \max_{0\le u\le 1}\left\|\left(\begin{smallmatrix}0 & 1\\ -u & 0\end{smallmatrix}\right)\right\| = \max_{0\le u\le 1}\left(\max_{\|x\|=1}\left\|\left(\begin{smallmatrix}x_2\\ -ux_1\end{smallmatrix}\right)\right\|\right) = \max_{0\le u\le 1}\max_{\|x\|=1}\sqrt{x_2^2+u^2x_1^2}$$
$$= 1$$

Aus $c^*p \neq 0$ folgt $k = 0$ und $\rho_k(\alpha t) = \rho_0(t) = e^t - 1$.
Eingesetzt in (i)-(iii), gewinnen wir z.B. für $t = \frac{\pi}{8}$ die Obermengen

(i)/(ii) $\mathcal{A}_{\frac{\pi}{8}}\left(\begin{smallmatrix}-1\\ 0\end{smallmatrix}\right) \subseteq e^{\frac{\pi}{8}}K \approx 1.481K$

(iii) $\mathcal{A}_{\frac{\pi}{8}}\left(\begin{smallmatrix}-1\\ 0\end{smallmatrix}\right) \subseteq \left(\begin{smallmatrix}-1\\ 0\end{smallmatrix}\right) + 2(e^{\frac{\pi}{8}} - 1)K \approx \left(\begin{smallmatrix}-1\\ 0\end{smallmatrix}\right) + 0.962K$

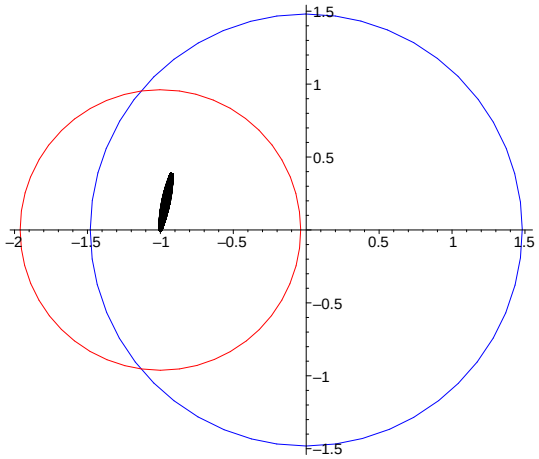

Abbildung 3.3: Obermengen von $\mathcal{A}_t(p)$

In Abbildung 3.3 sind die Obermengen (ii) und (iii) (blauer bzw. roter Kreis) zusammen mit der erreichbaren Menge $\mathcal{A}_{\frac{\pi}{8}}\left(\begin{smallmatrix}-1\\ 0\end{smallmatrix}\right)$ (schwarze Ellipse) eingezeichnet. Wie wir sehen können, liefert (iii) die beste Approximation.

Das nächste Lemma ist von elementarer Bedeutung. Es zeigt, dass sich jede Funktion $t \mapsto c^*x(t)$ lokal wie ein Monom von Grad k verhält, wenn p nicht komplett unkontrollierbar ist.

Lemma 3.58 (Fundamentallemma)**.** Für jeden Anfangswert $p \in \mathbb{R}^n$ gibt es zwei Alternativen

Fall 1 Es gilt

$$c^* A^j p = 0 \qquad \forall j \in \{0, 1, \ldots, n-1\} \ .$$

Dann sind die erreichbaren Mengen

$$\mathcal{A}_t(p) = \{e^{At} p\}$$

einelementig.

Fall 2 Es existiert ein

$$k := \min \left\{ 0 \le j \le n-1 \mid c^* A^j p \neq 0 \right\} \ .$$

Dann gibt es Konstanten $C > 0$ und $\Theta > 0$, so dass für jede zulässige Lösung $x : I \to \mathbb{R}^n$ des Anfangswertproblems

$$\dot{x}(t) = (A + u(t)c^*)x(t), \qquad u(t) \in \mathcal{U}, \quad t \in \mathbb{R}, \quad x(0) = p \quad \text{(AWP)}$$

gilt:

$$\left| c^* x(t) - \frac{c^* A^k p}{k!} t^k \right| \le C|t|^{k+1} \qquad \forall\, t \in I : |t| \le \Theta \tag{3.37}$$

Oder kurz:

$$c^* x(t) = \frac{c^* A^k p}{k!} t^k + O(t^{k+1}) \qquad \text{für } t \to 0.$$

Beweis. Bezeichnungen aus Satz 3.56 inklusive Beweis werden übernommen.

1. Im Fall 1 gilt sogar $c^* A^j p = 0$ für alle $j \ge 0$.
(Denn nach Cayley-Hamilton ist A^j eine Linearkombination der Form $\sum_{i=0}^{n-1} \lambda_i A^i$ mit $\lambda_i \in \mathbb{R}$, und folglich $c^* A^j p = \sum_{i=0}^{n-1} \lambda_i \underbrace{c^* A^i p}_{\overset{\text{Vor}}{=} 0} = 0$.)
Wir können also in der Gleichung (3.32) $k \to \infty$ gehen lassen, und erhalten so die erste Behauptung.

2. Aus $c^* A^j p \neq 0$ für ein minimales k $(0 \le k \le n-1)$ folgern wir

$$c^* Y_k(t) p \overset{(3.33)}{=} \sum_{i=0}^{k} \frac{1}{i!} t^i \underbrace{c^* A^i p}_{\substack{=0 \text{ für} \\ i \le k-1}} = \frac{1}{k!} t^k c^* A^k p \ .$$

Multiplikation in (3.34) von links mit c^* ergibt für eine beliebige zulässige Lösung $x(t) = \Phi(t;u)p$ von (AWP) auf einem Intervall I:

$$c^*x(t) = c^*\Phi(t;u)p = c^*Y_k(t)p + c^*R_k(t)p = \frac{1}{k!}t^k c^*A^k p + c^*R_k(t)p$$

3. Es genügt zu zeigen, dass $c^*R_k(t)p = O(t^{k+1})$ für $t \to 0$.
Wir rechnen nach, dass

$$\begin{aligned} |c^*R_k(t)p| &\le \|c^*\|\cdot\|p\|\cdot\|R_k(t)\| \overset{(3.35)}{\le} \|c^*\|\cdot\|p\|\cdot\rho_k(\alpha t) \\ &\le 2\|c^*\|\cdot\|p\|\cdot\frac{|\alpha t|^{k+1}}{(k+1)!} \qquad \text{für alle } t \text{ mit } |t| \le \tfrac{1}{\alpha} + \tfrac{1}{2\alpha}k. \end{aligned}$$

(Die Abschätzung des Restglieds in der letzten Ungleichung stammt aus [8, p.72].) Daher existiert eine Konstante $C := 2\|c^*\|\cdot\|p\|\cdot\frac{\alpha^{k+1}}{(k+1)!}$, so dass

$$|c^*R_k(t)p| \le C|t|^{k+1} \qquad \forall\, t \in I : |t| \le \tfrac{1}{\alpha} + \tfrac{1}{2\alpha}k\ .$$

□

Definition 3.59. Der lineare Raum

$$\mathcal{N} := \big\{p \in \mathbb{R}^n \mid c^*A^j p = 0 \ \ \forall j \ge 0\big\} = \big\{p \in \mathbb{R}^n \mid c^*A^j p = 0 \,\forall\, 0 \le j \le n-1\big\}$$

bezeichne die *Menge aller komplett unkontrollierbaren Zustände* von (3.30). Er ist A-invariant, d.h. $A\cdot\mathcal{N} \subseteq \mathcal{N}$. Ist $\mathcal{N} = \{0\}$, so heißt das System (3.30) *kontrollierbar*.
Wie wir im Fundamentallemma gesehen haben, sind die erreichbaren Mengen von komplett unkontrollierbaren Zuständen stets einelementig und unabhängig von $\mathcal{U}$.

Um $\mathcal{N}$ näher charakterisieren zu können, führen wir eine Notation ein, die in der linearen Kontrolltheorie gebräuchlich ist.

Definition und Hilfssatz 3.60. Für eine Teilmenge $U \subseteq \mathbb{R}^n$ bezeichne $\langle A;U\rangle$ den kleinsten A-invarianten Unterraum von $\mathbb{R}^n$, der U enthält.
Falls $0 \in U$ und L die lineare Hülle von U ist, so gilt

$$\langle A;U\rangle = \langle A;L\rangle = L + AL + A^2L + \ldots + A^{n-1}L\ .$$

Ist U das Bild einer Matrix B ist, so folgt

$$\langle A;U\rangle = \langle A;\mathrm{Im}B\rangle = \mathrm{Im}\big(B|AB|\ldots|A^{n-1}B\big)\ .$$

Bemerkung 3.61. Der Vektorraum $\mathcal{N}$ bildet ebenso die sog. „Menge aller unbeobachtbaren Zustände“ des linearen Kontrollsystems mit Ausgang

$$\begin{aligned} \dot{x}(t) &= Ax(t) + u(t), \qquad u(t) \in \mathcal{U} \ , \\ y(t) &= c^* x(t) \ . \end{aligned} \tag{3.38}$$

Dies rechtfertigt die Bezeichnung von c als *Beobachtungsvektor*.
Wir nennen (3.38) das „assoziierte lineare System“ von unserem bilinearen System (3.30). Insbesondere ist das Rang-1 System (3.30) genau dann kontrollierbar, wenn das assoziierte lineare System (3.38) beobachtbar ist.
Es ist wohlbekannt [9, p.129], dass $\mathcal{N} = \langle A^*; \mathrm{Im} c\rangle^\perp$. Ist z.B. $\mathcal{U} = B \cdot E^m$ für eine $n \times m$ Matrix B, so steht $\mathcal{N}$ senkrecht auf der Erreichbarkeitsmenge des dualen linearen Systems von (3.38)

$$\begin{aligned} \dot{x}(t) &= A^* x(t) + cv(t), \qquad v(t) \in E^m \ , \\ y(t) &= B^* x(t) \ . \end{aligned}$$

Folgerung 3.62 (1. Starke Invarianz)**.** $\mathcal{N}$ ist eine beidseitig invariante Menge. Das bedeutet, für jede zulässige Lösung $x : I \to \mathbb{R}^n$ von (AWP) gilt

entweder $x(t) \notin \mathcal{N}$ für alle $t \in I$, und $c^* x(\cdot)$ hat nur isolierte Nullstellen;

oder $x(t) \in \mathcal{N}$ für alle $t \in I$, und $c^* x(\cdot) \equiv 0$.

Beweis. 1. Angenommen, es ist $p \in \mathcal{N}$. Dann ist p auch komplett unkontrollierbar für das zeitumgekehrte System

$$\dot{x}(t) = -(A + u(t)c^*)x(t)$$

(denn $c^*(-A)^j p = 0 \quad \forall j \geq 0$). Wegen des Fundamentallemmas sind daher sowohl die erreichbaren Mengen $\mathcal{A}_t(p)$ des Ausgangssystems als auch die erreichbaren Mengen $\mathcal{C}_t(p)$ des zeitumgekehrten Systems einelementig mit $\mathcal{A}_t(p) = \{e^{At}p\}$ bzw. $\mathcal{C}_t(p) = \{e^{-At}p\}$. Folglich ist die maximale Lösung $x : \mathbb{R} \to \mathbb{R}^n$ von (AWP), unabhängig von der gewählten Kontrollfunktion, gegeben durch $x(t) = e^{At}p$. Insbesondere sind die Werte von $x(\cdot)$ stets in $\mathcal{N}$:

$$c^* A^j x(t) = c^* A^j e^{At} p = c^* e^{At} \underbrace{A^j p}_{:=\tilde{p}\in\mathcal{N}} = \sum_{i=0}^{\infty} \frac{t^i}{i!} \underbrace{c^* A^i \tilde{p}}_{=0} = 0 \ \ \forall j \geq 0, t \in \mathbb{R} \tag{3.39}$$

Es ist nun klar, dass $\mathcal{N}$ eine beidseitig invariante Menge ist.

2. Aus Lemma 3.13 wissen wir, dass auch das Komplement $\mathbb{R}^n \backslash \mathcal{N}$ beidseitig invariant ist, woraus die Entweder-oder-Unterscheidung folgt. Für $j = 0$ in

(3.39) erhalten wir den zweiten Fall.

3. Es sei $p \notin \mathcal{N}$ und $x(\cdot)$ eine zulässige Lösung von (AWP). Angenommen, es gibt eine Nullfolge $t_r \to 0$, so dass $c^*x(t_r) = 0$ für alle r genügend groß. Dann ergibt (3.37)

$$\left|\frac{c^*A^kp}{k!C}\right| \leq |t_r| \qquad \text{(hier ist } k \text{ die Zahl aus Lemma 3.58)}$$

für alle r genügend groß und eine Konstante $C > 0$. Dies ist ein Widerspruch zu $\frac{c^*A^kp}{k!C} \neq 0$. Unter Berücksichtigung der Autonomie des Systems folgern wir, dass $c^*x(\cdot)$ nur isolierte Nullstellen haben kann. □

Folgerung 3.63. Es sei $p \notin \mathcal{N}$, k die Zahl aus Lemma 3.58 und $x : I \to \mathbb{R}^n$ eine zulässige Lösung von (AWP). Dann ist das Vorzeichen von $c^*x(t)$ gleich dem Vorzeichen von c^*A^kp für alle positiven $t \in I$ nahe 0. Ist der Anfangswert p keine (isolierte) Nullstelle von $c^*x(\cdot\,)$, so ist die Aussage auch zum Zeitpunkt $t = 0$ gültig.

Beweis. Es sei ohne Einschränkung $\operatorname{sgn}(c^*A^kp) = +1$. Angenommen, es ist $\operatorname{sgn}(c^*x(t)) = -1$ für alle positiven t nahe 0. Laut Fundamentallemma existiert dann ein $C > 0$ und ein $\Theta > 0$, so dass

$$\frac{c^*A^kp}{k!}t^k \leq \frac{c^*A^kp}{k!}t^k - c^*x(t) = \left|c^*x(t) - \frac{c^*A^kp}{k!}t^k\right| \leq Ct^{k+1} \qquad \forall 0 \leq t \leq \Theta\ .$$

Dies führt für

$$0 < t < \underbrace{\frac{c^*A^kp}{k!C}}_{>0}$$

zum Widerspruch. □

Bemerkung 3.64. Wenn wir berücksichtigen, dass die Konstanten C und Θ aus dem letzten Beweis unabhängig von der gewählten Lösung $x(\cdot)$ sind, so erhalten wir bei näherer Betrachtung eine stärkere Version von Folgerung 3.63:
Falls $p \notin \mathcal{N}$, so gilt für jede zulässige Lösung $x : I \to \mathbb{R}^n$ von (AWP):

$$\operatorname{sgn}(c^*x(t)) = \operatorname{sgn}(c^*A^kp) \neq 0 \quad \text{für alle } t \in I \text{ mit } 0 < t < \min\left\{\Theta, \tfrac{|c^*A^kp|}{k!C}\right\}$$

Satz 3.65 (2. Starke Invarianz)**.** Es gilt

$$\mathcal{A}_t(p) \subseteq e^{At}p + \langle A; \mathcal{U}\rangle\ . \tag{3.40}$$

Insbesondere ist die Erreichbarkeitsmenge $\langle A; \mathcal{U}\rangle$ des assoziierten linearen Systems (3.38) eine beidseitig invariante Menge bezüglich des bilinearen Systems (3.30).

Beweis. Es bezeichne $\mathcal{L}$ die lineare Hülle von $\mathcal{U} \ni 0$. Dann ist

$$\langle A;\mathcal{U}\rangle = \langle A;\mathcal{L}\rangle = \mathcal{L} + A\mathcal{L} + \ldots + A^{n-1}\mathcal{L} \ . \tag{3.41}$$

Sei nun $x : [0,\infty) \to \mathbb{R}^n$ die Lösung von (AWP) zu einer zulässigen Kontrolle $u \in \mathcal{U}_r^{[0,\infty)}$. Dann liefert Variation der Konstanten (setze $b(t) := u(t)c^*x(t)$ in (B.27) ein)

$$x(t) = e^{At}p + \underbrace{\int_0^t \underbrace{e^{A(t-s)}u(s)}_{\in\langle A;\mathcal{U}\rangle}\underbrace{c^*x(s)}_{\in\mathbb{R}}\,\mathrm{ds}}_{\in\langle A;\mathcal{U}\rangle} \ . \tag{3.42}$$

Da $u(\cdot\,)$ beliebig wählbar, ist (3.40) gezeigt.

Aus $p \in \langle A;\mathcal{U}\rangle$ folgt $e^{At}p \in \langle A;\mathcal{U}\rangle$, und somit $x(t) \in \langle A;\mathcal{U}\rangle$ für alle $t \geq 0$. Diese Aussage können wir leicht verallgemeinern (Stichworte: Autonomie, zeitumgekehrtes System):
Ist $x(\cdot\,)$ eine zulässige Lösung von (AWP) auf einem Intervall I und $x(s) \in \langle A;\mathcal{U}\rangle$ zu einem Zeitpunkt $s \in I$, so folgt $x(t) \in \langle A;\mathcal{U}\rangle$ für alle Zeiten $t \in I$. □

Satz 3.66 (Affine Hülle)**.** Sei $t > 0$. Ist $p \notin \mathcal{N}$, so ist $e^{At}p + \langle A;\mathcal{U}\rangle$ die affine Hülle von $\mathcal{A}_t(p)$; sonst ist sie gleich $\{e^{At}p\}$.

Beweis. Es bezeichne $\tilde{\mathcal{L}}$ die lineare Hülle von $\mathcal{A}_t(p) - e^{At}p$.
Aus $p \in \mathcal{N}$ folgt $\mathcal{A}_t(p) = \{e^{At}p\}$, und die zweite Behauptung ist bewiesen. Im Falle $p \notin \mathcal{N}$ wird behauptet, dass $\langle A;\mathcal{U}\rangle$ die lineare Hülle von $\mathcal{A}_t(p) - e^{At}p$ ist. Nach Satz 3.65 gilt

$$\mathcal{A}_t(p) - e^{At}p \subseteq \langle A;\mathcal{U}\rangle \ .$$

Es bleibt daher zu zeigen, dass $\langle A;\mathcal{U}\rangle$ in der linearen Hülle $\tilde{\mathcal{L}}$ enthalten ist, oder gleichbedeutend

$$\tilde{\mathcal{L}}^{\perp} \subseteq \langle A;\mathcal{U}\rangle^{\perp} \ .$$

Da $\tilde{\mathcal{L}}$ eine Basis aus der Menge $\{\mathcal{A}_t(p) - e^{At}p\}$ besitzt, genügt es zu beweisen, dass jeder Vektor $q \in \mathbb{R}^n$, der orthogonal zu $\{\mathcal{A}_t(p) - e^{At}p\}$ ist, auch orthogonal zu $\langle A;\mathcal{U}\rangle$ ist. Es gelte also

$$q^*(x(t) - e^{At}p) = 0 \tag{3.43}$$

für jede zulässige Lösung $x(\cdot)$ mit Anfangswert $x(0) = p$.
Genauer, $x(\cdot)$ sei die Lösung zu einer stückweise konstanten Kontrolle $u : [0,t] \to \mathcal{U}$ mit festem Wert $v \in \mathcal{U}$ auf einem Teilintervall $[r, r+h]$ von $[0,t]$ und Wert 0 sonst. Mittels der Regel (2.17) berechnen wir:

$$x(s) = \Phi(s;u)p = \Phi(s-r;u^{-r})\cdot\Phi(r;u)p = e^{(A+vc^*)(s-r)}e^{Ar}p \;\; \forall s \in [r, r+h]$$

Aus (3.42) und (3.43) folgt

$$0 = q^*(x(t) - e^{At}p) = \int_r^{r+h} q^* e^{A(t-s)} v c^* e^{(A+vc^*)(s-r)} e^{Ar} p \,\mathrm{ds} \;.$$

Wir teilen durch $h > 0$ und lassen $h \to 0$ gehen. Dies ergibt auf der rechten Seite die Ableitung der Stammfunktion des Integranden an der Stelle r. Folglich ist

$$0 \overset{\text{HDI}}{=} q^* e^{A(t-r)} v \cdot c^* e^{Ar} p \qquad \forall r \in [0,t), v \in \mathcal{U} \;.$$

Da $p \notin \mathcal{N}$, hat $r \mapsto c^* e^{Ar} p$ nur isolierte Nullen (Folgerung 3.62). Aus Stetigkeitsgründen muß daher die analytische Funktion $r \mapsto q^* e^{A(t-r)} v$ identisch Null sein auf ganz $[0,t]$. Nach dem Identitätssatz für Potenzreihen sind die Koeffizienten $q^* A^k v$ $(k = 0, 1, \ldots)$ der zugehörigen Potenzreihenentwicklung gleich Null für alle $v \in \mathcal{U}$. Das impliziert $q^* A^k \mathcal{L} = \{0\}$ für $k = 0, 1, \ldots n-1$, und wegen (3.41) erhalten wir $q \in \langle A; \mathcal{U} \rangle^\perp$, was zu zeigen war. □

Spaltensysteme mit konvexem Steuerbereich

Wir untersuchen weiterhin das initialisierte System

$$\dot{x}(t) = (A + u(t)c^*)x(t), \qquad u(t) \in \mathcal{U}, \quad t \in \mathbb{R}, \quad x(0) = p \;. \qquad \text{(AWP)}$$

Fordern jedoch zusätzlich, dass der Steuerbereich konvex ist.

Voraussetzung 3.67. $\mathcal{U}$ ist kompakt, konvex und enthält 0.

Dies hat zur Folge, dass die erreichbaren Mengen kompakt und wegzusammenhängend sind (Folgerung 3.41), woraus wiederum die Existenz zeitoptimaler Steuerungen folgt.

Hilfssatz 3.68. Es sei $p \notin \mathcal{N}$. Falls es ein $t > 0$ gibt mit $0 \in c^* \mathcal{A}_t(p)$, so existiert ein erster positiver Zeitpunkt

$$\delta = \delta(p) := \min\{t > 0 \mid 0 \in c^* \mathcal{A}_t(p)\} \;,$$

zu dem die Hyperebene $\{x \in \mathbb{R}^n \mid c^* x = 0\}$ von p aus erreicht.

Beweis. Wir setzen zunächst

$$\delta = \inf\{t > 0 \mid 0 \in c^* \mathcal{A}_t(p)\}$$

und zeigen $0 \in c^* \mathcal{A}_\delta(p)$. Danach führen wir die Annahme „$\delta = 0$“ zum Widersprich, um den Beweis abzuschließen.
Nach Definition des Infimum gibt es zu jedem $k \in \mathbb{N}$ einen Zeitpunkt $t_k \geq \delta$ zusammen mit einer Lösung $x_k : [0, t_k] \to \mathbb{R}^n$ von (AWP) bezüglich einer zulässiger Kontrolle u_k, so dass $t_k > 0$, $t_k \to \delta$ (für $k \to \infty$) und $c^* x_k(t_k) = 0$. Nach Voraussetzung gilt $\delta < T < \infty$ für ein $T > 0$. Wegen Folgerung 3.41 ist die Menge

$$\mathcal{A}_{\leq T}(p) := \bigcup_{0 \leq t \leq T} \mathcal{A}_t(p) = \mathcal{A}_{\leq T} \cdot p$$

kompakt. Da die rechte Seite $f(x, u) := Ax + uc^* x$ stetig auf der kompakten Menge $\mathcal{A}_{\leq T}(p) \times \mathcal{U}$ ist, existiert eine obere Schranke $M \geq 0$, so dass

$$\|f(x, u)\| \leq M \qquad \forall\, x \in \mathcal{A}_{\leq T}(p), u \in \mathcal{U}\ .$$

Weiter gilt für k genügend groß

$$|c^* x_k(\delta) - \underbrace{c^* x_k(t_k)}_{=0}| \leq \|c^*\| \cdot \|x_k(\delta) - x_k(t_k)\| \leq \|c^*\| \cdot \int_\delta^{t_k} \underbrace{\|f(x_k(s), u_k(s))\|}_{\leq M} \mathrm{ds}$$
$$\leq \|c^*\| M (t_k - \delta)\ .$$

Es konvergiert somit $c^* x_k(\delta)$ gegen 0 für $k \to \infty$. Mit $\mathcal{A}_\delta(p)$ ist auch $c^* \mathcal{A}_\delta(p) \ni c^* x_k(\delta)$ abgeschlossen, woraus die Behauptung $0 \in c^* \mathcal{A}_\delta(p)$ folgt.
Laut Bemerkung 3.64 gibt es ein Intervall $(0, \beta)$ mit $\beta > 0$, so dass $\operatorname{sgn}(c^* x(t)) \neq 0$ und somit $c^* x(t) \neq 0$ für alle $t \in (0, \beta)$ und für jede zulässige Lösung $x(\cdot)$ von (AWP) auf $(0, \beta)$. Es kann also nicht $\delta = 0$ sein. □

Bezeichnung 3.69. Gemäß Hilfssatz 3.68 ist

$$\delta = \delta(p) := \begin{cases} \min\{t > 0 \mid 0 \in c^* \mathcal{A}_t(p)\}\ , & \text{falls } p \notin \mathcal{N} \text{ und } 0 \in c^* \mathcal{A}_t(p) \\ & \text{für ein } t > 0 \\ \infty\ , & \text{sonst} \end{cases}$$

eine wohldefinierte Größe mit $0 < \delta \leq \infty$.

Für einen Anfangswert $p \in \mathbb{R}^n$ und einem Endzeitpunkt $T \in \mathbb{R}_+$ führen wir die Lösungsmengen

$$\mathrm{H}(p, T) := \{x : [0, T] \to \mathbb{R}^n \mid x(\cdot) \text{ ist eine zulässige Lösung von (AWP)}\}$$

und

$$\mathrm{H}(p,\infty) := \{x : [0,\infty) \to \mathbb{R}^n \mid x(\cdot) \text{ ist eine zulässige Lösung von (AWP)}\}$$

ein. Es sei bemerkt, dass wegen Satz 3.38 die Menge $\mathrm{H}(p,T)$ eine nichtleere, kompakte und wegzusammenhängende Teilmenge von $C^0([0,T],\mathbb{R}^n)$ ist. Wir werden im nächsten Beweis sehen, dass $\mathrm{H}(p,\delta(p))$ noch eine weitere Eigenschaft besitzt: Sie ist konvex.

Satz 3.70 (Small-Time Konvexität)**.** Die erreichbaren Mengen $\mathcal{A}_t(p)$ sind konvex für alle t mit $0 < t \le \delta(p)$.

Beweis. 1. Ist $p \in \mathcal{N}$, so besteht $\mathcal{A}_t(p)$ für $t \ge 0$ nur aus dem Punkt $e^{At}p$ (siehe Lemma 3.58) und ist trivialerweise konvex.

2. Es sei ab jetzt $p \notin \mathcal{N}$. Für eine beliebige Lösung $x \in \mathrm{H}(p,\delta(p))$ hat die Ausgangsfunktion $c^*x(\cdot)$ nur isolierte Nullstellen (Folgerung 3.62). Die Werte der stetigen Funktion $c^*x(\cdot)$ haben auf $(0,\delta)$ stets das gleiche Vorzeichen—und zwar das Vorzeichen von c^*A^kp (Folgerung 3.63), wobei

$$k = \min\{0 \le j \le n-1 \mid c^*A^jp \ne 0\}\ . \tag{3.44}$$

Anders gesagt,

$$\operatorname{sgn}(c^*A^kp) = \operatorname{sgn}(c^*x(t)) \qquad \forall\, t \in (0,\delta),\ x \in \mathrm{H}(p,\delta(p))\ .$$

3. Es wird sogar gezeigt, dass die Lösungsmenge $\mathrm{H}(p,\delta(p))$ konvex ist, d.h.

$$x,y \in \mathrm{H}(p,\delta(p)), \lambda \in [0,1] \Longrightarrow z := \lambda x + (1-\lambda)y \in \mathrm{H}(p,\delta(p))\ .$$

Es seien $u,v : [0,\delta] \to \mathbb{R}^n$ die zu x,y gehörigen Kontrollen. Dann hat die absolut stetige Funktion z fast überall die Ableitung

$$\begin{aligned}\dot z &= \lambda\dot x + (1-\lambda)\dot y = \lambda(Ax + uc^*x) + (1-\lambda)(Ay + vc^*y)\\ &= Az + \underbrace{\big(\lambda uc^*x + \big(1-\lambda\big)vc^*y\big)}_{\stackrel{!}{=}wc^*z = w\left(\lambda c^*x + (1-\lambda)c^*y\right)}\ .\end{aligned}$$

Für die (fast überall definierte) Kontrolle

$$w := \frac{\lambda uc^*x + \big(1-\lambda\big)vc^*y}{\lambda c^*x + (1-\lambda)c^*y} = \frac{\lambda c^*x}{\lambda c^*x + (1-\lambda)c^*y}u + \frac{(1-\lambda)c^*y}{\lambda c^*x + (1-\lambda)c^*y}v \tag{3.45}$$

gilt daher $\dot{z}(t) = Az(t) + w(t)c^*z(t)$ fast überall.
Es fehlt nur noch die Zulässigkeit von w. Nach 2. haben $c^*x(t)$ und $c^*y(t)$ das gleiche Vorzeichen auf $(0, \delta)$. Also sind für $t \in (0, \delta)$ die Brüche in (3.45) positiv und addieren sich zu 1 auf. Da $\mathcal{U}$ konvex, haben wir $w(t) \in \mathcal{U}$ für alle t mit $0 \le t \le \delta$ höchstens bis auf die Randpunkte $t = 0$ und $t = \delta$ (falls $\delta < \infty$). An diesen Randpunkten setzen wir ohne Einschränkung $w(t) = 0$. Offenbar ist dann w eine messbare Funktion, definiert auf $[0, \delta]$ bzw. $[0, \infty)$ (falls $\delta = \infty$), mit Werten in $\mathcal{U}$, weshalb die zugehörige Lösung $z(\cdot)$ zulässig ist. Folglich gehören dessen Werte $z(t)$ zu $\mathcal{A}_t(p)$. □

Eine kurze Analyse des Beweises ermöglicht es, den Konvexitätsbereich weiter auszudehnen.

Definition und Folgerung 3.71. In Anlehnung an Folgerung 3.63 definieren wir den *Typ* $\sigma(p)$ von $p \in \mathbb{R}^n$ als

$$\sigma(p) := \lim_{t \overset{\ge}{\to} 0} \operatorname{sgn}(c^*x(t))$$

für eine (beliebige) zulässige Lösung $x : [0, T] \to \mathbb{R}^n$ von (AWP) mit Anfangswert p und Endzeit $T > 0$.
Ist $p \notin \mathcal{N}$, so gilt $\operatorname{sgn}(p) = \operatorname{sgn}(c^*A^kp) \neq 0$ für die Zahl k aus (3.44). Sonst ist $c^*x(\cdot) \equiv 0$, d.h. $\sigma(p) = 0$ (Folgerung 3.62).

Außerdem setzen wir

$$\omega(p) := \sup\{T > 0 \mid \operatorname{sgn}(c^*x(t)) \in \{0, \sigma(p)\} \ \forall\, t \in [0, T], x \in \mathrm{H}(p, T)\}.$$

Es ist stets $0 < \delta(p) \le \omega(p) \le \infty$ (z.B. $\omega(p) = \infty$, falls $p \in \mathcal{N}$).
Wegen Lemma 3.58 hat $c^*x(\cdot)$ nur isolierte Nullstellen, falls $p \notin \mathcal{N}$ und $x \in \mathrm{H}(p, \omega(p))$. Es folgt in diesem Fall, dass $c^*x(t) = \operatorname{sgn}(c^*A^kp)$ für alle t mit $0 \le t \le \omega(p)$ bis auf abzählbar viele Ausnahmen. Wie im dritten Beweisteil von Satz 3.70 können wir dann zu je zwei zulässigen Lösungen $x, y \in \mathrm{H}(p, \omega(p))$ und zu einem $\lambda \in [0, 1]$ eine zulässige Kontrolle $w(\cdot)$ bestimmen, definiert auf $[0, \omega(p)]$ bzw. $[0, \infty)$ (falls $\omega(p) = \infty$), die bis auf abzählbar viele Nullstellen von c^*x und c^*y durch (3.45) gegeben ist und sonst gleich Null ist. Auf diese Weise erhalten wir die Konvexität von $\mathcal{A}_t(p)$ auf dem Zeitbereich $[0, \omega(p)]$ bzw. $[0, \infty)$. $\omega(p)$ nennt man daher die *Konvexitätsausdehnung* von p.

Beispiel 3.72 (Lineare Systeme)**.** Wir führen eine Konvexitätsanalyse für das initialisierte lineare Kontrollsystem in $\mathbb{R}^n$

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad u(t) \in \mathcal{U}, \qquad x(0) = p \qquad (A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m})$$

durch. Wie in Beispiel 2.34 gezeigt wird, ist dieses System äquivalent zu einem Rang-1 System im $\mathbb{R}^{n+1}$

$$\begin{pmatrix}\dot{x}(t)\\ \dot{\varphi}(t)\end{pmatrix} = \left(\begin{pmatrix}A & 0\\ 0^* & 0\end{pmatrix} + \begin{pmatrix}Bu(t)\\ 0\end{pmatrix}\begin{pmatrix}0^* & 1\end{pmatrix}\right)\begin{pmatrix}x(t)\\ \varphi(t)\end{pmatrix}, u(t)\in\mathcal{U}, \begin{pmatrix}x(0)\\ \varphi(0)\end{pmatrix} = \begin{pmatrix}p\\ 1\end{pmatrix}$$

mit erweitertem Zustandsraum $\{(x,\varphi)\in\mathbb{R}^n\times\mathbb{R}\}$ und Beobachtungsvektor $c = \mathrm{e}_{n+1}$. Der Steuerbereich $\mathcal{U}$ sei eine Teilmenge des $\mathbb{R}^m$ und erfülle die Voraussetzung (3.67).
Offenbar gilt für jede zulässige Lösung $t\mapsto \binom{x(t)}{1}$ des letztgenannten Systems

$$c^*\begin{pmatrix}x(t)\\ 1\end{pmatrix} = \begin{pmatrix}0^* & 1\end{pmatrix}\begin{pmatrix}x(t)\\ 1\end{pmatrix} \equiv 1\ .$$

Dies impliziert

$$\delta\left(\binom{p}{1}\right) = \omega\left(\binom{p}{1}\right) = \infty \quad \text{und} \quad \sigma\left(\binom{p}{1}\right) = +1 \quad \forall p\in\mathbb{R}^n\ .$$

Die erreichbaren Mengen $\mathcal{A}_t\binom{p}{1}$ des Spalten-Systems sind daher konvex für alle $t\geq 0$.
Da $\mathcal{A}_t\binom{p}{1}$—bis auf Einbettung in die Hyperebene $\varphi = 1$ des $\mathbb{R}^{n+1}$—gleich der erreichbaren Menge „$\mathcal{A}_t^{\mathrm{lin}}(p)$“ des linearen Ausgangssystems ist, dürfen wir folgern, dass auch die erreichbaren Mengen des linearen Kontrollsystems stets konvex sind.

Beispiel 3.73. Wir betrachten wieder das Rang-1 Single-Input System (3.36)

$$\begin{pmatrix}\dot{x}_1\\ \dot{x}_2\end{pmatrix} = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} + u\begin{pmatrix}0\\ -1\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix},\ \ 0\leq u\leq 1,\ \ x(0) = \begin{pmatrix}-1\\ 0\end{pmatrix}.$$

Es sollen die erreichbaren Mengen $\mathcal{A}_t(p)$ von $p := \binom{-1}{0}$ auf Konvexität untersucht werden.
Zur Kontrollen $u\equiv 0$ und $u\equiv 1$ gehören die Lösungen $\binom{x_1(t)}{x_2(t)} = \binom{-1}{0}$ bzw. $\binom{x_1(t)}{x_2(t)} = \binom{-\cos t}{\sin t}$. Auswerten der beiden Lösungen zur Zeit $t=\pi$ liefert $(\pm 1, 0)\in\mathcal{A}_\pi(p)$. Wegen der starken Invarianz von $\mathcal{N}$ (Folgerung 3.62) gehört mit p auch jeder weitere erreichbare Punkt von p nicht zur Menge $\mathcal{N}$. Der Mittelpunkt $0 = \frac{1}{2}\binom{1}{0} + \frac{1}{2}\binom{-1}{0}\in\mathcal{N}$ ist daher zu keiner Zeit von p aus erreichbar, und folglich ist $\mathcal{A}_\pi(p)$ nicht konvex. (Die Abbildung 3.4 stellt eine rein empirische Approximation dieser erreichbaren Menge dar, für die etwa 50000 erreichbare Zustände berechnet wurden, die von stückweise konstanten Bang-Bang Funktionen mit bis zu vier zufällig bestimmten Sprungstellen angesteuert wurden.) Weil p kritisch ist mit monotonen erreichbaren Mengen

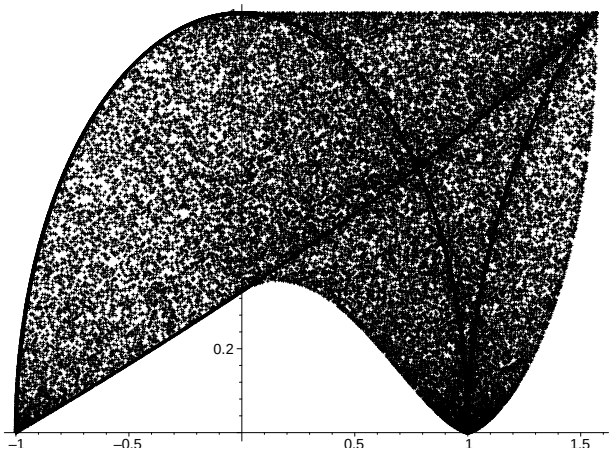

Abbildung 3.4: $\mathcal{A}_\pi(p)$ ist nicht konvex

(denn „$f(p,0) = 0$“), sind aufgrund der gleichen Argumentation auch die „späteren“ erreichbaren Mengen $\mathcal{A}_t(p), t \geq \pi$, nicht konvex.
Als Nächstes sollen die Größen $\delta(p), \sigma(p)$ und $\omega(p)$ berechnet werden.
Es ist $\sigma(p) = \operatorname{sgn}(c^*p) = -1$. Aus der Lösung $\binom{x_1(t)}{x_2(t)} = \binom{-\cos t}{\sin t}$ können wir ablesen, dass $\omega(p) \leq \frac{\pi}{2}$ gelten muß, da $c^*\binom{x_1(t)}{x_2(t)} = -\cos t > 0$ für $\pi > t > \frac{\pi}{2}$. Die Größe $\delta(p)$ ist der minimale Zeitpunkt, zu dem die Hyperebene $\{c^*\binom{x_1}{x_2} = x_1 = 0\}$ von p aus erreicht wird. Um die zugehörige zeitoptimale Kontrolle berechnen zu können, führen wir eine Koordinatentransformation im Zustandsraum $\mathbb{R}^n$ durch. Der Diffeomorphismus

$$\phi : \mathbb{R}^2 \backslash \{(x,0) \mid x \geq 0\} \longrightarrow \mathbb{R}_+ \times [-\pi, \pi] \backslash 0$$

führt die kartesischen Koordinaten (x,y) in die Polarkoordinaten (r,ρ) über:

$$\phi : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \sqrt{x^2+y^2} \\ \operatorname{sgn}(y) \arccos \frac{x}{x^2+y^2} \end{pmatrix} := \begin{pmatrix} r \\ \rho \end{pmatrix}, \qquad \phi^{-1} : \begin{pmatrix} r \\ \rho \end{pmatrix} \mapsto \begin{pmatrix} r\cos\rho \\ r\sin\rho \end{pmatrix} .$$

Die Funktionalmatrix von ϕ ist die Inverse der Funktionalmatrix der Umkehrabbildung:

$$\mathrm{D}\phi(x,y) = \left(\mathrm{D}\phi^{-1}(r,\rho)\right)^{-1} = \begin{pmatrix} \cos\rho & -r\sin\rho \\ \sin\rho & r\cos\rho \end{pmatrix}^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\rho & r\sin\rho \\ -\sin\rho & \cos\rho \end{pmatrix} .$$

Ist $\binom{x_1(t)}{x_2(t)}$ eine zulässige Lösung von (3.36) zur Kontrolle u, so löst die „transformierte Kurve“

$$\begin{pmatrix} r(t) \\ \rho(t) \end{pmatrix} := \phi \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$$

(fast überall) die Differentialgleichung

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}}\begin{pmatrix} r(t) \\ \rho(t)\end{pmatrix} &= \mathrm{D}\phi(x_1(t), x_2(t))\frac{\mathrm{d}}{\mathrm{dt}}\begin{pmatrix} x_1(t) \\ x_2(t)\end{pmatrix} \\ &= \begin{pmatrix} \cos\rho(t) & \sin\rho(t) \\ -\frac{1}{r(t)}\sin\rho(t) & \frac{1}{r(t)}\cos\rho(t)\end{pmatrix}\cdot\begin{pmatrix} x_2(t) \\ -x_1(t)u(t)\end{pmatrix} \\ &= \begin{pmatrix} \cos\rho(t) & \sin\rho(t) \\ -\frac{1}{r(t)}\sin\rho(t) & \frac{1}{r(t)}\cos\rho(t)\end{pmatrix}\cdot\begin{pmatrix} r(t)\sin\rho(t) \\ -r(t)\cos\rho(t)u(t)\end{pmatrix} \\ &= \begin{pmatrix} r(t)\sin\rho(t)\cos\rho(t)(1-u(t)) \\ -1+\cos^2\rho(t)-\cos^2\rho(t)u(t)\end{pmatrix}\end{aligned}$$

mit Anfangswert

$$\begin{pmatrix} r(0) \\ \rho(0)\end{pmatrix} = \phi\begin{pmatrix} -1 \\ 0\end{pmatrix} = \begin{pmatrix} 1 \\ \pi\end{pmatrix} .$$

Insbesondere ist das dynamische Verhalten des Winkels ρ unabhängig von r. Nämlich

$$\rho(t) = \rho(0) + \int_0^t \dot\rho(s)\mathrm{ds} = \pi - t + \int_0^t \underbrace{(1-u(s))\cos^2\rho(s)}_{\in[0,1]}\,\mathrm{ds} . \tag{3.46}$$

Ausgehend von $t = 0$ wollen wir in minimaler Zeit δ die y-Achse schneiden. Es muß also $\rho(\delta) = \pm\frac{\pi}{2}$ gelten. Wegen $\rho(t) \geq \pi - t$ ist $\rho(t) > \frac{\pi}{2}$ für alle $t \in [0, \frac{\pi}{2})$. Daher ist $u \equiv 1$ die optimale Steuerung und $\delta(p) = \omega(p) = \frac{\pi}{2}$. Folglich sind die erreichbaren Mengen $\mathcal{A}_t(p)$ auf dem Intervall $[0, \frac{\pi}{2}]$ konvex. In Abbildung 3.5 ist z.B. die konvexe Menge $\mathcal{A}_{\frac{\pi}{2}}(p)$ dargestellt.

Besitzt die erreichbare Menge $\mathcal{A}_t(p)$ des bilinearen Systems (3.30) ein nichtleeres Inneres, so ist $\langle A; \mathcal{U}\rangle = \mathbb{R}^n$, und das assoziierte lineare System (3.38) ist vollständig kontrollierbar (Satz 3.66). Die vollständige Kontrollierbarkeit dieses linearen Systems ist also eine notwendige Bedingung für $\mathrm{int}\,(\mathcal{A}_t(p)) \neq \emptyset$. Falls der Anfangszustand p nicht komplett unkontrollierbar ist, so ist diese Bedingung sogar hinreichend, wie die anschließende Folgerung zeigt.

Folgerung 3.74. Ist $p \notin \mathcal{N}$ und $\langle A; \mathcal{U}\rangle = \mathbb{R}^n$, so haben die erreichbaren Mengen $\mathcal{A}_t(p)$ zu jedem Zeitpunkt $t > 0$ ein nichtleeres Inneres.

Beweis. 1. Gemäß Satz 3.70 gibt es ein $\delta > 0$, so dass $\mathcal{A}_t(p)$ konvex ist für $t \in [0, \delta]$ und daher ein nichtleeres Inneres relativ zur affinen Hülle besitzt. Nach Satz 3.66 ist die affine Hülle gleich $e^{At}p + \langle A; U\rangle = \mathbb{R}^n$, woraus die Behauptung für alle $t \leq \delta$ folgt.

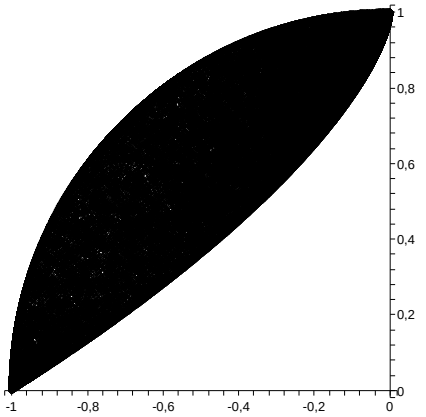

Abbildung 3.5: $\mathcal{A}_{\frac{\pi}{2}}(p)$ ist konvex

2. Für $t \geq \delta$ gilt aufgrund des Additionstheorems aus Folgerung 3.9

$$\mathcal{A}_t(p) = \mathcal{A}_t p = \mathcal{A}_{t-\delta} \cdot \mathcal{A}_\delta(p) \supseteq e^{A(t-\delta)} \cdot \mathcal{A}_\delta(p) \ .$$

(Hier wird die Voraussetzung $0 \in \mathcal{U}$ ausgenutzt.)
Wegen 1. hat $\mathcal{A}_\delta(p)$ ein nichtleeres Inneres und enthält somit die offene Umgebung V eines (inneren) Punktes $\tilde{p} \in \mathcal{A}_\delta(p)$. Die Abbildung $g : \mathbb{R}^n \to \mathbb{R}^n,\ x \mapsto e^{A(t-\delta)}x$ ist ein Diffeomorphismus. Es folgt, dass $g(V)$ eine offene Menge ist, die in $\mathcal{A}_t(p)$ enthalten ist. $g(\tilde{p}) = e^{A(t-\delta)}\tilde{p}$ ist also ein innerer Punkt von $\mathcal{A}_t(p)$. □

Das aus der Optimierung bekannte Pontryaginsche Maximumprinzip charakterisiert sogenannte *extremale Steuerungen*. Das sind Steuerungen, die den Anfangszustand p in einen Randpunkt der erreichbaren Menge $\mathcal{A}_t(p)$ zur Zeit $t \geq 0$ steuern. Dies wird im nächsten Abschnitt zur Herleitung des schwachen Bang-Bang Prinzips für spezielle Spalten-Kontrollsysteme führen, deren erreichbaren Mengen zudem strikt konvex sein werden (für $t \leq \omega(p)$).

Satz 3.75 (Maximumprinzip)**.** Es sei x_0 ein Punkt auf dem Rand $\delta\mathcal{A}_\theta(p)$ der kompakten erreichbaren Menge $\mathcal{A}_\theta(p)$ zu einem Zeitpunkt θ. Angenommen, es existiert ein $q \in \mathbb{R}^n \backslash \{0\}$, so dass

$$q^* x \leq q^* x_0 \qquad \forall x \in \mathcal{A}_\Theta(p) \ . \tag{3.47}$$

(q ist der äußere Normalenvektor einer Stützhyperebene durch x_0.) Dann gilt für jede zulässige Kontrolle $u(\cdot)$, die p nach x_0 zur Zeit θ steuert:

$$(q^* X_\theta X_t^{-1}) \cdot u(t) \cdot (c^* X_t p) = \max_{v \in \mathcal{U}} (q^* X_\theta X_t^{-1}) \cdot v \cdot (c^* X_t p) \quad \forall' t \in [0, \theta] \ , \tag{3.48}$$

wobei $t \mapsto X_t := \Phi(t;u)$ die Fundamentallösung zur Kontrolle u sei.

Beweis. Für irgendein Element $v \in \mathcal{U}$ und ein Teilintervall $[t, t+h] \subseteq [0,\theta]$ mit $h > 0$ betrachte die (zulässige) Kontrolle, welche auf $[t, t+h]$ den konstanten Wert v besitzt und sonst mit der extremalen Kontrolle u übereinstimmt. Die zugehörige Fundamentallösung $t \mapsto Z_t$ können wir anhand der Fundamentallösung $t \mapsto X_t$ (bezüglich u) und der Formel (2.17) auswerten:

$$Z_{t+h} = \Phi(h;v)\cdot Z_t = e^{(A+vc^*)h}\cdot X_t \ , \qquad X_\theta = \Phi(\theta-(t+h);u^{-(t+h)})\cdot X_{t+h} \ ,$$

und daher

$$Z_\theta = \Phi(\theta-(t+h);u^{-(t+h)})\cdot Z_{t+h} = X_\theta X_{t+h}^{-1} e^{(A+vc^*)h} X_t \ .$$

Wir setzen

$$x := Z_\theta p = X_\theta X_{t+h}^{-1} e^{(A+vc^*)h} X_t p \in \mathcal{A}_\theta(p) \ .$$

Weiter ist

$$x_0 = X_\theta p = X_\theta X_{t+h}^{-1} X_{t+h} X_t^{-1} X_t p \ .$$

Dies ergibt eingesetzt in die Ungleichung (3.47)

$$0 \le q^*(x_0 - x) = q^* X_\theta X_{t+h}^{-1} (X_{t+h} X_t^{-1} - e^{(A+vc^*)h}) X_t p \ .$$

Nun teilen wir durch $h > 0$ und formen um:

$$0 \le q^* X_\theta X_{t+h}^{-1} \left(\frac{(X_{t+h} - X_t)}{h}\cdot X_t^{-1} - \frac{(e^{(A+vc^*)h} - \mathrm{I})}{h} \right) X_t p$$

Es sei $h \mapsto X_{t+h}$ differenzierbar in $h = 0$. Dann gilt für $h \to 0$ fast überall

$$0 \le q^* X_\theta X_t^{-1} \Bigg(\underbrace{\frac{\mathrm{d}}{\mathrm{dh}} X_{t+h}\Big|_{h=0}}_{(A+u(t)c^*)X_t} \cdot X_t^{-1} - \underbrace{\frac{\mathrm{d}}{\mathrm{dh}} e^{(A+vc^*)h}\Big|_{h=0}}_{(A+vc^*)} \Bigg) X_t p$$

und folglich

$$0 \le q^* X_\theta X_t^{-1}\cdot (u(t) - v)\cdot c^* X_t p \ .$$

Da $v \in \mathcal{U}$ frei wählbar und diese Aussage für fast alle $t \in [0,\theta]$ korrekt ist, gilt schließlich

$$q^* X_\theta X_t^{-1}\cdot v\cdot c^* X_t p \le q^* X_\theta X_t^{-1}\cdot u(t)\cdot c^* X_t p \qquad \forall' t \in [0,\theta] \ \forall\, v \in \mathcal{U} \ ,$$

woraus die Behauptung folgt. □

Bemerkung 3.76. Die Funktion $y(t) := (X_t^{-1})^* X_\theta^* q$ aus Gleichung (3.48) ist die Lösung des sogenannten „adjungierten Systems“ von (AWP)

$$\dot{y} = -(A + uc^*)^* y, \qquad y(\theta) = q, \qquad t \in [0, \theta]$$

zur vorgegebenen Kontrolle u. (Dies ist sogar ein Spalten-Kontrollsystem der Form (2.25) mit Endbedingung.)
Denn $t \mapsto (X_t^{-1})^* := Y_t$ ist die Fundamentallösung des adjungierten Systems zur Kontrolle u. Um dies zu beweisen, genügt es zu zeigen, dass $Z_t := X_t^* \cdot Y_t$ konstant gleich I ist. In der Tat ist

$$\dot{Z}_t = (\dot{X}_t^*) Y_t + X_t^* \dot{Y}_t = X_t^* (A + uc^*)^* Y_t - X_t^* (A + uc^*)^* Y_t = 0, \; Z(0) = \mathrm{I} \; .$$

Bemerkung 3.77. Aus der konvexen Analysis ist bekannt [17, p.67], dass jeder Randpunkt einer abgeschlossenen konvexen Menge in $\mathbb{R}^n$ ein Stützpunkt ist. Falls $\mathcal{A}_\theta(p)$ konvex ist, so ist folglich die Bedingung (3.47) für jeden Randpunkt x_0 erfüllbar.

Lemma 3.78. Es sei $x : [0, \Theta] \to \mathbb{R}^n$ die zulässige Lösung zu einer extremalen Kontrolle u mit Anfangspunkt $x(0) = p$ und Endpunkt $x(\Theta) \in \delta\mathcal{A}_\Theta(p)$. Dann gilt

$$x(t) \in \delta\mathcal{A}_t(p) \qquad \forall t \in [0, \Theta] \; .$$

(Diese Aussage gilt auch für allgemeine bilineare Systeme.)

Beweis. 1. Wie gewohnt bezeichne $\Phi(\cdot\,; v)$ die Fundamentallösung zu einer zulässigen Kontrolle v. Dann ist die lineare Abbildung $q \mapsto \Phi(t; v) \cdot q$ ein Diffeomorphismus im $\mathbb{R}^n$, da die Matrix $\Phi(t; v)$ invertierbar ist. Wir folgern: Ist $P \subset \mathbb{R}^n$ eine offene Menge von Anfangswerten, so ist die Menge der Endpunkte $\Phi(t; v) \cdot P$, die ausgehend von Elementen aus P über die Kontrolle v zur Zeit t erreicht werden, offen.

2. Angenommen, es existiert ein $t \in [0, \Theta]$, so dass $x(t) \in \operatorname{int}(\mathcal{A}_t(p))$. Dann gilt $B_\epsilon(x(t)) \subset \mathcal{A}_t(p)$ für ein $\epsilon > 0$. In 1. haben wir gezeigt, dass die Menge $\Phi(\Theta - t; u^{-t}) \cdot B_\epsilon(x(t))$ offen ist, wobei $u^{-t}(\cdot) = u(\cdot + t)$ eine zeitverschobene Kontrolle ist. Wegen (2.17) und dem Additionstheorem aus Folgerung 3.9 ist

$$x(\Theta) = \Phi(\Theta; u) \cdot p = \Phi(\Theta - t; u^{-t}) \cdot \underbrace{\Phi(t; u) \cdot p}_{x(t)} \in \Phi(\Theta - t; u^{-t}) \cdot B_\epsilon(x(t))$$

und

$$\Phi(\Theta - t; u^{-t}) \cdot B_\epsilon(x(t)) \subset \mathcal{A}_{\Theta - t} \cdot \mathcal{A}_t \cdot p = \mathcal{A}_\Theta \cdot p = \mathcal{A}_\Theta(p) \; .$$

Es liegt also der Randpunkt $x(\Theta)$ in einer offenen Teilmenge von $\mathcal{A}_\Theta(p)$. Dies ist ein Widerspruch. □

3.5.2 Single-Input Systeme von Rang 1

Schließlich berücksichtigen wir noch bilineare Systeme in $\mathbb{R}^n$ der Form

$$\dot{x} = (A + ubc^*)x, \qquad |u(t)| \leq 1, \qquad x(0) = p \ , \tag{3.49}$$

gegeben durch eine Matrix $A \in \mathbb{R}^{n\times n}$ und Vektoren $b, c, p \in \mathbb{R}^n$.
Durch (3.49) ist offenbar ein spezielles Spalten-Kontrollsystem gegeben, dessen Steuerbereich $\mathcal{U} = [-1, 1]$ die Voraussetzung 3.67 erfüllt. Definitionen und Bezeichnungen aus dem vorherigen Abschnitt—wie z.B. die Konvexitätsausdehnung $\omega(p)$—übertragen wir auf (3.49). Alle Ergebnisse von dort sind anwendbar. Wir wollen nun die besondere Struktur des Systems nutzen, um dessen erreichbaren Mengen $\mathcal{A}_t(p)$ besser beschreiben zu können.

Erinnerung 3.79. Mit (3.49) assoziieren wir ein lineares System in $\mathbb{R}^n$ mit Ausgang:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), \qquad |u(t)| \leq 1 \ , \\ y(t) &= c^*x(t) \ . \end{aligned} \tag{3.50}$$

Falls (3.50) beobachtbar ist, d.h. die Menge $\mathcal{N} = \langle A^*; \mathrm{Im}c\rangle^\perp$ aller unbeobachtbaren Zustände ist gleich $\{0\}$, so ist (3.49) kontrollierbar mit 0 als einzigem komplett unkontrollierbaren Zustand. Aufgrund der starken Invarianz von $\mathcal{N}$ ist 0 von keinem Zustand ungleich Null erreichbar, und die Funktion $t \mapsto c^*x(t)$ besitzt nur isolierte Nullstellen für jede zulässige Lösung mit Anfangswert $p \neq 0$ (Folgerung 3.62). Außerdem gibt es für alle $p \neq 0$ eine sog. Konvexitätsausdehnung $\omega(p) > 0$ gemäß Definition 3.71, so dass die erreichbaren Mengen $\mathcal{A}_t(p)$ mit $0 \leq t \leq \omega(p)$ konvex sind.
Ist (3.50) zudem noch vollständig kontrollierbar, d.h. $\langle A; \mathrm{Im}b\rangle = \mathbb{R}^n$, so ist im Fall $p \neq 0$ die affine Hülle von $\mathcal{A}_t(p)$ der gesamte $\mathbb{R}^n$ (Satz 3.66) und es gilt $\mathrm{int}\,(\mathcal{A}_t(p)) \neq \emptyset$ für alle $t > 0$ (Folgerung 3.74).

Um lästige Fallunterscheidungen zu vermeiden, fordern wir ab jetzt, dass die folgenden Bedingungen eingehalten werden.

Voraussetzung 3.80. 1. Das assoziierte lineare System (3.50) sei beobachtbar und vollständig kontrollierbar.

2. Der Anfangszustand p sei ungleich 0.

Beispiel 3.81. Das gut bekannte System (3.36) mit Parametern $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right)$ und $c = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ erfüllt die Bedingung 3.80:

$$\begin{aligned} \langle A; \mathrm{Im}b\rangle &= \mathrm{Im}\big(b|Ab\big) = \mathrm{Im}\left(\begin{smallmatrix} 0 & -1 \\ -1 & 0 \end{smallmatrix}\right) = \mathbb{R}^2 \\ \langle A^*; \mathrm{Im}c\rangle^\perp &= \ker\big(c|A^*c\big) = \ker\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right) \end{aligned}$$

Beispiel 3.82. Beim Pontryaginschen Maximumprinzip ist das adjungierte System von (3.49)

$$\dot{x} = -(A + ubc^*)^* x = -(A^* + ucb^*)x, \qquad |u(t)| \leq 1 \tag{3.51}$$

von besonderer Bedeutung (Bemerkung 3.76). Dies ist ebenfalls ein Single-Input System von Rang 1, dessen assoziiertes lineares System bis auf Vorzeichen mit dem assoziierten linearen System bezüglich des dualen Systems von (3.50)

$$\begin{aligned} \dot{x}(t) &= A^* x(t) + cu(t), \qquad |u(t)| \leq 1 \ , \\ y(t) &= b^* x(t) \end{aligned} \tag{3.52}$$

übereinstimmt. Aus der Theorie linearer Systeme ist bekannt, dass ein lineaeres Kontrollsystem mit Ausgang genau dann beobachtbar und kontrollierbar ist, wenn das duale System kontrollierbar und beobachtbar ist. Das adjungierte bilineare System (3.51) genügt somit der Voraussetzung 3.80.

Sprechweise 3.83. Wenn wir im nächsten Satz sagen, dass die Kontrolle $u : [0, \Theta] \to [-1, 1]$ eine „stückweise konstante Bang-Bang Funktion“ ist, so meinen wir, dass u fast überall mit einer stückweise konstanten Funktion (Definition A.2) der Form $\tilde{u} : [0, \Theta] \to \{-1, +1\}$ übereinstimmt.

Der folgende Satz stützt sich auf die eindeutige Lösbarkeit der beim Maximumprinzip auftretenden Optimierungsaufgabe (3.48).

Satz 3.84. Es sei $x_0 \in \delta\mathcal{A}_\theta(p)$ ein Randpunkt mit $0 \leq \theta \leq \omega(p)$, $\theta < \infty$. Dann ist jede zulässige Kontrolle $u(\cdot)$, die den Anfangszustand p nach x_0 zur Zeit θ steuert, eine stückweise konstante Bang-Bang Funktion, die bis auf eine Nullmenge eindeutig durch die Parameter (p, x_0, θ) bestimmt ist.

Beweis. 1. Da $\mathcal{A}_\theta(p)$ konvex ist (beachte $\theta \leq \omega(p)$), existiert ein $q \in \mathbb{R}^n \backslash \{0\}$, so dass die Bedingung (3.47) erfüllt ist. Das Maximumprinzip 3.75 liefert

$$(q^* X_\theta X_t^{-1} b) \cdot u(t) \cdot (c^* X_t p) = \max_{-1 \leq v \leq +1} (q^* X_\theta X_t^{-1} b) \cdot v \cdot (c^* X_t p) \quad \forall' t \in [0, \theta] \ ,$$

wobei $t \mapsto X_t$ die Fundamentallösung zur extremalen Kontrolle u ist. Weil es sich bei den eingeklammerten Faktoren um reelle Zahlen handelt, können wir anhand des Vorzeichens dieser Faktoren die Werte der Kontrolle u ableiten. Dies klappt natürlich nur, wenn die Vorzeichen ungleich Null sind.

2. Nach Folgerung 3.62 hat der rechte Faktor $c^* X_t p$ nur isolierte Nullstellen. Da $[0, \Theta]$ ein kompaktes Intervall ist, kann er höchstens endlich viele Nullstellen auf diesem Intervall besitzen. Außerhalb dieser Nullstellen hat $c^* X_t p$

das Vorzeichen $\sigma(p) = c^* A^k p$ mit $k = \min\{0 \le j \le n-1 \mid c^* A^j p \neq 0\}$ (Folgerung 3.71).

3. Dies gilt analogerweise auch für den linken Faktor $q^* X_\theta X_t^{-1} b$. Um dies nachzuweisen, betrachten wir das adjungierte System

$$\dot{y} = -(A^* + ucb^*)y \ .$$

Da $q^* X_\theta X_t^{-1} b$ eine reelle Zahl ist, können wir ohne Einschränkung transponieren und erhalten $b^*(X_t^{-1})^* q_1$ für $q_1 := X_\theta^* q$. Wie in der Bemerkung 3.76 festgestellt, ist $y(t) := (X_t^{-1})^* q_1$ die Lösung des adjungierten Systems zur Kontrolle u, welche $y(0) = q_1$ erfüllt. Laut Beispiel 3.82 ist auch das assoziierte lineare System zum adjungierten System kontrollierbar. Da $q_1 \neq 0$ (X_θ^* hat vollen Rang), ist q_1 nicht komplett unkontrollierbar bezüglich des adjungierten Systems. Die Funktion $b^* y(\cdot\,)$ hat somit höchstens endlich viele Nullstellen auf $[0, \Theta]$ (Folgerung 3.62 angewandt auf (3.51)).
Wir wenden Folgerung 3.63 auf die komplett unkontrollierbaren Punkte $y(t)$—sie übernehmen die Rolle von $p \notin \mathcal{N}$—an, um das Vorzeichen von $b^* y(t)$ außerhalb der Nullstellen zu bestimmen. Dies ergibt, dass für fast alle (mit höchstens endlich vielen Ausnahmen) $t \in [0, \Theta]$ ein kleinstes $k = k(t) \in \mathbb{N}$ mit $0 \le k \le n-1$ und $b^*(A^k)^* y(t) \neq 0$ existiert, so dass

$$\operatorname{sgn}(b^* y(t)) = \operatorname{sgn}\left(b^*(A^k)^* y(t)\right) \ .$$

Insbesondere haben wir somit gezeigt, dass für den linken Faktor aus dem Maximumprinzip

$$\operatorname{sgn}\left(q^* X_\theta X_t^{-1} b\right) = \operatorname{sgn}(b^* y(t)) \neq 0 \qquad \forall' t \in [0, \Theta]$$

gilt.

4. Die Eingabegröße $u(t) \in [-1, +1]$ maximiert den Ausdruck

$$(q^* X_\theta X_t^{-1} b) \cdot v \cdot (c^* X_t p) \quad \text{fast überall.}$$

In 2. und 3. wurde bewiesen, dass die beiden eingeklammerten Faktoren für fast alle $t \in [0, \Theta]$ mit höchstens endlich viele Ausnahmen ungleich Null sind. Notwendigerweise gilt dann fast überall

$$u(t) = \operatorname{sgn}\left(q^* X_\theta X_t^{-1} b\right) \cdot \operatorname{sgn}\left(c^* X_t p\right) = \operatorname{sgn}\left(q^* X_\theta X_t^{-1} b\right) \cdot \sigma(p) \ . \tag{3.53}$$

Offenbar ist durch (3.53) eine stückweise konstante Funktion mit Werten in $\{+1, -1\}$ gegeben, weshalb die Kontrolle u eine stückweise konstante Bang-Bang Funktion ist.

5. Betrachte zwei zulässige Kontrollen $u, v : [0, \Theta] \to [-1, 1]$ mit zugehörigen Lösungen x bzw. y, die den Anfangswert p zum Randpunkt x_0 (zur Zeit Θ) steuern. Da $\Theta \leq \omega(p)$, ist nach dem dritten Beweisteil von Satz 3.70 auch $\frac{1}{2}(x + y)$ eine zulässige Lösung. Als zugehöriger Kontrolle können wir die Funktion

$$w := \frac{c^*x}{c^*x + c^*y}u + \frac{c^*y}{c^*x + c^*y}v \tag{3.54}$$

wählen, die bis auf den endlich vielen Nullstellen von c^*x und c^*y wohldefiniert ist. Wegen $\frac{1}{2}\big(x(\Theta) + y(\Theta)\big) = x_0$ steuert w ebenfalls x_0 an, und nach 4. sind die Kontrollen u, v, w stückweise konstante Bang-Bang Funktionen.

6. Wir zeigen jetzt, dass $u(t)$ und $v(t)$ fast überall übereinstimmen. Dazu seien die Kontrollen u, v, w aus 5. o.B.d.A. stückweise konstant—gemäß Definition A.2—mit Werten in $\{-1, +1\}$. Das Intervall $[0, \Theta]$ zerlegen wir folgendermaßen:
Da die Kontrollen u, v, w stückweise konstant sind und die Ausgangsfunktionen c^*x, c^*y nur endlich viele Nullstellen besitzen auf $[0, \Theta]$, existiert eine endliche Zerlegung

$$0 = \lambda_1 < \lambda_2 < \ldots < \lambda_r = \Theta \ ,$$

so dass auf jedem Intervall $(\lambda_i, \lambda_{i+1})$ alle Kontrollen konstant sind mit Werten in der Menge $\{-1, 1\}$ und so, dass die Werte der Ausgangsfunktionen auf $(\lambda_i, \lambda_{i+1})$ ungleich Null sind. Wir wählen ein solches Intervall $(\lambda_i, \lambda_{i+1})$ mit $1 \leq i \leq r - 1$ und beweisen, dass dort die Kontrollen u und v dieselben Werte besitzen. Nach Konstruktion sind u, v, w konstant gleich $+1$ oder -1 auf $(\lambda_i, \lambda_{i+1})$. Ohne Einschränkung gelte dort $w = v$, d.h. (3.54) wird für alle $t \in (\lambda_i, \lambda_{i+1})$ zu

$$\begin{aligned} v(t) &= \frac{c^*x(t)}{c^*x(t) + c^*y(t)}u(t) + \frac{c^*y(t)}{c^*x(t) + c^*y(t)}v(t) \\ &\iff \big(c^*x(t) + c^*y(t)\big)\, v(t) = c^*x(t)u(t) + c^*y(t)v(t) \\ &\iff c^*x(t)\,\big(u(t) - v(t)\big) = 0 \ . \end{aligned}$$

Man beachte, dass $c^*x(t)$ und $c^*y(t)$ auf $(\lambda_i, \lambda_{i+1})$ dasselbe Vorzeichen ungleich Null besitzen (denn $\theta \leq \omega(p)$), weshalb aus der letzten Gleichung $u(t) = v(t)$ für alle $t \in (\lambda_i, \lambda_{i+1})$ folgt. Da dies auf jedem Intervall der obigen Zerlegung von $[0, \Theta]$ gilt, ist die Behauptung „$u = v$ f.ü.“ bewiesen. □

Bevor wir uns mit der strikten Konvexität erreichbarer Mengen auseinandersetzen, wird ein Hilfsresultat aus der konvexen Analysis [17, p.16] angegeben.

Hilfssatz 3.85. Es sei X eine konvexe Teilmenge des $\mathbb{R}^n$. Dann ist

$$\{\lambda x + (1-\lambda)y \mid 0 \leq \lambda \leq 1\} \subset \{y\} \cup \operatorname{int} X$$

für alle $x \in \operatorname{int} X$ und $y \in \overline{X}$.

Folgerung 3.86 (Strikte Konvexität)**.** Die erreichbaren Mengen $\mathcal{A}_\Theta(p)$ sind unter der Voraussetzung

$$\Theta \leq \omega(p), \qquad 0 \leq \Theta < \infty$$

sogar strikt konvex, d.h.

$$x_0, y_0 \in \mathcal{A}_\Theta(p), x_0 \neq y_0, \lambda \in (0,1) \Longrightarrow \lambda x_0 + (1-\lambda)y_0 \in \operatorname{int}(\mathcal{A}_\Theta(p)) \ . \tag{3.55}$$

Beweis. 1. Da $\mathcal{A}_\Theta(p)$ konvex ist, können wir den Hilfssatz 3.85 anwenden. Der besagt einerseits, dass die Bedingung (3.55) erfüllt ist, falls zumindest einer der Punkte x_0 oder y_0 im Inneren von $\mathcal{A}_\Theta(p)$ liegt. Wir können also davon ausgehen, dass x_0 und y_0 Randpunkte sind. Andererseits erlaubt er es, die Bedingung (3.55) abzuschwächen: $\mathcal{A}_\Theta(p)$ ist genau dann strikt konvex, falls für $x_0, y_0 \in \mathcal{A}_\Theta(p)$ und $x_0 \neq y_0$ stets $\frac{1}{2}(x_0 + y_0) \in \operatorname{int}(\mathcal{A}_\Theta(p))$ gilt.

2. Angenommen, es existieren Randpunkte $x_0, y_0 \in \delta\mathcal{A}_\Theta(p)$ mit $x_0 \neq y_0$, so dass $\frac{1}{2}(x_0 + y_0) \in \delta\mathcal{A}_\Theta(p)$. Dann gibt es zulässige Kontrollen u, v und w, die den Anfangszustand p nach x_0, y_0 bzw. $\frac{1}{2}(x_0+y_0)$ zur Zeit θ steuern. Nach Satz 3.84 sind sie eindeutig bestimmte (bis auf Nullmenge), stückweise konstante Bang-Bang Funktionen. Insbesondere können wir (bis auf die Nullstellen von c^*x und c^*y)

$$w = \frac{c^*x}{c^*x + c^*y}u + \frac{c^*y}{c^*x + c^*y}v$$

fordern, wobei x und y die Lösungen zu u bzw. v sind. Wie in Beweisteil 6 von Satz 3.84 folgern wir, dass fast überall $u = v$ ist. Aber dann würden x und y übereinstimmen, und $x_0 = x(\Theta) = y(\Theta) = y_0$ würde zum Widerspruch führen. □

Hilfssatz 3.87. Sei $u : [0, \Theta] \to [-1, 1]$ eine zulässige zeitoptimale Steuerung, die den Anfangszustand p in minimaler Zeit Θ in den Zustand $q \in \mathcal{A}_\Theta(p)$ steuert. Falls $\Theta \leq \omega(p)$, so gilt $x(\Theta) \in \delta\mathcal{A}_\Theta(p)$.

Beweis. Angenommen, es ist $q \in \operatorname{int}(\mathcal{A}_\Theta(p))$. Nach Folgerung 3.74 und Voraussetzung 3.80 ist die Menge $\mathcal{A}_\Theta(p)$ kompakt und besitzt ein nichtleeres

Inneres. Es existiert also ein Simplex mit $n+1$ Ecken $q_0, q_1, \ldots, q_n$ aus $\mathcal{A}_\Theta(p)$, dessen Schwerpunkt der Punkt q ist. Dann gibt es zulässige Lösungen $x_k : [0,\Theta] \to \mathbb{R}^n$, $0 \le k \le n$, mit $x_k(0) = p$ und $x_k(\Theta) = q_k$. Das Simplex ist die konvexe Hülle seiner Ecken. Wir bezeichnen es durch

$$S(\Theta) := \operatorname{cvx}\left(x_0(\Theta), \ldots, x_n(\Theta)\right) \ .$$

Da die Funktionen $x_k(\cdot\,)$ stetig sind, ist q auch im Inneren des Simplex $S(t)$ für ein $t < \Theta$, das genügend nahe Θ ist. Wegen $t \le \omega(p)$ ist $\mathcal{A}_t(p)$ konvex. Es folgt $q \in S(t) \subseteq \mathcal{A}_t(p)$. Der Endpunkt q ist daher bereits zur Zeit $t < \Theta$ erreichbar, im Widerspruch zur Minimalität von Θ. □

Als Nächstes wird die Restriktion $\Theta \le \omega(p)$ wegfallen.

Satz 3.88 (Schwaches Bang-Bang Prinzip)**.** Jede zulässige extremale und auch jede zulässige zeitoptimale Kontrolle ist eine stückweise konstante Bang-Bang Funktion, d.h sie stimmt fast überall mit einer stückweise konstanten Funktion überein, deren Werte gleich ± 1 sind. Insbesondere ist das schwache Bang-Bang Prinzip (Definition 3.47) gültig.

Beweis. 1. Die messbare Funktion $u : [0,T] \to [-1,+1]$ steuere den Anfangszustand $p \neq 0$ zu einem Randpunkt $q \in \delta\mathcal{A}_T(p)$ (zur Zeit T). Die Funktion $x : [0,T] \to \mathbb{R}^n$ sei die zu u gehörige Lösung. In Lemma 3.78 wird gezeigt, dass der Wert $x(t)$ ein Randpunkt von $\mathcal{A}_t(p)$ für alle $t \in [0,T]$ ist.

2. Nach Voraussetzung 3.80 gilt $p \notin \mathcal{N}$. Da $\mathcal{N}$ stark invariant ist, folgt $x(t) \notin \mathcal{N}$ für alle $t \in [0,T]$. Wir erinnern uns an die Größe δ aus Bezeichnung 3.69. Für alle $t \in [0,T]$ setzen wir $\delta_t := \delta(x(t))$. Es gilt stets $0 < \delta_t \le \omega(x(t))$. Falls $T \le \omega(p)$, so folgt bereits aus Satz 3.84, dass die extremale Kontrolle u eine stückweise konstante Bang-Bang Funktion ist. Es sei also ohne Einschränkung $T > \omega(p)$. Die Intervalle $U_t := (t, t+\delta_t)$ bilden dann eine offene Überdeckung des kompakten Intervalls $[\delta_0, T]$, aus der wir eine endliche Teilüberdeckung $U_{t_1}, U_{t_2}, \ldots, U_{t_k}$ wählen können. Nachdem wir überlappte Intervalle entfernt und die Zeitpunkte t_i „sortiert“ haben, erhalten wir eine Zerlegung

$$0 := t_0 < t_1 < t_2 < \ldots < t_k < T := t_{k+1} \ ,$$

so dass $t_{i+1} - t_i < \delta_{t_i}$ für alle $0 \le i \le k$. Es genügt zu zeigen, dass u auf jedem Intervall $[t_i, t_{i+1}]$ eine stückweise konstante Bang-Bang Funktion ist.

3. Aus dem Additionstheorem folgt

$$\mathcal{A}_{t_{i+1}-t_i} \cdot x(t_i) \subseteq \mathcal{A}_{t_{i+1}-t_i} \cdot \mathcal{A}_{t_i} p = \mathcal{A}_{t_{i+1}}(p) \ni x(t_{i+1}) \ .$$

Der Zustand $x(t_{i+1})$ ist also nicht nur ein Randpunkt von $\mathcal{A}_{t_{i+1}}(p)$, sondern auch von $\mathcal{A}_{t_{i+1}-t_i} \cdot x(t_i)$. Die Kontrolle $u(\cdot)$ steuert auf $[t_i, t_{i+1}]$ den Zustand $x(t_i)$ zum Randpunkt $x(t_{i+1}) \in \delta\left(\mathcal{A}_{t_{i+1}-t_i} \cdot x(t_i)\right)$ in der Zeit $t_{i+1} - t_i \leq \omega\left(x(t_i)\right)$. Aus Satz 3.84 folgt, dass u eine stückweise konstante Bang-Bang Funktion auf $[t_i, t_{i+1}]$ ist. Da i beliebig vorgegeben ist, gilt dies auch auf dem gesamten Intervall $[0, T]$.

4. Es sei $v : [0, T] \to [-1, +1]$ eine zulässige Steuerung, die den Zustand p in minimaler Zeit T zum Zielpunkt q steuert. Es bezeichne $z(t) := \Phi(t; v)p$ die zugehörige Lösung. Wir zerlegen das Intervall $[0, T]$ analog zu 2. in Teilintervalle $[t_i, t_{i+1}]$ $(i = 0, 1, \ldots, k)$, so dass $t_{i+1} - t_i \leq \omega\left(z(t_i)\right)$ für alle i. Ohne das Bellmannsche Optimalitätsprinzip zitieren zu wollen, sollte es klar sein, dass die Restriktionen von v auf diese Teilintervalle zeitoptimal sind. Wenden wir Hilfssatz 3.87 auf die Restriktionen an, so folgt, dass sie extremal sind, und daher den stückweise konstanten Bang-Bang Funktionen zuzuordnen sind. Folglich gilt dies auch für v als endliche Konkatenation der Restriktionen. □

Beim Kompaktheitsschluss im letzten Beweis geht die Eindeutigkeit extremaler Kontrollen, die auf Intervallen außerhalb des Konvexitätsbereichs $[0, \omega(p)]$ definiert sind, meist verloren.

Beispiel 3.89. Betrachte wieder das Single-Input Rang-1 System aus Beispiel (3.36)

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x + u \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} x, \qquad 0 \leq u \leq 1, \qquad 0 \leq t \leq 2\pi \ .$$

Der Zustand $p = \left(\begin{smallmatrix} -1 \\ 0 \end{smallmatrix}\right)$ liegt auf dem Rand von $\mathcal{A}_{2\pi}(p)$. Um dies nachzuweisen, zeigen wir, dass p das einzige Element aus $\mathcal{A}_{2\pi}(p)$ ist, welches auf der negativen x-Achse liegt. Wir führen dazu wieder die Polarkoordinaten (r, ρ) ein. Um ausgehend von p wieder die negative x-Achse in der Zeit 2π zu erreichen, muß eine koordinaten-transformierte Lösung die Winkelgleichung (basierend auf (3.46))

$$\pi + 2k\pi = \rho(0) = \rho(2\pi) = -\pi + \int_0^{2\pi} \underbrace{(1 - u(s)) \cos^2 \rho(s)}_{\in [0,1]} \mathrm{ds} \ , \qquad k \in \mathbb{Z}$$

erfüllen. Dies ist nur möglich für $k = 0$ und $k = -1$, woraus $u = 0$ bzw. $u = 1$ (fast überall) folgen würde. Offenbar steuern die Kontrollen $u = 0$ und $u = 1$ den Punkt p zu sich selbst in der Zeit 2π. Der Zustand p ist daher ein Randpunkt, der von zwei unterschiedlichen extremalen Kontrollen

angesteuert wird. Die Eindeutigkeit extremaler Kontrollen ist nur auf dem Zeitintervall $[0, \frac{\pi}{2}]$ sichergestellt.

In Abbildung 3.6 sehen wir erreichbare Punkte aus der Menge $\mathcal{A}_{2\pi}(p)$. Diese gehören zu stückweise konstanten Bang-Bang Funktionen mit bis zu fünf Sprungstellen. Die Anzahl der verwendeten Sprungstellen ist durch die Färbung der Punkte ablesbar (z.B. grün $\simeq$ 5 Sprünge). Da das approximative Bang-Bang Prinzip gültig ist, können wir bei einer genügend großen Anzahl an Sprungstellen die erreichbare Menge $\mathcal{A}_{2\pi}(p)$ approximieren. Die Abbildung soll veranschaulichen, dass nur $p = \left(\begin{smallmatrix} -1 \\ 0 \end{smallmatrix}\right)$ auf der negativen x-Achse liegt.

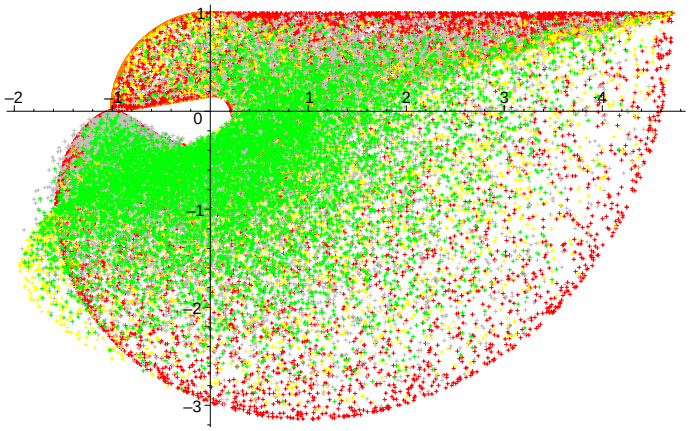

Abbildung 3.6: Erreichbare Zustände aus $\mathcal{A}_{2\pi}(p)$

Bemerkung 3.90. O. Hàjek zeigt in [10, p.307ff], dass die Theorie zu Single-Input Systemen von Rang 1 auch auf Spalten-Kontrollsysteme der Form

$$\dot{x} = (A + uc^*)x, \qquad u(t) \in \mathcal{U}, \qquad \mathcal{X} = \mathbb{R}^n$$

mit Steuerbereich

$$\mathcal{U} = \left\{ \sum_{k=1}^{m} \mu_k b_k \;\middle|\; -1 \leq \mu_k \leq 1 \right\}, \qquad b_1, \ldots, b_m \in \mathbb{R}^n$$

übertragbar ist. Es gibt für solche Systeme i.A. kein assoziiertes lineares System, das beobachtbar und kontrollierbar sein könnte. Man fordert stattdessen, dass die Vektoren b_k die sog. „Normalitätsbedingung“

$$\langle A + uc^*; b_k \rangle = \mathbb{R}^n \qquad \forall u \in \mathcal{U},\ 1 \leq k \leq m$$

erfüllen, um beispielsweise das schwache Bang-Bang Prinzip herleiten zu können.

Kapitel 4

Ein-Ausgangsverhalten bilinearer Systeme

4.1 Einführung

In diesem Kapitel wollen wir das Ein-Ausgangsverhalten bilinearer Systeme untersuchen. Dazu betrachten wir ein initialisiertes bilineares System Σ mit Ausgang, gegeben durch die Gleichungen

$$\begin{aligned} \dot{x}(t) &= Ax(t) + \sum_{i=1}^{m} N_i x(t) u_i(t) + Bu(t), \qquad t \in [0,T] \ , \qquad \text{(AWP)} \\ y(t) &= Cx(t) \ , \\ x(0) &= x_0 \ , \end{aligned}$$

mit Zustandsraum $\mathcal{X} = \mathbb{R}^n$, Steuerbereich $\mathcal{U} = \mathbb{R}^m$ und Messbereich $\mathcal{Y} = \mathbb{R}^q$. $A, B, C, N_1, \ldots, N_m$ sind konstante Matrizen von angemessener Form, d.h. $B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{q\times n}$ und $A, N_1, \ldots, N_m \in \mathbb{R}^{n\times n}$. Unser Interesse konzentriert sich auf das Zeitintervall $[0,T]$ mit fester Anfangszeit 0 und fester Endzeit $T > 0$. Die zulässigen Eingangsfunktionen seien genau die messbaren, lokal (essentiell) beschränkten Funktionen über diesem Intervall. Indem wir Kontrollfunktionen, die fast überall gleich sind, miteinander identifizieren, können wir den Banachraum $L^\infty([0,T],\mathbb{R}^m)$ als die Menge zulässiger Kontrollfunktionen wählen.
Wie in Abschnitt 1.4 gezeigt, können wir das Ein-Ausgangsverhalten durch eine Responsefunktion oder eine Ein-Ausgangsfunktion charakterisieren. Hält man Anfangs- und Endzeit fest, so ist die Responsefunktion aus Definition

1.3 durch die Abbildung

$$\begin{aligned}\lambda: \quad [0,T] \times L^\infty([0,T],\mathbb{R}^m) &\longrightarrow \mathbb{R}^q \\ \lambda(t;u) &:= Cx(t;x_0,u)\end{aligned}$$

gegeben, wobei $x(\cdot\,;x_0,u)$ die maximale Lösung von (AWP) zur Kontrolle u bezeichne. Die Responsefunktion ist nichts anderes als die punktweise Auswertung der Ein-Ausgangsfunktion

$$\begin{aligned}\Psi: \quad L^\infty([0,T],\mathbb{R}^m) &\longrightarrow C^0([0,T],\mathbb{R}^q) \\ \Psi(u)(t) &:= Cx(t;x_0,u)\ .\end{aligned}$$

Falls keine Ausgangsfunktion „$y(t) = Cx(t)$“ in der Systemgleichung angegeben ist, so gehen wir davon aus, dass alle Zustandsgrößen messbar sind (d.h. $\mathcal{Y} = \mathcal{X}$). Die Ausgangsfunktion soll dann implizit durch die Einheitsmatrix $C := \mathrm{I}$ des $\mathbb{R}^{n\times n}$ bestimmt sein.

Definition 4.1. Wir bezeichnen eine Folge von Partialsummen $s_k := \sum_{i=0}^{k} x_i$, dessen Summanden x_i Elemente aus einem Banachraum $(X, \|\cdot\|)$ sind, als Reihe $\sum_{i=0}^{\infty} x_i$ im Banachraum X. Die Reihe heißt *konvergent*, wenn die Folge (s_k) in X konvergiert. In diesem Fall wird der Grenzwert ebenfalls mit $\sum_{i=0}^{\infty} x_i$ bezeichnet.
Ferner heißt die Reihe *absolut konvergent*, falls die Reihe der Normen $\sum_{i=0}^{\infty} \|x_i\|$ konvergiert.
Da die Folge der Partialsummen einer absolut konvergenten Reihe eine Cauchy-Folge im vollständigen Raum X ist, folgt, dass jede absolut konvergente Reihe in einem Banachraum konvergiert.

Definition 4.2. Man sagt, die Ein-Ausgangsfunktion

$$\Psi: \quad L^\infty([0,T],\mathbb{R}^m) \longrightarrow C^0([0,T],\mathbb{R}^q)$$

hat eine *Volterra-Reihen-Repräsentation*, falls es zu jedem Multiindex $(i_1 \dots i_k)$ von beliebiger Länge $k > 0$, mit Elementen $i_1, \dots, i_k$ aus der Menge $\{1, \dots, m\}$, eine Funktion $\omega_{i_1 \dots i_k}$ gibt, so dass alle Funktionen $\omega_{i_1 \dots i_k}$ zusammen mit einer weiteren Funktion ω_0 den folgenden Bedingungen genügen:

1. ω_0 ist eine stetige Funktion der Form $\omega_0 : [0,T] \to \mathbb{R}^q$.

2. Jedes $\omega_{i_1 \dots i_k}$ ist eine stetige Funktion der Form $\omega_{i_1 \dots i_k} : D_k \to \mathbb{R}^q$, definiert über der Menge

$$D_k := \{(t, s_1, \dots, s_k) \mid 0 \le s_k \le \dots \le s_1 \le t \le T\}\ .$$

3. Es existiert ein $\delta > 0$, so dass für alle $t \in [0,T]$ gilt:

$$\Phi(u)(t) = \omega_0(t) + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \underbrace{\int_0^t \int_0^{s_1} \ldots \int_0^{s_{k-1}}}_{k} \omega_{i_1\ldots i_k}(t,s_1,\ldots,s_k) u_{i_1}(s_1)\ldots u_{i_k}(s_k) \mathrm{ds_k}\ldots \mathrm{ds_1} \ , \tag{4.1}$$

falls $\|u\|_\infty < \delta$. (Dabei ist $\int_0^{s_0} := \int_0^t$.)

4. Die Reihe in 3. ist für jede zulässige Kontrolle $u(\cdot)$ mit $\|u\|_\infty < \delta$ eine konvergente Reihe im Banachraum $\big(C([0,T],\mathbb{R}^q), \|\cdot\|_0\big)$. (Insbesondere konvergiert sie gleichmäßig über $[0,T]$.)

Sprechweise 4.3. Die Funktionen ω_0 und $\omega_{i_1\ldots i_k}$ werden auch (Volterra-) *Kerne* der Ordnung 0 bzw. k genannt. Die Reihe im dritten Definitionsteil heißt *Volterra-Reihe*.

Bezeichnung 4.4. Falls angebracht, sollen die Abkürzungen

$$\mathrm{ds}^k := \mathrm{ds_k ds_{k-1}}\ldots \mathrm{ds_1} \ ,$$
$$\omega_{i_1\ldots i_k}(t,s^k) := \omega_{i_1\ldots i_k}(t,s_1,\ldots,s_k)$$

die Schreibweise vereinfachen.

Bemerkung 4.5. In Teil 4 der Definition steckt implizit die Forderung, dass die Summanden der Volterra-Reihe stetige Funktionen in t sind (für eine vorgegebene Kontrolle u). Man kann sich leicht überlegen, dass dies bereits durch die Stetigkeit der Kerne impliziert wird. Denn für alle $k \in \mathbb{N}$ und jeden Multiindex $(i_1\ldots i_k)$, mit Elementen aus $\{1,\ldots,m\}$, ist die Funktion

$$\begin{aligned} [0,T]\times\mathbb{R}^k &\longrightarrow \mathbb{R}^q \\ (t,s) &\longmapsto \chi_{D_k}(t,s_1,\ldots,s_k)\cdot\omega_{i_1\ldots i_k}(t,s_1,\ldots,s_k)\cdot u_{i_1}(s_1)\ldots u_{i_k}(s_k) \end{aligned}$$

bis auf höchstens eine Sprungstelle von χ_{D_k} stetig in t (bei festem s), integrierbar in s (bei festem t) und essentiell beschränkt, zumal der Kern $\omega_{i_1\ldots i_k}$ stetig über der kompakten Menge D_k ist. Aufgrund des Satzes von der majorisierten Konvergenz [7, p.98] folgt, dass all diese Funktionen

$$t \longmapsto \sum_{i_1,\ldots,i_k=1}^{m} \int_0^t \int_0^{s_1} \ldots \int_0^{s_{k-1}} \omega_{i_1\ldots i_k}(t,s^k) u_{i_1}(s_1)\ldots u_{i_k}(s_k)\mathrm{ds}^k$$

(d.h die Summanden der Volterra-Reihe) stetig über $[0,T]$ sind.

Wir setzen uns die folgenden Ziele:

- Berechnung der Volterra-Reihen-Repräsentation für das System (AWP).
- Nachweis der Eindeutigkeit dieser Repräsentation.
- Herleitung eines Kriteriums für die Endlichkeit von Volterra-Reihen.

Im letzten Abschnitt dieses Kapitels wollen wir kurz auf die Anwendbarkeit von Volterra-Reihen eingehen. Dies wird unsere Bemühungen rechtfertigen.

4.2 Existenz der Volterra-Reihen-Repräsentation

Wir wollen die Volterra-Reihen-Repräsentation der Ein-Ausgangsfunktion von (AWP) herleiten. Dazu müssen die Kerne berechnet werden. Ein wichtiges Hilfsmittel ist Variation der Konstanten. Setzt man

$$b(t) := \sum_{i=1}^{m} N_i x(t) u_i(t) + Bu(t)$$

in die Formel (B.27) ein für ein vorgegebenes $u \in L^\infty([0,T],\mathbb{R}^m)$ mit zug. Lösung $x(\cdot)$ von (AWP), so erhält man

$$\Psi(u)(t) = Ce^{At}x_0 + \int_0^t Ce^{A(t-s)}\Big(\sum_{i=1}^{m} N_i x(s)u_i(s) + Bu(s)\Big)\mathrm{ds}\ . \tag{4.2}$$

Beispiel 4.6 (Lineare Systeme)**.** Sind die Matrizen $N_1,\dots,N_m$ gleich Null, so ist die Volterra-Reihen-Repräsentation gefunden:

$$\Psi(u)(t) = Ce^{At}x_0 + \int_0^t Ce^{A(t-s)}Bu(s)\mathrm{ds} = Ce^{At}x_0 + \sum_{i=1}^{m}\int_0^t Ce^{A(t-s)}b^i u_i(s)\mathrm{ds}\ , \tag{4.3}$$

wobei b^i die i-te Spalte von B sei. Die Kerne sind

$$\omega_0(t) = Ce^{At}x_0\ , \tag{4.4a}$$

$$\omega_i(t,s) = Ce^{A(t-s)}b^i\ , \tag{4.4b}$$

$$\omega_{i_1\dots i_k}(t,s^k) = 0 \qquad \forall k \geq 2\ . \tag{4.4c}$$

Sonst läßt sich keine direkte Abhängigkeit der Ausgangsfunktion von der Eingangsfunktion herleiten, da auf der rechten Seite von (4.2) der Zustandsvektor x auftaucht. Ein indirekter Ansatz soll helfen die Kerne zu bestimmen: Man approximiert bilineare Systeme durch lineare, deren Kerne aus (4.4) bekannt sind, und berechnet die Kerne der Approximation.

Satz 4.7 (Approximationseigenschaft bilinearer Systeme)**.** Sei $x(\cdot)$ die Lösung von (AWP) bzgl. der Eingangsfunktion $u \in L^\infty([0,T],\mathbb{R}^m)$.
Dann existiert eine Folge stetiger Funktionen $(x_k)_{k\in\mathbb{N}} \subset C([0,T],\mathbb{R}^n)$ mit

$$\lim_{k\to\infty} x_k = x \tag{4.5}$$

im normierten Raum $\big(C([0,T],\mathbb{R}^n), \|\cdot\|_0\big)$.
Die Funktionen erhält man sukzessiv als Lösungen der linearen Differentialgleichungen

$$\begin{aligned} \dot{x_0}(t) &= Ax_0(t) + Bu(t)\ , \\ \dot{x_k}(t) &= Ax_k(t) + \sum\nolimits_{i=1}^{m} N_i x_{k-1}(t)u_i(t) + Bu(t) \qquad (k=1,2,\ldots) \end{aligned} \tag{4.6}$$

mit gemeinsamer Anfangsbedingung $x_k(0) = x_0$ für $k \geq 0$.

Beweis. Ich führe die „Fehlerfunktion“

$$z_k(t) = x(t) - x_k(t) \qquad (k=0,1,\ldots) \tag{4.7}$$

ein. Es ist zu zeigen, dass $\|z_k(t)\|$ gleichmäßig gegen Null konvergiert im Intervall $[0,T]$ (für $k\to\infty$).
Ohne Einschränkung[1] sei $\|u(t)\| < \delta$ für alle $t \in [0,T]$.
Variation der Konstanten (B.27) liefert:

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)} \left(\sum_{i=1}^{m} N_i x(s)u_i(s) + Bu(s) \right) \mathrm{ds}$$

und

$$x_k(t) = e^{At}x_0 + \int_0^t e^{A(t-s)} \left(\sum_{i=1}^{m} N_i x_{k-1}(s)u_i(s) + Bu(s) \right) \mathrm{ds} \qquad (k=1,2,\ldots) \tag{4.8}$$

Dies setzt man in (4.7) ein und rechnet aus:

$$z_k(t) = \int_0^t e^{A(t-s)} \left(\sum_{i=1}^{m} N_i z_{k-1}(s)u_i(s) \right) \mathrm{ds}$$

[1] Wähle beschränkten Repräsentanten aus $L^\infty([0,T],\mathbb{R}^m)$

Es gilt folgende Abschätzung für alle $t, s \in [0,T], a \in \mathbb{R}^n$:

$$\begin{aligned}\left\| e^{A(t-s)} \sum_{i=1}^{k} N_i a u_i(s) \right\| &\leq \|e^{A(t-s)}\| \cdot \left\| \sum_{i=1}^{m} N_i a u_i(s) \right\| \\ &\leq \|e^{A(t-s)}\| \cdot \sum_{i=1}^{m} (\|N_i a\| \cdot |u_i(s)|) \\ &\leq \underbrace{\max_{t,s\in[0,T]} \|e^{A(t-s)}\| \cdot \delta \cdot \left(\sum_{i=1}^{m} \|N_i\| \right)}_{:=M} \cdot \|a\| \leq M\|a\|\end{aligned}$$

Also

$$\begin{aligned}\|z_k(t)\| &\leq \int_0^t \left\| e^{A(t-s)} \sum_{i=1}^{m} N_i \underbrace{z_{k-1}(s)}_{a} u_i(s) \right\| \mathrm{ds} \leq M \int_0^t \|z_{k-1}(s)\| \mathrm{ds} \\ &\leq M^2 \int_0^t \int_0^{s_1} \|z_{k-2}(s_2)\| \mathrm{ds_2 ds_1} \leq \dots \\ &\leq M^k \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \|z_0(s_k)\| \mathrm{ds_k ds_{k-1} \dots ds_1} \ .\end{aligned}$$

Mit $x_0(\cdot), x(\cdot)$ ist auch $z_0(\cdot)$ stetig über $[0,T]$ und es gibt daher eine Konstante $K > 0$, so dass

$$\|z_0(t)\| \leq K \qquad \forall t \in [0,T] \ .$$

Und wir folgern für alle t aus $[0,T]$:

$$\|z_k(t)\| \leq M^k K \underbrace{\int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \mathrm{ds_k ds_{k-1} \dots ds_1}}_{\frac{t^k}{k!}} \leq \frac{(MT)^k}{k!} K \quad (\text{für } k \in \mathbb{N}_0)$$

Schließlich

$$\sum_{k=0}^{\infty} \|z_k\|_0 \leq \sum_{k=0}^{\infty} \frac{(MT)^k}{k!} K = K e^{MT} < \infty$$

und somit $\|z_k\|_0 \to 0$ für $k \to \infty$. □

Wir führen die Funktionenfolge $(p_k)_{k\in\mathbb{N}}$ mit

$$p_0 := x_0 \ , \qquad p_k := x_k - x_{k-1} \qquad (k = 1, 2, \dots)$$

ein, um den Limes (4.5) als Reihe in $C([0,T], \mathbb{R}^n)$ auszuwerten:

$$x = \lim_{l\to\infty} x_l = \lim_{l\to\infty} \left(\underbrace{x_0}_{p_0} + \sum_{k=1}^{l} (x_k - x_{k-1}) \right) = \sum_{i=0}^{\infty} p_k \tag{4.9}$$

Insbesondere gilt dann punktweise

$$x(t) = \sum_{k=0}^{\infty} p_k(t) \qquad \forall t \in [0,T] \; . \tag{4.10}$$

Die Funktion $p_0(t)$ ist die Lösung von (4.6), d.h.

$$p_0(t) = e^{At}x_0 + \sum_{i=1}^{m} \int_0^t e^{A(t-s)} b^i u_i(s) \mathrm{ds} \; . \tag{4.11}$$

Durch sukzessive Substituitionen mittels (4.8) erhalten wir eine ähnliche Darstellung für die restlichen Funktionen $p_k(t)$, falls $k \geq 1$:

$$\begin{aligned}
p_k(t) = x_k(t) - x_{k-1}(t) &\overset{(4.8)}{=} \sum_{i_1=1}^{m} \int_0^t e^{A(t-s_1)} N_{i_1} \underbrace{p_{k-1}(s_1)}\, u_{i_1}(s_1) \mathrm{ds}_1 \\
&\overset{(4.8)}{=} \sum_{i_1=1}^{m} \int_0^t e^{A(t-s_1)} N_{i_1} \left[\sum_{i_2=1}^{m} \int_0^{s_1} e^{A(s_1-s_2)} N_{i_2} \underbrace{p_{k-2}(s_2)}\, u_{i_2}(s_2) \mathrm{ds}_2 \right] u_{i_1}(s_1) \mathrm{ds}_1 \\
&= \sum_{i_1=1}^{m} \int_0^t e^{A(t-s_1)} N_{i_1} \left[\sum_{i_2=1}^{m} \int_0^{s_1} e^{A(s_1-s_2)} N_{i_2} \left[\sum_{i_3=1}^{m} \int_0^{s_2} e^{A(s_2-s_3)} N_{i_3} \dots \right. \right. \\
&\dots \left[\sum_{i_k=1}^{m} \int_0^{s_{k-1}} e^{A(s_{k-1}-s_k)} N_{i_k} \left[e^{As_k} x_0 + \sum_{i_{k+1}=1}^{m} \int_0^{s_k} e^{A(s_k-s_{k+1})} b^{i_{k+1}} \cdot \right. \right. \\
&\qquad \left. \left. \cdot \, u_{i_{k+1}}(s_{k+1}) \mathrm{ds}_{k+1} \right] u_{i_k}(s_k) \mathrm{ds}_k \right] \dots u_{i_3}(s_3) \mathrm{ds}_3 \Big] u_{i_2}(s_2) \mathrm{ds}_2 \Big] u_{i_1}(s_1) \mathrm{ds}_1
\end{aligned} \tag{4.12}$$

Wir nutzen die Linearitäten in (4.12), um die Integrale und das Pluszeichen nach außen zu ziehen, und setzen das Ergebnis zusammen mit (4.11) in die Reihe (4.10) ein. Auf diese Weise gewinnen wir eine Darstellung des Zustands $x(t)$, die nur von der Eingabefunktion $u(\cdot)$ und dem Anfangszustand x_0 abhängt. Derartige Darstellungen nennt man *Input-to-State Repräsenta-*

tionen:

$$x(t) = e^{At}x_0 + \tag{4.13a}$$

$$+\sum_{i=1}^{m}\int_0^t e^{A(t-s)}b^i u_i(s)\mathrm{ds} + \tag{4.13b}$$

$$+\sum_{k=1}^{\infty}\sum_{i_1,\ldots,i_{k+1}=1}^{m}\int_0^t\int_0^{s_1}\ldots\int_0^{s_k} e^{A(t-s_1)}N_{i_1}e^{A(s_1-s_2)}N_{i_2}\ldots \tag{4.13c}$$

$$\ldots e^{A(s_{k-1}-s_k)}N_{i_k}e^{A(s_k-s_{k+1})}b^{i_{k+1}}u_{i_1}(s_1)u_{i_2}(s_2)\ldots u_{i_{k+1}}(s_{k+1})\mathrm{ds}^{k+1}$$

$$+\sum_{k=1}^{\infty}\sum_{i_1,\ldots,i_k=1}^{m}\int_0^t\int_0^{s_1}\ldots\int_0^{s_k} e^{A(t-s_1)}N_{i_1}e^{A(s_1-s_2)}N_{i_2}\ldots \tag{4.13d}$$

$$\ldots e^{A(s_{k-1}-s_k)}N_{i_k}e^{As_k}x_0u_{i_1}(s_1)u_{i_2}(s_2)\ldots u_{i_k}(s_k)\mathrm{ds}^k$$

Der Ausdruck lässt sich weiter vereinfachen, indem wir (4.13b) und (4.13c) zusammenfassen zu

$$\sum_{k=0}^{\infty}\sum_{i_1,\ldots,i_{k+1}=1}^{m}\int_0^t\int_0^{s_1}\ldots\int_0^{s_k} e^{A(t-s_1)}N_{i_1}e^{A(s_1-s_2)}N_{i_2}\ldots$$

$$\ldots e^{A(s_{k-1}-s_k)}N_{i_k}e^{A(s_k-s_{k+1})}b^{i_{k+1}}u_{i_1}(s_1)u_{i_2}(s_2)\ldots u_{i_{k+1}}(s_{k+1})\mathrm{ds}^{k+1}\ ,$$

und den Index k inkrementiern

$$\sum_{k=1}^{\infty}\sum_{i_1,\ldots,i_k=1}^{m}\int_0^t\int_0^{s_1}\ldots\int_0^{s_{k-1}} e^{A(t-s_1)}N_{i_1}e^{A(s_1-s_2)}N_{i_2}\ldots$$

$$\ldots e^{A(s_{k-2}-s_{k-1})}N_{i_{k-1}}e^{A(s_{k-1}-s_k)}b^{i_k}u_{i_1}(s_1)u_{i_2}(s_2)\ldots u_{i_k}(s_k)\mathrm{ds}^k\ .$$

Indem wir die rechte Seite von (4.13) von links mit der Matrix C multiplizieren, erhalten wir die Ausgangsgröße $y(t) = Cx(t)$. Sie entspricht dem Wert der Responsefunktion $\lambda(t;u)$.

Lemma 4.8 (Zerlegungseigenschaft der Responsefunktion)**.** Die Responsefunktion des bilinearen Systems (AWP) läßt sich zerlegen in die Summe

$$\lambda(t;u) = \lambda_{x_0}(t;u) + \lambda_u(t;u) + \lambda_{x_0u}(t;u)$$

von drei Termen, die sich jeweils in Volterra-Reihen entwickeln lassen:

$$\begin{aligned}
\lambda_{x_0}(t;u) &= Ce^{At}x_0\ ,\\
\lambda_u(t;u) &= \sum_{k=1}^{\infty}\sum_{i_1,\dots,i_k=1}^{m}\int_0^t\int_0^{s_1}\dots\int_0^{s_{k-1}} Ce^{A(t-s_1)}N_{i_1}e^{A(s_1-s_2)}N_{i_2}\dots\\
&\qquad \dots e^{A(s_{k-2}-s_{k-1})}N_{i_{k-1}}e^{A(s_{k-1}-s_k)}b^{i_k}u_{i_1}(s_1)u_{i_2}(s_2)\dots u_{i_k}(s_k)\mathrm{ds}^k,\\
\lambda_{x_0u}(t;u) &= \sum_{k=1}^{\infty}\sum_{i_1,\dots,i_k=1}^{m}\int_0^t\int_0^{s_1}\dots\int_0^{s_k} Ce^{A(t-s_1)}N_{i_1}e^{A(s_1-s_2)}N_{i_2}\dots\\
&\qquad \dots e^{A(s_{k-1}-s_k)}N_{i_k}e^{As_k}x_0u_{i_1}(s_1)u_{i_2}(s_2)\dots u_{i_k}(s_k)\mathrm{ds}^k
\end{aligned}$$

Der erste Term $\lambda_{x_0}(t;u)$ ist der sogenannte *Zero-Input Response*—er entspricht dem Wert der Responsefunktion $\lambda(t;0)$ zur Kontrolle $u=0$.
Der zweite Term $\lambda_u(t;u)$ ist der sogenannte *Zero-State Response*—er bestimmt den Teil der Ausgangsgröße $y(t) := \lambda(t;u)$, der nicht vom Anfangszustand x_0 abhängt.

Satz 4.9 (Existenz der Volterra-Reihen-Repräsentation)**.** Für alle zulässigen Kontrollfunktionen $u(\cdot)$ ist die Reihe

$$\begin{aligned}
&\Psi(u)(t) = \omega_0(t)+\\
&\sum_{k=1}^{\infty}\sum_{i_1,\dots,i_k=1}^{m}\int_0^t\int_0^{s_1}\dots\int_0^{s_{k-1}} \omega_{i_1\dots i_k}(t,s_1,\dots,s_k)u_{i_1}(s_1)\dots u_{i_k}(s_k)\mathrm{ds_k}\dots\mathrm{ds_1}
\end{aligned}$$

mit den Kernen

$$\begin{aligned}
\omega_0(t) &= Ce^{At}x_0\ , \qquad (4.14)\\
\omega_{i_1}(t,s_1) &= Ce^{A(t-s_1)}(b^{i_1}+N_{i_1}e^{As_1}x_0)\ ,\\
\omega_{i_1\dots i_k}(t,s^k) &= Ce^{A(t-s_1)}\cdot N_{i_1}e^{A(s_1-s_2)}\cdot\\
&\qquad \cdot N_{i_2}e^{A(s_2-s_3)}\cdot\\
&\qquad \vdots\\
&\qquad \cdot N_{i_{k-1}}e^{A(s_{k-1}-s_k)}\cdot (b^{i_k}+N_{i_k}e^{As_k}x_0) \qquad \forall k\geq 2
\end{aligned}$$

absolut konvergent in $C([0,T],\mathbb{R}^q)$, d.h. $\sum_{k=0}^{\infty}\|\dots\|_0<\infty$.
Die Volterra-Reihen-Repräsentation der Ein-Ausgangsfunktion des bilinearen Systems (AWP) ist gefunden.

Beweis. 1. Die Konvergenz in $C([0,T],\mathbb{R}^q)$ folgt bereits aus Satz 4.7. Denn die Funktionenfolge $x_k(\cdot)$ von dort entspricht nach Konstruktion (4.9) genau

den Partialsummen der Volterra-Reihe (bis auf Linksmultiplikation mit der Matrix C). Weil jede absolut konvergente Reihe im Banachraum $C([0,T],\mathbb{R}^q)$ konvergiert, wird dies gleich nochmals bewiesen. Wegen Lemma 4.8 sind dann alle Bedingungen aus Definition 4.2 durch die gewählten Kerne erfüllt.

2. Wir zeigen zunächst, dass die Kerne auf ihrem Definitionsbereich eine Wachstumsbedingung erfüllen:
Es existieren reelle Zahlen $K \geq 0$ und $M > 0$, so dass

$$\|\omega_{i_1\dots i_k}(t,s_1,\dots,s_k)\| \leq KM^k \tag{4.15}$$

für alle $0 \leq s_k \leq \dots \leq s_1 \leq t \leq T$, $k \geq 0$ und allen Multiindizes $(i_1 \dots i_k)$ mit Elementen $i_1,\dots,i_k$ aus der Menge $\{1,\dots,m\}$.
Setze dazu

$$\begin{aligned} K_A &:= \max_{t\in[0,T]} \|e^{At}\| \ , \\ K_N &:= \max\{1,\|N_1\|,\dots,\|N_m\|\} \ , \\ K_B &:= \max\left\{\|b^1\|,\dots,\|b^m\|\right\} \ . \end{aligned}$$

Es gilt:

$$\begin{aligned} \|\omega_{i_1\dots i_k}(t,s^k)\| =& \|Ce^{A(t-s_1)}N_{i_1}e^{A(s_1-s_2)}\dots N_{i_{k-1}}e^{A(s_{k-1}-s_k)}(N_{i_k}e^{As_k}x_0+b^{i_k})\| \\ \leq& \|C\|\cdot\|e^{A(t-s_1)}\|\cdot\|N_{i_1}\|\cdot\|e^{A(s_1-s_2)}\|\dots\|N_{i_{k-1}}\|\cdot\|e^{A(s_{k-1}-s_k)}\|\cdot \\ & \cdot(\|N_{i_k}\|\cdot\|e^{As_k}\|\cdot\|x_0\|+\|b^{i_k}\|) \\ \leq& \|C\|{K_A}^k{K_N}^{k-1}(K_NK_A\|x_0\|+K_B) \\ =& {K_A}^k{K_N}^k(\|C\|K_A\|x_0\|+\|C\|K_BK_N^{-1}) \end{aligned}$$

Wähle nun $K := \|C\|K_A\|x_0\| + \|C\|K_BK_N^{-1}$ und $M := K_AK_N$, um die Bedingung (4.15) zu erfüllen.
Jetzt können wir das Integral abschätzen:

$$\begin{aligned} &\left\| \int_0^t\int_0^{s_1}\dots\int_0^{s_{k-1}} \omega_{i_1\dots i_k}(t,s^k)u_{i_1}(s_1)\dots u_{i_k}(s_k)\mathrm{ds_k}\dots\mathrm{ds_1}\right\| \\ &\leq \int_0^t\int_0^{s_1}\dots\int_0^{s_{k-1}} \|\omega_{i_1\dots i_k}(t,s^k)\|\cdot|u_{i_1}(s_1)|\cdots|u_{i_k}(s_k)|\mathrm{ds_k}\dots\mathrm{ds_1} \\ &\leq \|u_{i_1}\|_\infty\cdots\|u_{i_k}\|_\infty KM^k\frac{t^k}{k!} \\ &\leq K\frac{(MT\|u\|_\infty)^k}{k!} \qquad \forall t\in[0,T] \end{aligned} \tag{4.16}$$

Insgesamt gilt

$$\begin{aligned}&\big\|\omega_0(t)\big\|_0 + \\ &\sum_{k=1}^{\infty}\Big\| \sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \omega_{i_1\dots i_k}(t,s^k) u_{i_1}(s_1)\dots u_{i_k}(s_k) ds^k \Big\|_0 \\ &\le K \sum_{k=0}^{\infty} \frac{(mMT\|u\|_\infty)^k}{k!} \le K \exp(mMT\|u\|_\infty) < \infty \ .\end{aligned}$$

Die absolute Konvergenz ist bewiesen. □

Bemerkung 4.10. Es gibt eine weitere Möglichkeit, die Volterra-Reihen-Repräsentation von Ψ herzuleiten. Die Vorgehensweise ist sehr elegant bei homogenen Systemen und läßt sich problemlos auf nichtautonome Systeme anwenden. Betrachte etwa das homogene bilineare System in $\mathbb{R}^n$ mit Ausgang

$$\begin{aligned}&\dot{x}(t) = A(t)x(t) + \sum_{i=1}^{m} u_i(t) N_i(t) x(t) \ , \quad u(\cdot) \in L^\infty([0,T],\mathbb{R}^m) \ , \quad x(t) \in \mathbb{R}^n \ , \\ &y(t) = Cx(t) \ , \quad y(t) \in \mathbb{R}^q \ , \quad t \in [t_0, T] \ , \\ &x(t_0) = x_0 \ ,\end{aligned}$$

gegeben durch messbare, (essentielle) beschränkte Funktionen $A, N_1, \dots, N_m : [0,T] \to \mathbb{R}^{n\times n}$ und einer $n \times n$ Matrix C.
Mit $X(\cdot) := \Phi(\cdot\,; t_0, u)$ bezeichnen wir die Fundamentallösung des assoziierten Matrixsystems zur Kontrolle u. Durch die Substituition

$$Z(t) := \Phi(t_0; t, 0) X(t)$$

erhalten wir ein „äquivalentes“ Matrixsystem, in dessen Systemgleichung die Matrix $A(t)$ eliminiert ist. Ähnlich wie in (2.22), ist dieses System folgendermaßen bestimmt:

$$\dot{Z}(t) = \underbrace{\left(\sum_{k=1}^{m} u_k(t) \Phi(t_0; t, 0) N_k(t) \Phi(t; t_0, 0) \right)}_{:= U(t)} Z(t) = U(t) Z(t) \ .$$

(Diese Gleichung erhält man direkt durch Ableiten mittels Produktregel unter Berücksichtigung der Regeln aus Folgerung 2.21.)
Zur Vereinfachung setzen wir $F_k(t) := \Phi(t_0; t, 0) N_k(t) \Phi(t; t_0, 0)$. Wir können

nun $Z(t)$ in eine Neumann-Reihe entwickeln. Dies ergibt

$$Z(t) = \mathrm{I} + \sum_{k=1}^{\infty} \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \dots \int_{t_0}^{s_{k-1}} U(s_1)U(s_2)\dots U(s_k)\mathrm{ds_k}\dots\mathrm{ds_2ds_1} =$$
$$\mathrm{I} + \sum_{k=1}^{\infty} \sum_{i_1,\dots,i_k=1}^{m} \int_{t_0}^{t} \int_{t_0}^{s_1} \dots \int_{t_0}^{s_{k-1}} F_{i_1}(s_1)\dots F_{i_k}(s_k)\; u_{i_1}(s_1)\dots u_{i_k}(s_k)\mathrm{ds}^k \ .$$

Wegen $y(t) = Cx(t) = CX(t)x_0 = C\Phi(t;t_0,0)Z(t)x_0$, folgt

$$y(t) = C\Phi(t;t_0,0)x_0 + \sum_{k=1}^{\infty} \sum_{i_1,\dots,i_k=1}^{m} \int_{t_0}^{t} \int_{t_0}^{s_1} \dots \int_{t_0}^{s_{k-1}} \omega_{i_1\dots i_k}(t,s_1,\dots,s_k)\cdot$$
$$\cdot u_{i_1}(s_1)u_{i_2}(s_2)\dots u_{i_k}(s_k)\mathrm{ds}^k$$

mit den Kernen

$$\begin{aligned}\omega_{i_1\dots i_k}(t,s^k) = C\ &\cdot\Phi(t;s_1,0)N_{i_1}(s_1)\cdot\\ &\cdot\Phi(s_1;s_2,0)N_{i_2}(s_2)\cdot\\ &\cdot\Phi(s_2;s_3,0)N_{i_3}(s_3)\cdot\\ &\vdots\\ &\cdot\Phi(s_{k-1};s_k,0)N_{i_k}(s_k)\cdot\ \Phi(s_k;t_0,0)x_0 \qquad \forall\, k \geq 1\ .\end{aligned}$$

Indem wir $\Phi(s_{k-1};s_k,0) := e^{A(s_{k-1}-s_k)}$ und $t_0 := 0$ setzen, gewinnen wir genau die Volterra-Reihe aus Satz 4.9—angewandt auf das homogene System (AWP) mit $B = 0$.
Mit Variation der Konstanten könnten wir auch die Volterra-Reihen-Repräsentation bezüglich inhomogener bilinearer Systeme der Form

$$\begin{aligned}&\dot{x}(t) = A(t)x(t) + \sum_{i=1}^{m} u_i(t)N_i(t)x(t) + B(t)u(t)\ , \quad u(t) \in \mathbb{R}^m\ ,\ x(t) \in \mathbb{R}^n\ ,\\ &y(t) = Cx(t)\ , \quad y(t) \in \mathbb{R}^q\ , \quad t \in [t_0,T]\ ,\\ &x(t_0) = x_0\ ,\end{aligned}$$

mit zusätzlicher messbarer und (essentiell) beschränkter Funktion $B : [0,T] \to \mathbb{R}^{n\times m}$, herleiten. Dies wäre allerdings sehr mühsam, weswegen nur das Er-

gebnis in Form der zugehörigen Volterra-Kerne angegeben wird:

$$\begin{aligned}
\omega_0(t) &= C\Phi(t;t_0,0)x_0 \ , \\
\omega_{i_1}(t,s_1) &= C\Phi(t;s_1,0)\left(b^{i_1}(s_1)+N_{i_1}(s_1)\Phi(s_1;t_0,0)x_0\right) \ , \\
\omega_{i_1\dots i_k}(t,s^k) &= C\Phi(t;s_1,0)\cdot N_{i_1}(s_1)\Phi(s_1;s_2,0)\cdot \\
&\qquad \cdot N_{i_2}(s_2)\Phi(s_2;s_3,0)\cdot \\
&\qquad \cdot N_{i_3}(s_3)\Phi(s_3;s_4,0)\cdot \\
&\qquad \vdots \\
&\qquad \cdot N_{i_{k-1}}(s_{k-1})\Phi(s_{k-1};s_k,0)\cdot \\
&\qquad\qquad \cdot\left(b^{i_k}(s_k)+N_{i_k}(s_k)\Phi(s_k;t_0,0)x_0\right),
\end{aligned}$$

wobei $b^{i_k}(t) = B(t)\cdot \mathrm{e}_{i_k}$ und $k \geq 2$.

Beispiel 4.11. Zur Übung berechnen wir die Volterra-Reihen-Repräsentation des Oszillator-Modells (leicht modifiziert) aus Beispiel 2.15

$$\begin{pmatrix}\dot{x}_1\\ \dot{x}_2\end{pmatrix} = \begin{pmatrix}0 & 1\\ -\omega^2 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} + u\begin{pmatrix}0 & 0\\ -1 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \ ,$$

$$y = \begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}, \quad u(t)\in\mathbb{R}, \ \ \mathcal{X}=\mathbb{R}^2$$

bezüglich der Anfangsbedingung $x(0) = \left(\begin{smallmatrix}0\\ \omega\end{smallmatrix}\right)$.
(Vorsicht: Hier bezeichnet ω ausnahmsweise eine physikalische Konstante.)
MAPLE berechnet die Matrixexponentialfunktion

$$e^{At} = \sum_{k=0}^{\infty}\frac{t^k}{k!}\begin{pmatrix}0 & 1\\ -\omega^2 & 0\end{pmatrix}^k = \begin{pmatrix}\cos(\omega t) & \frac{1}{\omega}\sin(\omega t)\\ -\omega\sin(\omega t) & \cos(\omega t)\end{pmatrix} \ .$$

Wir wählen in (4.14) $C = \mathrm{e}_1^*$, $b^{i_k} = 0$, $N_1 = (-\mathrm{e}_2)\mathrm{e}_1^*$ und $x_0 = \mathrm{e}_2\omega$, um die Kerne $\tilde{\omega}_0$ und $\tilde{\omega}_{i_1\dots i_k}$ zu bestimmen. (Dabei ist e_k der k-te kanonische Einheitsvektor des $\mathbb{R}^2$.) Es folgt

$$\begin{aligned}
\tilde{\omega}_0(t) &= Ce^{At}x_0 = \mathrm{e}_1^*e^{At}\mathrm{e}_2\omega = \sin(\omega t) \ , \\
\tilde{\omega}_{1,1,\dots,1}(t,s^k) &= \mathrm{e}_1^*e^{A(t-s_1)}(-\mathrm{e}_2)\cdot \mathrm{e}_1^*e^{A(s_1-s_2)}(-\mathrm{e}_2)\cdot \\
&\qquad \cdot \mathrm{e}_1^*e^{A(s_2-s_3)}(-\mathrm{e}_2)\cdot \\
&\qquad \vdots \\
&\qquad \cdot \mathrm{e}_1^*e^{A(s_{k-1}-s_k)}(-\mathrm{e}_2)\cdot \mathrm{e}_1^*e^{As_k}\mathrm{e}_2\omega \\
&= \frac{(-1)^k}{\omega^k}\sin\left(\omega(t-s_1)\right)\left(\prod_{l=2}^{k}\sin\left(\omega(s_{l-1}-s_l)\right)\right)\sin(\omega s_k), \ k\geq 1, \\
\tilde{\omega}_{i_1\dots i_k}(t,s^k) &= 0 \ , \qquad \text{sonst} \ .
\end{aligned}$$

Schließlich ergibt sich:

$$\Psi(u)(t) = \sin(\omega t) + \\ + \sum_{k=1}^{\infty} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \frac{(-1)^k}{\omega^k} \sin\left(\omega(t-s_1)\right) \left(\prod_{l=2}^{k} \sin\left(\omega(s_{l-1}-s_l)\right) \right) \cdot \\ \cdot \sin(\omega s_k) u(s_1) u(s_2) \dots u(s_k) \mathrm{ds_k} \dots \mathrm{ds_2} \mathrm{ds_1}$$

4.3 Eindeutigkeit der Volterra-Reihen-Repräsentation

Indem wir die Volterra-Reihe aus Definition 4.2 abbrechen, können wir die Ein-Ausgangsfunktion Ψ approximieren. Der Fehlerterm dieser Approximation läßt sich mit Hilfe des Landau-Symbols beschreiben.

Bezeichnung 4.12. Es seien Ψ_1,Ψ_2 zwei Abbildungen von $L^\infty([0,T],\mathbb{R}^m)$ nach $C^0([0,T],\mathbb{R}^q)$ und $u, u_0 \in L^\infty([0,T],\mathbb{R}^m)$. Die Notation

$$\Psi_1(u) = \Psi_2(u) + o(\|u\|_\infty^l) \text{ für } u \to u_0$$

bedeutet, dass für alle $\epsilon > 0$ ein $\delta > 0$ existiert, so dass

$$\|\Psi_1(u) - \Psi_2(u)\|_0 < \epsilon \|u\|_\infty^l$$

für alle u mit $\|u - u_0\|_\infty < \delta$.

Definition 4.13. Man sagt, die Ein-Ausgangsfunktion

$$\Psi : L^\infty([0,T],\mathbb{R}^m) \to C^0([0,T],\mathbb{R}^q)$$

hat eine *Volterra-Entwicklung der Länge l*, falls es zu jedem Multiindex $(i_1 \dots i_k)$ der Länge $k \in \{1,\dots,l\}$, mit Elementen $i_1,\dots,i_k$ aus der Menge $\{1,\dots,m\}$, eine Funktion $\omega_{i_1\dots i_k}$ gibt, so dass alle Funktionen $\omega_{i_1\dots i_k}$ zusammen mit einer weiteren Funktion ω_0 die folgenden Bedingungen erfüllen:

1. ω_0 ist eine stetige Funktion der Form $\omega_0 : [0,T] \to \mathbb{R}^q$.

2. Jedes $\omega_{i_1\dots i_k}$ ist eine stetige Funktion der Form $\omega_{i_1\dots i_k} : D_k \to \mathbb{R}^q$, definiert über der Menge

$$D_k := \{(t, s_1, \dots, s_k) \mid 0 \le s_k \le \dots \le s_1 \le t \le T\} \ .$$

3. Es ist

$$\Psi(u)(t) = \omega_0(t) + \sum_{k=1}^{l} \sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \omega_{i_1\dots i_k}(t, s^k) u_{i_1}(s_1) \dots u_{i_k}(s_k) \mathrm{d}s^k + o(\|u\|_\infty^l) \tag{4.17}$$

für $u \to 0$. (Dabei ist $\int_0^{s_0} := \int_0^t$.)

Satz 4.14. Die Ein-Ausgangsfunktion $\Psi : L^\infty([0,T], \mathbb{R}^m) \to C^0([0,T], \mathbb{R}^q)$ eines Systems Σ besitze eine Volterra-Reihen-Repräsentation. Angenommen, es existieren reelle Zahlen $K \geq 0$ und $M > 0$, so dass die Volterra-Kerne dieser Repräsentation auf ihrem Definitionsbereich der Wachstumsbedingung

$$\|\omega_0(t)\| \leq K, \qquad \|\omega_{i_1\dots i_k}(t, s^k)\| \leq k! K M^k \tag{4.18}$$

für alle $k \in \mathbb{N}$ und allen Multiindizes genügen.
Dann hat Ψ eine Volterra-Reihen Entwicklung beliebiger Länge $l \geq 0$, welche durch die l-te Partialsumme gegeben ist.

Beweis. Wir übernehmen die Bezeichnungen aus Definition 4.2.

1. Bricht man die Volterra-Reihe aus Definition 4.2 nach dem l-ten Summanden ab, so bleibt ein Restterm der Form

$$R_{l+1}(u)(t) = \sum_{k=l+1}^{\infty} \sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \omega_{i_1\dots i_k}(t, s^k) u_{i_1}(s_1) \dots u_{i_k}(s_k) \mathrm{d}s^k \ .$$

Um die Behauptung zu beweisen, zeigen wir, dass

$$R_{l+1}(u) = o(\|u\|_\infty^l) \ \text{ für } u \to 0 \ .$$

2. Wir führen für alle $k \in \mathbb{N}$ die Konstanten

$$C_0 := K \qquad \text{und} \qquad C_k := K(mTM)^k$$

ein. Analog zu (4.16) gewinnen wir aus der Wachstumsbedingung (4.18) die Abschätzungen

$$\max_{t\in[0,T]} \left(\sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \|\omega_{i_1\dots i_k}(t, s^k)\| \mathrm{ds_k ds_{k-1} \dots ds_1} \right) \leq C_k$$

und $\max_{t\in[0,T]} \|\omega_0(t)\| \leq C_0$ für alle $k \in \mathbb{N}$ und allen Multiindizes.

3. Wie jede geometrische Reihe hat auch die Potenzreihe $\sum_{k=0}^{\infty} C_k z^k$ einen positiven Konvergenzradius. Insbesondere ist

$$C := \sum_{k=l+1}^{\infty} C_k r^{k-(l+1)} < \infty \quad \text{für ein } r > 0.$$

Es folgt für alle $t \in [0,T]$ und $u \in L^\infty([0,T],\mathbb{R}^m)$ mit $\|u\|_\infty \le r$:

$$\begin{aligned}\|R_{l+1}(u)(t)\| &\le \sum_{k=l+1}^{\infty} \left(\sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \|\omega_{i_1\dots i_k}(t,s^k)\| \mathrm{d}s^k \cdot \|u\|_\infty^k \right) \\ &\le \sum_{k=l+1}^{\infty} C_k \|u\|_\infty^{l+1} r^{k-(l+1)} = C \cdot \|u\|_\infty^{l+1}\end{aligned}$$

Dies impliziert

$$\lim_{\substack{\|u\|\to 0 \\ u\neq 0}} \frac{\|R_{l+1}(u)\|_0}{\|u\|_\infty^l} = 0$$

und somit die Behauptung $R_{l+1}(u) = o(\|u\|_\infty^l)$ für $u \to 0$. □

Bemerkung 4.15. Wir können aus der Wachstumsbedingung (4.18) und der Stetigkeit der Kerne die absolute Konvergenz einer Volterra-Reihe in $C^0([0,T],\mathbb{R}^q)$ ableiten, falls die Kontrollen u genügend klein sind. Wie wir im letzten Beweis gesehen haben, wird nämlich die Reihe der Normen durch eine geometrische Reihe majorisiert. Dass die Kerne unseres bilinearen Systems (AWP) die Bedingung (4.18) erfüllen, wird im Beweis von Satz 4.9 gezeigt. Die Tatsache, dass der Restterm solcher Systeme von Ordnung $o(\|u\|_\infty^l)$ für $u \to 0$ ist, wird die Grundlage bei der Analyse der Volterra-Reihen-Repräsentation hinsichtlich der Eindeutigkeit seiner Kerne bilden.

Satz 4.16 (Eindeutigkeit)**.** Es seien Ψ_1,Ψ_2 zwei Ein-Ausgangsfunktionen der Form $\Psi : L^\infty([0,T],\mathbb{R}^m) \to C^0([0,T],\mathbb{R}^q)$, welche eine Volterra-Entwicklung der Länge l besitzen. Ist

$$\Psi_1(u) = \Psi_2(u) + o(\|u\|_\infty^l) \qquad \text{für } u \to 0\ , \tag{4.19}$$

so stimmen die Kerne $\omega_{i_1\dots i_k}(t,s^k)$ der Volterra-Entwicklung von Ψ_1 und die Kerne $\tilde{\omega}_{i_1\dots i_k}(t,s^k)$ der Volterra-Entwicklung von Ψ_2 überein, d.h

$$\omega_0 = \tilde{\omega}_0 \qquad \text{und} \qquad \omega_{i_1\dots i_k} = \tilde{\omega}_{i_1\dots i_k} \qquad \forall 1 \le k \le l\ .$$

(Zur Vereinfachung, werde durch $(i_1 \dots i_k)$ ein beliebiger Multiindex der Länge k mit Elementen in $\{1,\dots,m\}$ repräsentiert.)

Beweis. 1. Wähle die Steuerung $u \equiv 0$, um $\omega_0(t) = \tilde{\omega}_0(t)$ zu zeigen.

2. Als nächstes zeigen wir

$$\int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} [\omega_{i_1\dots i_k}(t, s^k) - \tilde{\omega}_{i_1\dots i_k}(t, s^k)] u_{i_1}(s_1) \dots u_{i_k}(s_k) \mathrm{d}s^k = 0$$
$$\forall u \in L^\infty([0,T], \mathbb{R}^m), \quad \forall 1 \le k \le l \ . \quad (4.20)$$

Angenommen, dies ist falsch. Dann bezeiche $r \in \{1, \dots, l\}$ den kleinstmöglichen Index, so dass

$$\int_0^t \int_0^{s_1} \dots \int_0^{s_{r-1}} [\omega_{i_1\dots i_r}(t, s^r) - \tilde{\omega}_{i_1\dots i_r}(t, s^r)] u_{i_1}(s_1) \dots u_{i_r}(s_r) \mathrm{d}s^r = z \neq 0$$

für eine Kontrolle u.
Setze $\nu_n := \frac{1}{n} u$ für eine natürliche Zahl n und wähle ein beliebiges $\epsilon > 0$. Nach Voraussetzung (4.19) und der Minimalitätseigenschaft von r gilt bzgl. der Kontrollfunktion ν_n

$$\Big\| \sum_{k=r}^{l} \big(\tfrac{1}{n}\big)^k \sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} [\omega_{i_1\dots i_k}(t, s^k) - \tilde{\omega}_{i_1\dots i_k}(t, s^k)] \cdot$$
$$\cdot u_{i_1}(s_1) \dots u_{i_k}(s_k) \mathrm{d}s^k \Big\|$$
$$< \epsilon (\tfrac{1}{n} \|u\|_\infty)^l$$

für alle $n \in \mathbb{N}$ genügend groß. Multiplikation mit n^r ergibt:

$$\Big\| z + \sum_{k=r+1}^{l} n^{r-k} \dots \Big\| < \epsilon n^{r-l} \|u\|_\infty^l$$

Die linke Seite der Ungleichung konvergiert gegen $\|z\| > 0$ für $n \to \infty$, während die rechte beliebig klein gemacht werden kann (ϵ frei wählbar). Das ist ein Widerspruch. Also existiert kein solches r.

3. Um den Beweis abzuschließen, zeigen wir, dass (4.20) die Gleichung $\omega_{i_1\dots i_k}(t, s^k) = \tilde{\omega}_{i_1\dots i_k}(t, s^k)$ impliziert für $1 \le k \le l$. Exemplarisch wird nur der Fall $k = 2$ vorgeführt. Der allgemeine Fall wird analog bewiesen. Sei jetzt $a \in (0, T]$. Wir definieren die Eingabefunktion u komponentenweise

$$u_i(s) := \begin{cases} n, & \text{falls } a - \frac{1}{n} \le s \le a \\ 0, & \text{sonst} \end{cases}$$

$(i = 1, \dots m)$ und setzen sie in (4.20) ein. Wir stellen fest, indem wir a und n variieren, dass

$$\int_{a-\frac{1}{n}}^{a} \int_{a-\frac{1}{n}}^{s_1} [\omega_{i_1,i_2}(t, s_1, s_2) - \tilde{\omega}_{i_1,i_2}(t, s_1, s_2)] n^2 \mathrm{ds}_2 \mathrm{ds}_1 = 0$$

für alle $0 < a \leq t \leq T$ und alle $n \in \mathbb{N}$. Folglich gilt unter Ausnutzung der Substituitionsregel und des Satzes der majorisierten Konvergenz

$$\begin{aligned}
&\lim_{n\to\infty} n^2 \int_{a-\frac{1}{n}}^{a} \int_{a-\frac{1}{n}}^{s_1} [\omega_{i_1,i_2}(t, s_1, s_2) - \tilde{\omega}_{i_1,i_2}(t, s_1, s_2)] \, \mathrm{ds}_2 \mathrm{ds}_1 = 0 \\
&\Rightarrow \lim_{n\to\infty} n^2 \int_{-\frac{1}{n}}^{0} \int_{-\frac{1}{n}}^{s_1} [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, s_1 + a, s_2 + a) \, \mathrm{ds}_2 \mathrm{ds}_1 = 0 \\
&\Rightarrow \lim_{n\to\infty} \int_{-1}^{0} \int_{-1}^{s_1} [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, \frac{s_1}{n} + a, \frac{s_2}{n} + a) \, \mathrm{ds}_2 \mathrm{ds}_1 = 0 \\
&\Rightarrow \int_{-1}^{0} \int_{-1}^{s_1} \lim_{n\to\infty} [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, \frac{s_1}{n} + a, \frac{s_2}{n} + a) \, \mathrm{ds}_2 \mathrm{ds}_1 = 0 \\
&\Rightarrow [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, a, a) \int_{0}^{1} \int_{0}^{s_1} 1 \, \mathrm{ds}_2 \mathrm{ds}_1 = \frac{1}{2!} [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, a, a) = 0 \\
&\Rightarrow \omega_{i_1,i_2}(t, a, a) - \tilde{\omega}_{i_1,i_2}(t, a, a) = 0 \; .
\end{aligned}$$

Da $\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}$ stetig ist, ist auch $\omega_{i_1,i_2}(t, 0, 0) - \tilde{\omega}_{i_1,i_2}(t, 0, 0)$. Also

$$\omega_{i_1,i_2}(t, a, a) - \tilde{\omega}_{i_1,i_2}(t, a, a) = 0 \qquad \forall 0 \leq a \leq t \leq T \; .$$

4. Nun verwenden wir in (4.20) die Kontrolle

$$u_i(s) := \begin{cases} n, & \text{falls } a - \frac{1}{n} \leq s \leq a \\ n, & \text{falls } b - \frac{1}{n} \leq s \leq b \\ 0, & \text{sonst} \end{cases}$$

für $i = 1, \dots, m$, wobei $0 < b < a < t < T$. Wir lassen wieder $n \to \infty$ gehen und erhalten diesmal

$$\begin{aligned}
0 &= \lim_{n\to\infty} \big(n^2 \int_{a-\frac{1}{n}}^{a} \int_{a-\frac{1}{n}}^{s_1} \dots + n^2 \int_{b-\frac{1}{n}}^{b} \int_{b-\frac{1}{n}}^{s_1} \dots + n^2 \int_{a-\frac{1}{n}}^{a} \int_{b-\frac{1}{n}}^{b} \dots \big) \\
&= \tfrac{1}{2} [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, a, a) + \tfrac{1}{2} [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, b, b) + [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, a, b) \\
&\overset{3.}{=} [\omega_{i_1,i_2} - \tilde{\omega}_{i_1,i_2}](t, a, b) \; .
\end{aligned}$$

Aus Stetigkeitsgründen folgt

$$\omega_{i_1,i_2}(t, a, b) - \tilde{\omega}_{i_1,i_2}(t, a, b) = 0 \qquad \forall 0 \leq b \leq a \leq t \leq T \; .$$

Die Kerne stimmen somit auf ihrem Definitionsbereich überein, was zu zeigen war. □

Folgerung 4.17. Angenommen, wir fordern, dass die Kerne einer Volterra-Reihen-Repräsentation die Wachstumsbedingung (4.18) erfüllen müssen. Dann ist Volterra-Reihen-Repräsentation der Ein-Ausgangsfunktion

$$\Psi : L^\infty([0,T],\mathbb{R}^m) \to C^0([0,T],\mathbb{R}^q)$$

eines Systems Σ eindeutig, sofern existent. Insbesondere gilt dies für unser bilineares System (AWP).

Beweis. Nach Satz 4.14 und Voraussetzung hat jede Ein-Ausgangsfunktion Ψ, die eine Volterra-Reihen-Repräsentation mit Kernen $\omega_{i_1\dots i_k}$ besitzt, eine Volterra-Reihen Entwicklung beliebiger Länge l, gegeben durch die l-te Partialsumme ihrer Volterra-Reihe. Indem wir die Volterra-Reihe einer—möglicherweise anderen—Repräsentation von Ψ nach dem l-ten Summanden abrechen, erhalten wir eine weitere Ein-Ausgangsfunktion Ψ_2, die

$$\Psi(u) = \Psi_2(u) + o(\|u\|_\infty^l) \qquad \text{für } u \to 0$$

erfüllt. Die Eindeutigkeit gemäß Satz 4.16 impliziert, dass die Kerne $\omega_{i_1\dots i_k}$ für $k \le l$ eindeutig bestimmt sind. Da l beliebig wählbar, folgt die Behauptung. □

4.4 Endliche Volterra-Entwicklung

Definition 4.18. Wir sagen, dass die Ein-Ausgangsfunktion Ψ eine *endliche Volterra-Entwicklung* der Länge l besitzt, wenn sie eine Volterra-Entwicklung der Länge l ohne Restterm $o(\|u\|_\infty^l)$ hat.

Häufig teilt man bilineare Systeme mit Ausgang in zwei Klassen ein:

(i) Die Klasse der *schwachen* bilinearen Systeme, deren Ein-Ausgangsfunktion eine endliche Volterra-Entwicklung besitzt. Solche Systeme lassen sich als ein Gebilde endlich vieler linearer Teilsysteme auffassen, die wie in Satz 4.7 ineinander „verschachtelt“ sind.

(ii) Die Klasse der *starken* bilinearen Systeme, deren Ein-Ausgangsfunktion keine endliche Volterra-Entwicklung besitzt.

Wir wollen nun ein Kriterium für die Endlichkeit einer Volterra-Reihe herleiten. Dazu führen wir Begriffe aus der Algebra ein.

Definition 4.19. $\{A_1,\dots,A_m\}$ sei eine Menge reeller $n\times n$-Matrizen. Dann bezeichne $\{A_1,\dots,A_m\}_{\mathrm{AA}}$ die kleinste (assoziative) Unteralgebra von $\mathbb{R}^{n\times n}$, welche die Matrizen A_1 bis A_m enthält.
Eine assoziative Algebra heißt *nilpotent*, falls ein $l \in \mathbb{N}$ existiert, so dass jedes Produkt von l beliebigen Elementen dieser Algebra verschwindet.

Satz 4.20. Die Volterra-Reihe mit den Kernen aus (4.14) ist endlich, falls die assoziative Algebra

$$\{\mathrm{ad}_A^k N_j \mid 0 \le k \le n^2 - 1,\ 1 \le j \le m\}_{\mathrm{AA}}$$

nilpotent ist, wobei der Operator ad_A^k durch

$$\mathrm{ad}_A^0 N = N,\ \mathrm{ad}_A^1 N = [A, N],\ \mathrm{ad}_A^2 N = [A, [A, N]],\ \mathrm{ad}_A^k N = [A, \mathrm{ad}_A^{k-1} N]$$

gegeben ist. Insbesondere hat dann die Ein-Ausgangsfunktion Ψ des bilinearen Systems (AWP) eine endliche Volterra-Entwicklung.

Beweis. Die Kerne der Volterra-Reihe sind von der Form

$$\begin{aligned}\omega_{i_1 \dots i_{k+1}}(t, s^{k+1}) = Ce^{At} &\cdot e^{-As_1} N_{i_1} e^{As_1} \cdot \\ &\cdot e^{-As_2} N_{i_2} e^{As_2} \cdot \\ &\cdot e^{-As_3} N_{i_3} e^{As_3} \cdot \\ &\vdots \\ &\cdot e^{-As_k} N_{i_k} e^{As_k} \cdot e^{-As_{k+1}} (b^{i_{k+1}} + N_{i_{k+1}} e^{As_{k+1}} x_0)\ .\end{aligned}$$

Anhand der Baker-Hausdorff-Formel (3.19) analysieren wir die Struktur. Es gilt

$$e^{-At} N_j e^{At} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{ad}_A^k N_j \qquad \forall t \in \mathbb{R},\ j = 1, \dots, m\ .$$

Mit den Argumenten aus dem Beweis von Folgerung 3.32 (Stichworte: Cayley-Hamilton, Abgeschlossenheit) kommen wir zum Schluss

$$e^{-At} N_j e^{At} \in \{\mathrm{ad}_A^k N_j \mid 0 \le k \le n^2 - 1,\ 1 \le j \le m\}_{\mathrm{AA}}\ .$$

Der oben angegebene Kern $\omega_{i_1 \dots i_{k+1}}$ enthält also ein Produkt von k Elementen aus der Algebra $\{\mathrm{ad}_A^k N_j \mid 0 \le k \le n^2 - 1,\ 1 \le j \le m\}_{\mathrm{AA}}$. Ist die Algebra nilpotent, so verschwinden folglich die Kerne höherer Ordnung und die Volterra-Reihe ist endlich. □

Das folgende Beispiel stammt aus [26, p.344].

Beispiel 4.21. Betrachte das System im $\mathbb{R}^5$:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{pmatrix} = \begin{pmatrix} u_1 \\ x_1 \\ u_2 \\ x_3 \\ x_1 u_2 \end{pmatrix}, \qquad x(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \in E^2$$

In der Notation $\dot{x} = Ax + u_1N_1x + u_2N_2x + Bu$ ist $A = E_{21} + E_{43}$, $N_1 = 0$, $N_2 = E_{51}$, $b^1 = (1,0,0,0,0)$ und $b^2 = (0,0,1,0,0)$. Als Nächstes werden die Erzeuger der Algebra $\{\mathrm{ad}_A^k N_j \mid 0 \le k \le 5^2 - 1,\ 1 \le j \le 2\}_{\mathrm{AA}}$ berechnet:

$$\mathrm{ad}_A^0 N_2 = N_2 = E_{51}\ , \qquad \mathrm{ad}_A^1 N_2 = AN_2 - N_2A = 0$$

Es bleibt $\{E_{51}\}_{\mathrm{AA}}$. Offenbar ist die Matrix E_{51} nilpotent und somit auch unsere Algebra. Jedes Produkt von zwei Elementen dieser Algebra ist gleich Null. Wie im Beweis von Satz 4.20 verschwinden alle Kerne $\omega_{i_1\dots i_k}$ mit $k \ge 3$. Die zugehörige Volterra-Reihe ist endlich. Natürlich hätte man dies auch aus der Tatsache, dass A,N_1 und N_2 stikte untere Dreiecksmatrizen sind, folgern können.

Mit MAPLE können wir die Volterra-Reihe bestimmen, indem wir die Kerne von Ordnung kleiner 3 berechnen:

$$x(t) = \int_0^t \begin{pmatrix} 1 \\ t-s \\ 0 \\ 0 \\ 0 \end{pmatrix} u_1(s)\mathrm{ds} + \int_0^t \begin{pmatrix} 0 \\ 0 \\ 1 \\ t-s \\ 0 \end{pmatrix} u_2(s)\mathrm{ds} + \int_0^t \int_0^{s_1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} u_2(s_1)u_1(s_2)\mathrm{ds}_2\mathrm{ds}_1$$

Folgerung 4.22. Angenommen, es sind entweder alle Matrizen aus der Menge $\{A, N_1, \dots, N_m\}$ strikte obere Dreiecksmatrizen oder alle Matrizen dieser Menge sind strikte untere Dreiecksmatrizen. Dann ist die Algebra $\{\cdot\}_{\mathrm{AA}}$ trivialerweise nilpotent, und die Ein-Ausgangsfunktion Ψ von (AWP) besitzt eine endliche Volterra-Entwicklung.

Bemerkung 4.23. Vergleichbar mit der Kalman-Zerlegung aus der linearen Kontrolltheorie, gibt es auch für autonome bilineaere Systeme der Form (AWP) eine kanonische Zerlegung des Zustandsraums in die direkte Summe von bestimmten Unterräumen, welche in [21, p.30] definiert werden. Wir können die Vereinigung der Basen dieser Unterräume als die neue Basis des Zustandsraums wählen, und erhalten durch diesen Koordinatenwechsel ein äquivalentes blineares System, das durch die transformierten Matrizen TAT^{-1}, TN_jT^{-1}, TB und CT^{-1} für ein $T \in \mathrm{GL}(n;\mathbb{R})$ gegeben ist $(j = 1, \dots m)$. Der Clou ist, dass TAT^{-1} und TN_jT^{-1} $(j = 1, \dots m)$ obere Dreiecksmatrizen sind. Sind es sogar strikte obere Dreiecksmatrizen, so ist laut Folgerung 4.22 die Volterra-Entwicklung des (un-)transformierten Systems endlich.

Beispiel 4.24. Die Matrizen $A = E_{34}$ und $N = E_{12} + E_{23}$ in Sussmanns Beispiel 3.34 sind strikte obere Dreiecksmatrizen. Offenbar sind Produkte von drei oder mehr Elementen der Algebra $\{\dots\}_{\mathrm{AA}}$ gleich Null. Es müssen

also höchstens die Kerne $\omega_{i_1 \dots i_k}$ mit $k \leq 3$ berechnet werden.
Die Volterra-Entwicklung der Fundamentallösung ist

$$\Phi(t;u) = \begin{pmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&t\\0&0&0&1 \end{pmatrix} + \int_0^t \begin{pmatrix} 0&1&0&0\\0&0&1&s\\0&0&0&0\\0&0&0&0 \end{pmatrix} u(s)\mathrm{d}s + \int_0^t \int_0^{s_1} \begin{pmatrix} 0&0&1&s_2\\0&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix} u(s_1)u(s_2)\mathrm{d}s^2.$$

4.5 Anwendungen

In diesem Abschnitt wollen wir kurz einige Anwendungen angeben, welche die Volterra-Reihen-Repräsentation von bilinearen Systemen der Form[2]

$$\begin{aligned} \dot{x}(t) &= Ax(t) + \sum_{i=1}^{m} N_i x(t) u_i(t) + Bu(t), \qquad t \in [0,T]\ , \qquad \text{(AWP)}\\ x(0) &= x_0, \quad \mathcal{X} = \mathbb{R}^n,\ \mathcal{U} = \mathbb{R}^m \end{aligned}$$

nutzen.

Numerische Approximation der Lösungstrajektorien bilinearer Systeme

Die Volterra-Entwicklung einer zulässigen Lösung von (AWP) ist eine naheliegende Alternative zur diskreten Approximation solcher Lösungen durch numerische Ein- oder Mehrschrittverfahren (z.B. Runge-Kutta-Verfahren). Indem man die Länge der Volterra-Entwicklung geeignet groß wählt, kann eine beliebig gute Approximation bestimmt werden. Ist die Volterra-Entwicklung endlich, so kann man aus der Kenntnis der Volterra-Kerne sogar die exakte Lösungstrajektorie für jede zulässige Kontrolle berechnen.

Folgerung 4.25. Es seien $\epsilon > 0$ und $\delta > 0$ vorgegeben. Die Konstanten K, M werden aus dem Beweis von Satz 4.9 übernommen. Wir wählen ein $l \in \mathbb{N}$, so dass $\rho_l(mMT\delta) \leq \frac{\epsilon}{K}$, wobei $\rho_l(s) := \sum_{j=l+1}^{\infty} \frac{s^j}{j!}$ die Restfunktion der Exponentialreihe ist, und berechnen die Kerne $\omega_0(t)$ und $\omega_{i_1 \dots i_k}(t, s^k)$ aus Satz 4.9 für alle $1 \leq k \leq l$. Dann existiert für jede Lösung $x(\cdot)$ von (AWP), bezüglich einer messbaren Kontrollfunktion u mit $\|u\|_\infty \leq \delta$, eine Approximation

$$\tilde{x}(t) := \omega_0(t) + \sum_{k=1}^{l} \sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \omega_{i_1 \dots i_k}(t, s_1, \dots, s_k) u_{i_1}(s_1) \dots u_{i_k}(s_k) \mathrm{ds_k} \dots \mathrm{ds_1}\ ,$$

so dass $\|x(t) - \tilde{x}(t)\| \leq \epsilon$ für alle $t \in [0,T]$.

[2] Hier ist $C = \mathrm{I}$ die Matrix in der Ausgangsfunktion.

Beweis. Es folgt aus Abschätzung (4.16) für alle $t \in [0,T]$

$$\begin{aligned}
&\|x(t) - \tilde{x}(t)\| \\
&= \left\| \sum_{k=l+1}^{\infty} \sum_{i_1,\dots,i_k=1}^{m} \int_0^t \int_0^{s_1} \dots \int_0^{s_{k-1}} \omega_{i_1\dots i_k}(t, s^k) u_{i_1}(s_1) \dots u_{i_k}(s_k) \mathrm{d}s^k \right\| \\
&\leq K \sum_{k=l+1}^{\infty} \frac{(mMT\|u\|_\infty)^k}{k!} \leq K \cdot \rho_l(mMT\delta) \ ,
\end{aligned}$$

wobei die Kerne $\omega_{i_1\dots i_k}(t, s^k)$ wie in Satz 4.9 definiert sind. □

Bemerkung 4.26. Durch die Einführung einer Schrittweitensteuerung oder durch die Verringerung der oberen Schranken K, M aus Abschätzung (4.16) könnten wir eine numerische Approximation bestimmen, die mit Kernen niedrigerer Ordnung auskommt. Dies würde den Rechenaufwand erheblich reduzieren.

Beispiel 4.27. Wir wenden uns dem Single-Input System aus Beispiel (3.36) zu:

$$\begin{pmatrix} \dot{x}_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + u \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad 0 \leq u \leq 1, \quad x(0) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

mit den Parametern

$$A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right), \qquad N = \left(\begin{smallmatrix} 0 & 0 \\ -1 & 0 \end{smallmatrix}\right), \qquad B = 0, \qquad x_0 = \left(\begin{smallmatrix} -1 \\ 0 \end{smallmatrix}\right)$$

Eine Approximation auf dem Zeitintervall $[0, \frac{\pi}{2}]$ (d.h. $T = \frac{\pi}{2}$) wird angestrebt. Anhand einer kurzen Rechnung lassen sich schrittweise die Parameter K, M aus dem Beweis von Satz 4.9 bestimmen:

$$\begin{aligned}
&\|x_0\| = 1, \quad e^{At} = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right), \quad K_A := \max_{t\in[0,T]} \|e^{At}\| \approx 2.057 \ , \\
&K_N := \max\{1, \|N\|\} = 1, \quad K_B := 0 \ , \\
&K = K_A \cdot \|x_0\| \approx 2.057, \quad M = K_A K_N \approx 2.057
\end{aligned}$$

Wir wählen die Steuerbeschränkung $\delta = 1$ und die Fehlertoleranz $\epsilon = 0.03$. Die maximale Ordnung l der benötigten Kerne muß dann die Ungleichung

$$K \cdot \rho_l(mMT\delta) \leq \epsilon \overset{\text{approx.}}{\Longleftrightarrow} 2.057 \cdot \rho_l(3.23113) \leq 0.03$$

erfüllen. Für $l = 10$ erhalten wir beispielsweise auf der linken Seite den aufgerundeten Wert 0.028 . Wir berechnen demzufolge die Kerne $\omega_{i_1\dots i_k}(t, s^k)$

aus der Volterra-Reihen-Repräsentation für alle $1 \leq k \leq 10$. Dies ergibt

$$\begin{aligned}
\omega_0(t) &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} \ , \\
\omega_1(t, s_1) &= \begin{pmatrix} t - s_1 \\ 1 \end{pmatrix} \ , \\
\omega_{1\ldots1}(t, s^k) &= \begin{pmatrix} (-1)^{k+1}(t-s_1)(s_1-s_2)\ldots(s_{k-1}-s_k) \\ (-1)^{k+1}(s_1-s_2)(s_2-s_3)\ldots(s_{k-1}-s_k) \end{pmatrix} \qquad \forall\, k \geq 2 \ .
\end{aligned}$$

Eine geeignete Approximation zu einer vorgegebenen zulässigen Kontrolle u lautet

$$\begin{aligned}
\tilde{x}(t) &:= \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \int_0^t \begin{pmatrix} t-s \\ 1 \end{pmatrix} u(s)\, \mathrm{ds} + \\
&+ \sum_{k=2}^{10} \int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} \begin{pmatrix} (-1)^{k+1}(t-s_1)(s_1-s_2)\ldots(s_{k-1}-s_k) \\ (-1)^{k+1}(s_1-s_2)(s_2-s_3)\ldots(s_{k-1}-s_k) \end{pmatrix} u(s_1)\ldots u(s_k)\, \mathrm{ds}^k \ .
\end{aligned}$$

Zur Veranschaulichung wählen wir die Bang-Bang Kontrolle

$$u(t) := \begin{cases} 1 \ , & 0 \leq t < 0.25\pi \\ 0 \ , & 0.25\pi \leq t < 0.375\pi \\ 1 \ , & t \geq 0.375\pi \end{cases} \ ,$$

und vergleichen den Verlauf der exakten Lösungstrajektorie mit dem Verlauf der approximierten Trajektorie bezüglich der Kontrolle u. (Die exakte Lösung ist etwa für $t \geq 0.375\pi$ gegeben durch

$$\Phi(t - 0.375\pi; 1) \cdot \Phi(0.125\pi; 0) \cdot \Phi(0.25\pi; 1) \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} \ ,$$

wobei $\Phi(\cdot\,; u)$ die Fundamentallösung von (AWP) bezeichnet.)

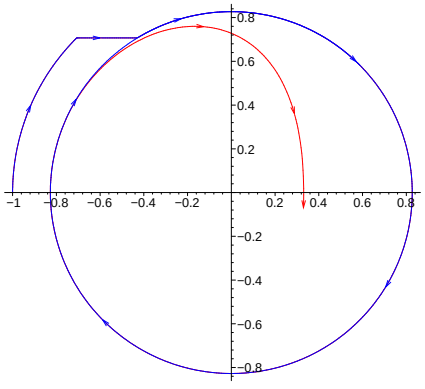

Abbildung 4.1: Approximierte Lösung

Die Abbildung 4.1 soll zeigen, dass die Approximation (rote Kurve) bis zum

Zeitpunkt $t = 2.375\pi$ nahezu perfekt ist. Doch anstelle den Kreis zu schließen, wie die optimale blaue Lösungskurve, bricht die rote Kurve nach unten weg. Dennoch ist das Ergebnis unerwartet gut—schließlich wollten wir nur auf dem Intervall $[0, \frac{\pi}{2}]$ approximieren.

Charakterisierung erreichbarer Mengen

Volterra-Reihen ermöglichen eine explizite Darstellung der erreichbaren Mengen von bilinearen Systemen der Form (AWP). Diese ist bei Systemen mit endlicher Volterra-Entwicklung besonders gut interpretierbar. Betrachte z.B. die erreichbare Menge $\mathcal{A}_1(0)$ des Systems aus Beispiel 4.21:

$$\mathcal{A}_1(0) = \left\{ \begin{pmatrix} \int_0^1 u_1(s)\mathrm{ds} \\ \int_0^1 (1-s)u_1(s)\mathrm{ds} \\ \int_0^1 u_2(s)\mathrm{ds} \\ \int_0^1 (1-s)u_2(s)\mathrm{ds} \\ \int_0^1 \int_0^{s_1} u_2(s_1)u_1(s_2)\mathrm{ds}_2\mathrm{ds}_1 \end{pmatrix} \;\middle|\; u_1, u_2 : [0,1] \to [-1,1] \text{ messbar} \right\}$$

Mit einfachen Überlegungen können wir nun $\mathcal{A}_1(0)$ charakterisieren. Zur Vereinfachung werden nur geometrische Eigenschaften der Projektion von $\mathcal{A}_1(0)$ auf die $x_1x_2x_5$-Ebene angegeben:

- Die Projektion von $\mathcal{A}_1(0)$ auf die x_1-Achse ist gleich $[-1, 1]$.
- Die Projektion auf die x_2- bzw. x_5-Achse ist gleich $[-\frac{1}{2}, \frac{1}{2}]$.
- Die Projektion auf die x_1x_2-Ebene ist konvex.
- Die Projektion auf die $x_1x_2x_5$-Ebene ist punktsymmetrisch zum Ursprung, achsensymmetrisch bzgl. der x_5-Achse und flächensymmetrisch bzgl. der x_1x_2-Ebene.

Abbildung 4.2 zeigt die Projektion von $\mathcal{A}_1(0)$ auf die x_1x_2-Ebene. Dies ist anscheinend eine konvexe Menge innerhalb des Gebiets $[-1, 1] \times [-\frac{1}{2}, \frac{1}{2}]$.

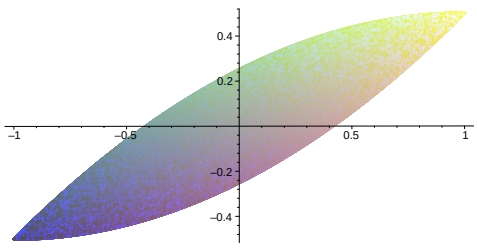

Abbildung 4.2: Projektion aus der Vogelperspektive

Um die Symmetrieeigenschaften zu veranschaulichen, ist die Projektion von $\mathcal{A}_1(0)$ auf die $x_1x_2x_5$-Ebene aus zwei verschiedenen Perspektiven in Abbildung 4.3 dargestellt.
Natürlich handelt es sich bei den Plots nur um Approximationen, für die immerhin mehr als 100000 erreichbare Punkte berechnet wurden, die von stückweise konstanten Bang-Bang Funktionen $u : [0,1] \to \{\binom{\pm 1}{\pm 1}\}$ mit bis zu 10 zufällig bestimmten Sprungstellen angesteuert worden waren.

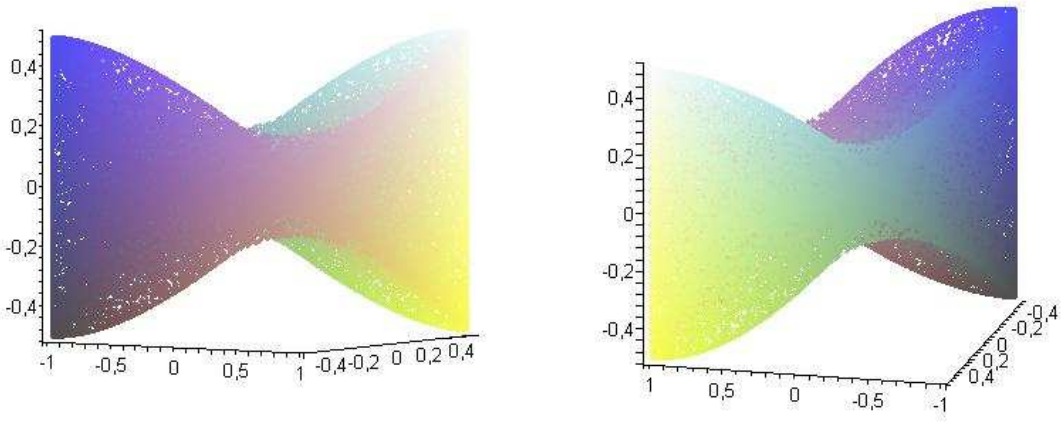

Abbildung 4.3: Projektion auf $x_1x_2x_5$-Ebene

Nikolay Kirov untersucht in [26, p.344] dasselbe System und berechnet dort vergleichbare Plots. Er verwendet ebenfalls stückweise konstanten Kontrollen, und setzt diese in die endliche Volterra-Entwicklung ein. Im Gegensatz zu dieser Vorgehensweise habe ich die erreichbaren Punkte meiner Approximation über die Fundamentallösungen $\Phi\left(\cdot\,;\binom{\pm 1}{\pm 1}\right)$ des homogenisierten Systems kalkuliert.

Ausblick

Die Diplomarbeit war gedacht als Einführung in die Klasse der bilinearen Systeme. Der Leser sollte sich vom Umfang der Arbeit nicht täuschen lassen—viele Aspekte wurden nur kurz oder gar nicht angesprochen. Ich möchte daher einige möglichen Anwendungen und Weiterentwicklungen aufzeigen.

In Abschnitt 3.2 wurde ein Zusammenhang zur Lie-Theorie hergestellt. Wie man in [13], [2] und [3] nachlesen kann, können Resultate aus der Lie-Theorie verwendet werden, um für autonome homogene bilineare Systeme eine Kontrolltheorie analog zu der von linearen Systemen mit Prinzipien wie Kontrollierbarkeit und Beobachtbarkeit aufzubauen. Eine Schlüsselfrage ist, ob die erreichbare Menge von der Einheitsmatrix (bzgl. des assoziierten Matrixsystems) die gesamte Lie-Gruppe bildet, auf der das System definiert ist.
Ergebnisse aus Abschnitt 3.5—insbesondere das Maximumprinzip, die Konvexitätsaussagen und das schwache Bang-Bang Prinzip—könnten die Grundlage für Optimierungsverfahren im Rahmen der bilinearen Systeme von Rang 1 schaffen.
Die Bemerkung 3.90 verweist auf einen Artikel von Otomar Hàjek [10], in welchem die Theorie aus Abschnitt 3.5.2 auf eine größere Kategorie von Spalten-Kontrollsystemen verallgemeinert wird. Die Klasse der Zeilen-Kontrollsysteme scheint dagegen nicht näher untersucht zu sein ...
In Kapitel 4 berechneten wir die Volterra-Reihen-Repräsentation der Ein-Ausgangsfunktion bilinearer Systeme. Die Ad-Hoc-Anwendungen aus Abschnitt 4.5 verdienen größere Aufmerksamkeit und könnten systematisch ausgebaut werden. Die Lösungen von bilinearen Systemen können nicht nur in Volterra-Reihen entwickelt werden: Ein Vergleich zur sog. Fliess-Entwicklung [12] wäre interessant.
Bei der Berechnung der Volterra-Kerne sind wir immer davon ausgegangen, dass die Zustandsraumdarstellung des vorgegebenen bilinearen Systems vollständig bekannt ist. Ein Physiker, der ein weitgehend unbekanntes System anhand seines Ein-Ausgangsverhaltens analysieren will, wird genau andersherum vorgehen: Anhand von Messdaten und Experimenten kann er eventu-

ell die Ein-Ausgangsfunktion des untersuchten Systems durch eine endliche Volterra-Entwicklung approximieren, die er unter Umständen durch ein bilineares System realisieren kann (d.h. dieses bilineare System besitzt genau das Ein-Ausgangsverhalten des unbekannten Systems). Es stellt sich automatisch die Frage, ob eine endliche Volterra-Entwicklung durch ein bilineares System realisiert werden kann. Roger W. Brockett leitet in [4] eine notwendige und hinreichende Bedingung her. Weitere Informationen und Anregungen aus dem Gebiet der Realisierung und Identifikation findet man in [22] und [21].

Anhang A

Konzepte aus der Funktionalanalysis

A.1 Normierte Räume

Es sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, $\mathcal{U}$ sei ein metrischer Raum und I ein reelles Intervall.

Definition A.1. Das Paar $(X, \|\cdot\|)$, bestehend aus einem $\mathbb{K}$-Vektorraum X und einer Norm $\|\cdot\|$, ist ein *normierter Raum*. Um Verwechslungen zu vermeiden, schreiben wir auch $\|\cdot\|_X$ statt $\|\cdot\|$.
Ein normierter Raum wird zum *metrischen Raum* (X, d), wenn man die Metrik definiert als $d(x, y) := \|x - y\|$.
Ein vollständig normierter Raum heißt *Banachraum*. Banachräume, deren Norm durch ein Skalarprodukt induziert wird ($\|x\| := \langle x, x\rangle^{1/2}$), nennt man *Hilberträume*.

Bevor wir uns normierten Funktionenräumen zuwenden, nutzen wir die Gelegenheit, um einige relevante Funktionenklassen und Begriffe aus der Lebesgue-Theorie einzuführen.

Definition A.2. Eine Funktion $f : [a, b] \to \mathcal{U}$ (mit $a < b$) heißt *Treppenfunktion*, falls es eine endliche Zerlegung

$$a = a_1 < a_2 < \ldots < a_k = b$$

von $[a, b]$ gibt, so dass die Restriktion $f|_{(a_i, a_{i+1})}$ konstant ist ($i = 1, \ldots, k-1$).
Wir sagen, eine Funktion $f : I \to \mathcal{U}$ ist *stückweise konstant*, falls jede Restriktion von f auf ein kompaktes Teilintervall von I eine Treppenfunktion ist.

Sprechweise A.3. Wir sagen, dass eine „Eigenschaft“ *fast überall* (kurz: f.ü.) gültig ist, falls die Menge der Elemente, für die diese Eigenschaft nicht gilt, eine (Lebesgue-) Nullmenge ist.
So stimmen z.B. zwei Funktionen $f, g : I \to \mathcal{U}$ genau dann fast überall überein, wenn die Menge

$$\{t \in I \mid f(t) \neq g(t)\}$$

eine Nullmenge ist.
Gelegentlich sagt man auch, dass eine Eigenschaft für *fast alle* Elemente gültig ist, anstelle zu sagen, dass sie fast überall gültig ist.

Definition A.4. Eine Funktion $f : I \to \mathcal{U}$ heißt (Lebesgue-) messbar, falls es eine Folge stückweise konstanter Funktionen $(f_i)_{i\in\mathbb{N}}$ gibt, die fast überall gegen f konvergiert (d.h. die Menge $\{t \in I \mid f_i(t) \not\to f(t)\}$ ist eine Nullmenge).

Definition A.5. Eine messbare Funktion $f : I \to \mathcal{U}$ ist *essentiell beschränkt*, falls eine kompakte Menge $K \subseteq \mathcal{U}$ existiert, so dass $f(t) \in K$ für fast alle $t \in I$. Sie ist *lokal* essentiell beschränkt, falls jede Restriktion von f auf ein beschränktes Teilintervall von I essentiell beschränkt ist.

Beispiel A.6. Hier sind einige normierte Räume aufgelistet, die in dieser Arbeit verwendet werden:

1. $\mathbb{R}^n$, versehen mit der euklidischen Norm

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2},$$

ist ein Hilbertraum. Zusammen mit der Maximumsnorm

$$\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$$

bildet er einen Banachraum.

2. Der Raum $C^0([a,b], \mathbb{R}^n)$ der stetigen Funktionen von $[a,b]$ nach $(\mathbb{R}^n, \|\cdot\|)$ ist ein Banachraum bezüglich der Supremumsnorm

$$\|f\|_\infty = \sup_{t\in[a,b]} \|f(t)\| .$$

Um eine Verwechslung mit der Maximumsnorm oder der essentiellen Supremumsnorm zu vermeiden, schreiben wir auch $\|f\|_0$.

3. Der Raum $\mathbb{R}^{n\times m}$ der reellen $n \times m$- Matrizen, versehen mit der Operatornorm

$$\|A\| = \sup_{\substack{x\in\mathbb{R}^m:\\ \|x\|_{\mathbb{R}^m}\leq 1}} \|Ax\|_{\mathbb{R}^n} \ ,$$

ist ein Banachraum.

4. Es bezeichne $\mathcal{L}^\infty(I,\mathbb{R}^n)$ den Raum der messbaren, essentiell beschränkten Funktionen von I nach $\mathbb{R}^n$. Dann ist die Menge

$$L^\infty(I,\mathbb{R}^n) = \mathcal{L}^\infty(I,\mathbb{R}^n)\big/\sim$$

der Äquivalenzklassen[1] von Funktionen aus $\mathcal{L}^\infty(I,\mathbb{R}^n)$, versehen mit der essentiellen Supremumsnorm

$$\begin{aligned}\|f\|_\infty &= \operatorname{ess.sup}_{t\in I} \|f(t)\|_{\mathbb{R}^n} \\ &= \inf\{M>0 \mid \{t\in I \mid \|f(t)\|_{\mathbb{R}^n}\geq M\} \text{ ist eine Nullmenge}\} \ ,\end{aligned}$$

ein Banachraum.

5. Der Raum der quadratintegrablen Funktionen

$$\mathcal{L}_2([a,b],\mathbb{R}^n) = \left\{ f:[a,b]\to\mathbb{R}^n \text{ messbar} \;\middle|\; \int_a^b \|f(s)\|_{\mathbb{R}^n}\mathrm{ds} < \infty \right\}$$

wird zu Hilbertraum $L_2([a,b],\mathbb{R}^n) = \mathcal{L}_2([a,b],\mathbb{R}^n)\big/\sim$ mit Skalarprodukt

$$\langle f,g\rangle_{L_2} = \int_a^b \langle f(s),g(s)\rangle_{\mathbb{R}^n}\mathrm{ds} \qquad \forall f,g\in L_2([a,b],\mathbb{R}^n) \ ,$$

wenn man die Funktionen, die fast überall gleich sind, miteinander identifiziert.

6. Die Menge $\Omega(\mathbb{R}^n) := \{A\subset\mathbb{R}^n \mid A \text{ kompakt, nichtleer}\}$ bildet zusammen mit dem Hausdorffabstand

$$d_H(A,B) := \max\left\{ \sup_{a\in A}\left(\inf_{b\in B}\|a-b\|_{\mathbb{R}^n}\right) \ , \ \sup_{b\in B}\left(\inf_{a\in A}\|a-b\|_{\mathbb{R}^n}\right)\right\}$$

einen metrischen Raum. Insbesondere ist $d_H(A,B)$ die kleinste Zahl $\epsilon\geq 0$, so dass $A\subseteq\overline{B_\epsilon(B)}$ und $B\subseteq\overline{B_\epsilon(A)}$.
(Hier bezeichnet die letzte Menge den Abschluß von $B_\epsilon(A) = \bigcup_{a\in A} B_\epsilon(a)$, wobei $B_\epsilon(a) = \{b\in\mathbb{R}^n \mid \|a-b\|_{\mathbb{R}^n} < \epsilon\}$ die offene Einheitskugel von $(\mathbb{R}^n,\|\cdot\|_{\mathbb{R}^n})$ mit Radius ϵ ist.)

[1] $f\sim g$ bedeutet $f(t)=g(t)$ für fast alle $t\in I$

Bemerkung A.7. Je zwei Normen $\|\cdot\|_a$ und $\|\cdot\|_b$ auf dem $\mathbb{R}^n$ (bzw. $\mathbb{R}^{n\times m}$) sind äquivalent, d.h. es gibt positive Konstanten C_1, C_2, so dass

$$C_1\|x\|_b \leq \|x\|_a \leq C_2\|x\|_b \qquad \forall x \in \mathbb{R}^n \text{ (bzw. } \mathbb{R}^{n\times m}) \ .$$

Es folgt, dass auf diesen Räumen sämtliche topologischen Aussagen unabhängig von der gewählten Norm gelten.
Wenn nicht anders vorausgesetzt, ist die Norm eines Vektors $x \in \mathbb{R}^n$ stets die euklidische Norm, d.h. $\|x\| = \|x\|_2$.

Die folgenden Rechenregeln sollten dem Leser wohlbekannt sein. Zur Erinnerung:

(i) Im $\mathbb{R}^n$ gilt die Ungleichung von Cauchy-Schwartz:

$$|\langle x, y\rangle|^2 = |x^*y|^2 \leq \|x\|_2 \cdot \|y\|_2 \quad \forall x, y \in \mathbb{R}^n$$

(ii) Aus der Definition der Operatornorm folgt direkt:

$$\|Ax\| \leq \|A\| \cdot \|x\| \qquad \forall A \in \mathbb{R}^{n\times m}, x \in \mathbb{R}^n$$

(iii) Ist $A \in \mathbb{R}^{r\times n}$, $B \in \mathbb{R}^{n\times m}$ und $x \in \mathbb{R}^n$, so ist wegen (ii)

$$\|ABx\| \leq \|A\| \cdot \|Bx\| \leq \|A\| \cdot \|B\| \cdot \|x\|$$

und daher

$$\|AB\| \leq \|A\| \cdot \|B\| \ .$$

Den Satz von Arzelà-Ascoli wird uns als zentrales Hilfsmittel zum Beweis der lokalen Existenz von Caratheory-Lösungen dienen. Wir werden gleich sehen, wie dieser Satz in der Funktionalanalysis [27] formuliert wird.

Definition und Lemma A.8. Für einen metrischen Raum (X, d) sind äquivalent:

(i) X ist *(überdeckungs-) kompakt*, d.h. jede offene Überdeckung von X besitzt eine endliche Teilüberdeckung.

(ii) X ist *folgenkompakt*, d.h. jede Folge in X besitzt eine konvergente Teilfolge mit Grenzwert in X.

Satz A.9 (Arzelà-Ascoli)**.** Es sei X ein kompakter metrischer Raum und $C^0(X, \mathbb{K}^n)$ der Banachraum aller stetigen Funktionen von X nach $\mathbb{K}^n$, versehen mit der Supremumsnorm $\|\cdot\|_\infty$.
Eine Teilmenge $M \subset C^0(X, \mathbb{K}^n)$ ist kompakt, falls sie die folgenden Eigenschaften erfüllt:

(i) M ist beschränkt.

(ii) M ist abgeschlossen.

(iii) M ist gleichgradig stetig.

Gelten nur (i) und (iii), so ist zumindest der Abschluß $\overline{M}$ von M in $C^0(X, \mathbb{K}^n)$ kompakt.

Beweis. Siehe [27, p.68].

Folgerung A.10. Ist X ein abgeschlossenes Intervall $[a, b]$ in $\mathbb{R}$, und M eine Folge $(x_k)_{k\in\mathbb{N}}$ stetiger Funktionen von $[a, b]$ nach $\mathbb{R}^n$, so folgt aus den Sätzen A.8 und A.9:
Falls $(x_k)_{k\in\mathbb{N}}$ gleichmäßig beschränkt und gleichgradig stetig ist, so existiert eine Teilfolge von $(x_k)_{k\in\mathbb{N}}$, die gleichmäßig gegen eine stetige Funktion $x : [a, b] \to \mathbb{R}^n$ konvergiert.

A.2 Absolut stetige Funktionen

Es sei $[a, b]$ ein reelles Intervall mit $a < b$.

Der Hauptsatz der Differential- und Integralrechnung in seiner einfachsten Form [8, Satz 19.3] besagt, dass für stetig differenzierbare Funktionen $f : [a, b] \to \mathbb{R}$ die Integralgleichung

$$f(t) = f(a) + \int_a^t f'(s)\mathrm{ds} \qquad \forall t \in [a, b] \tag{A.1}$$

gilt.
Will man den Hauptsatz auf stetige Funktionen übertragen, die nur fast überall differenzierbar sind, so muss f eine weitere Voraussetzung erfüllen.

Definition A.11. Eine Funktion $f : [a, b] \to \mathbb{R}$ heißt *absolut stetig*, wenn zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert, so dass für alle $n \in \mathbb{N}$ und jeder endlichen Kette

$$a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n \text{ mit } a_i, b_i \in [a, b] \ (i = 1, \ldots, n)$$

die Implikation

$$\sum_{i=1}^{n}(b_i - a_i) < \delta \Longrightarrow \sum_{i=1}^{n} |f(b_i) - f(a_i)| < \epsilon \tag{A.2}$$

gilt.

Beispiel A.12. Lipschitz-stetige Funktionen sind absolut stetig.
Denn erfüllt $f : [a, b] \to \mathbb{R}$ die Lipschitzbedingung mit Konstante L, so gilt

$$\sum_{i=1}^{n} |f(b_i) - f(a_i)| \leq \sum_{i=1}^{n} L(b_i - a_i) = L \sum_{i=1}^{n} (b_i - a_i)$$

und der rechte Ausdruck wird mit $\sum_{i=1}^{n}(b_i - a_i)$ beliebig klein.

Bezeichnung A.13. Mit $\mathcal{L}^1([a, b])$ bezeichnen wir die Menge aller über $[a, b]$ Lebesgue-integrierbaren Funktionen, die fast überall auf $[a, b]$ definiert sind.

Satz A.14 (Hauptsatz der Differential- und Integralrechnung)**. (i)** Sei $f(\cdot)$ eine absolut stetige Funktion über $[a, b]$.
Dann ist $f' \in \mathcal{L}^1([a, b])$ und

$$f(t) = f(a) + \int_a^t f'(s)\mathrm{ds}$$

für alle $t \in [a, b]$.

(ii) Hat die Funktion f über $[a, b]$ die Form

$$f(t) = f(a) + \int_a^t \rho(s)\mathrm{ds}$$

für ein $\rho \in \mathcal{L}^1([a, b])$, so ist f absolut stetig über $[a, b]$.
(In diesem Fall ist wegen (i) $\rho(t) = f'(t)$ f.ü. auf $[a, b]$.)

Folgerung A.15.

$$f \text{ absolut stetig über } [a, b] \iff f' \in \mathcal{L}^1([a, b]) \text{ und } f(t) = f(a) + \int_a^t f'(s)\mathrm{ds} \qquad \forall\, t \in [a, b]$$

Bemerkung A.16. Jedes Paar absolut stetiger Funktionen $f, g : [a, b] \to \mathbb{R}$ besitzt die folgenden Eigenschaften:

1. f ist stetig und von beschränkter Variation.

2. $f + g$ und $f \cdot g$ sind absolut stetig. Ist g stets ungleich Null, so ist auch der Quotient $\frac{f}{g}$ absolut stetig.

3. $f'(t) = 0$ fast überall $\Rightarrow f$ konstant.

4. $f'(t) \leq 0$ fast überall $\Rightarrow f$ monoton fallend.

Es sei bemerkt, dass es stetige Funktionen (z.B. die „Cantorsche Funktion“) gibt, die nicht absolut stetig sind.

Satz A.17. Ist $g : [a, b] \to [c, d]$ absolut stetig und f Lipschitz-stetig, so ist auch die Komposition $f \circ g$ absolut stetig.

Beweis. $f : [a, b] \to \mathbb{R}$ erfülle die Lipschitzbedingung mit Konstante L. Es gilt:

$$\sum_{i=1}^{n} |f(g(b_i)) - f(g(a_i))| \leq \sum_{i=1}^{n} L|g(b_i) - g(a_i)| = L \sum_{i=1}^{n} |g(b_i) - g(a_i)|$$

Die rechte Seite der Ungleichung wird mit $\sum_{i=1}^{n}(b_i - a_i)$ beliebig klein, da g absolut stetig ist. □

Definition und Lemma A.18. Eine vektorwertige Funktion $f : [a, b] \to \mathbb{R}^n$ ist genau dann absolut stetig, falls die folgenden Bedingungen erfüllt sind:

(i) Die Ableitung

$$f'(t) = \lim_{\substack{h \to 0 \\ h \neq 0}} \frac{f(t+h) - f(t)}{h}$$

existiert für fast alle $t \in [a, b]$ und ist integrierbar über $[a, b]$.

(ii) Es gilt die Integralgleichung

$$f(t) = f(a) + \int_a^t f'(s)\mathrm{ds} \qquad \forall t \in [a, b] \ .$$

(Es wird stets komponentenweise integriert.)

A.3 Schwache Konvergenz

Es sei H ein Hilbertraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$.
Wir führen einen aus der Funktionalanalysis bekannten Konvergenzbegriff ein, der eine Abschwächung zur Konvergenz bezüglich der Norm darstellt.

Definition A.19. Eine Folge (f_k) in H *konvergiert schwach* gegen $f \in H$, falls

$$\langle f_k, h\rangle \to \langle f, h\rangle \quad \text{für } k \to \infty \qquad \forall h \in H \ .$$

Schreibe kurz: $f_k \xrightarrow{w} f$.

Bemerkung A.20. 1. Konvergiert eine Folge (f_k) in H gegen $f \in H$ bezüglich der Norm von H, so spricht man von „starker Konvergenz“. Wir schreiben kurz: $f_k \to f$ oder $f_k \xrightarrow{s} f$. Es gilt aufgrund der Stetigkeit des Skalarprodukt, dass aus starker Konvergenz $f_k \to f$ die schwache Konvergenz $f_k \xrightarrow{w} f$ folgt.

2. Ist K eine dichte Teilmenge von H und (f_k) eine beschränkte Folge in H, so konvergiert (f_k) genau dann schwach gegen $f \in H$, wenn

$$\langle f_k, h\rangle \to \langle f, h\rangle \quad \text{für } k \to \infty \qquad \forall h \in K$$

gilt.

3. Schwach konvergente Folgen besitzen genau einen Grenzwert.

Satz A.21 (Banach-Saks)**.** Angenommen, $f_n \xrightarrow{w} f$ in H. Dann existiert eine Teilfolge $(f_{n_k}) \subseteq (f_n)$ derart, dass für das arithmetische Mittel $F_N := \frac{1}{N}\sum_{k=1}^{N} f_{n_k}$ gilt: $F_N \xrightarrow{s} f$ für $N \to \infty$.

Beweis. Siehe [27, p.223].

Definition A.22. Die *schwache Topologie* τ_w ist die „schwächste“ Topologie auf H, für welche die Abbildungen

$$\begin{aligned} L_h : \quad & H \longrightarrow \mathbb{K} \\ & f \longmapsto \langle f, h\rangle \end{aligned}$$

stetig sind für alle $h \in H$. (Wie diese Topologie konstruiert wird, sieht man in [27, p.315ff] oder [1, p.217].)

Bemerkung A.23. Versieht man H mit der schwachen Topologie τ_w, so erhält man einen „lokal konvexen topologischen Vektorraum“ (Definition in [1, p.142]). Insbesondere ist (H, τ_w) ein Hausdorffraum, in welchem der (topologische) Konvergenzbegriff mit dem Begriff der schwachen Konvergenz übereinstimmt.

Sprechweise A.24. Teilmengen von H heißen *schwach (folgen-) kompakt* oder *schwach abgeschlossen*, wenn sie (folgen-) kompakt bzw. abgeschlossen bezüglich der schwachen Topologie auf H sind.

Bemerkung A.25. Aus der Topologie ist bekannt, dass im Falle H separabel (dann erfüllt (H, τ_w) das „1. Abzählbarkeitsaxiom“) das Folgenkriterium für schwache Abgeschlossenheit nicht nur notwendig, sondern auch hinreichend ist. Das bedeutet, eine Teilmenge $M \subset H$ ist genau dann schwach abgeschlossen, wenn aus $(f_k) \subset M$ und $f_k \xrightarrow{w} f \in H$ folgt, dass $f \in M$.
Außerdem sind in diesem Fall schwach folgenkompakt und schwach kompakt äquivalente Eigenschaften.

Folgerung A.26. Für jede konvexe Teilmenge M eines separablen Hilbertraumes H gilt:

$$M \text{ schwach abgeschlossen} \iff M \text{ abgeschlossen}$$

Beweis. 1. Es sei M schwach abgeschlossen und $(f_k) \subset M$ eine konvergente Folge in M mit Grenzwert $f \in H$. Es folgt $(f_k) \xrightarrow{w} f$ und $f \in M$.

2. Es sei M abgeschlossen und $(f_k) \subset M$ eine schwach konvergente Folge in M mit Grenzwert $f \in H$. Nach dem Satz von Banach-Saks A.21 existiert eine gegen f konvergierende Folge (F_N), deren Elemente eine Konvexkombination von Elementen aus M bilden. Da M konvex und abgeschlossen ist, muß $f \in M$ gelten. □

Satz A.27. Die abgeschlossene Einheitskugel $\overline{B_1(0)} \subset H$ ist schwach folgenkompakt, d.h. jede Folge in $\overline{B_1(0)}$ besitzt eine schwach konvergente Teilfolge mit Grenzwert in $\overline{B_1(0)}$.

Beweis. Siehe [1, p.219]. (Es wird die „Reflexität“ des Hilbertraumes H genutzt.)

Folgerung A.28. Es sei $M \subset H$ abgeschlossen, beschränkt und konvex. Dann ist M schwach folgenkompakt. (Im Falle H separabel, ist M sogar schwach kompakt.)

Beweis. Da M abgeschlossen und konvex ist, ist es zumindest schwach abgeschlossen im Sinne des Folgenkriteriums (Satz von Banach-Saks). Aus der Beschränktheit von M folgt, dass ein $\epsilon > 0$ existiert, so dass $M \subset \overline{B_\epsilon(0)}$. Nach Satz A.27 ist $\overline{B_\epsilon(0)}$ schwach folgenkompakt. Insgesamt folgt, dass jede Folge $(f_n) \subset M$ eine schwach konvergente Teilfolge (f_{n_k}) besitzt mit einem Grenzwert $f \in M$, d.h. $f_{n_k} \xrightarrow{w} f$ in M. □

Anhang B

Caratheodory-Systeme

B.1 Systeme ohne Kontrolle

Es sei I ein Intervall in $\mathbb{R}$.
Betrachte eine gewöhnliche Differentialgleichung der Form

$$\dot{x}(t) = f(x(t), t), \qquad f : \mathbb{R}^n \times I \to \mathbb{R}^n \ . \tag{B.1}$$

Die „klassische Lösung“ von (B.1) ist eine differenzierbare Funktion $x(\cdot\,)$ auf einem Intervall $J \subseteq I$, welche diese Gleichung in jedem Punkt $t \in J$ erfüllt. Ist die rechte Seite $f(x,t)$ stetig, so sind die klassischen Lösungen genau durch die Integralgleichung

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(s), s)\mathrm{ds}, \qquad t_0 \in J \tag{B.2}$$

(dabei sei $\int_{t_0}^{t} = -\int_{t}^{t_0}$ für $t_0 > t$) gegeben.
Ist allerdings die rechte Seite unstetig, so ist der klassische Lösungsbegriff zu stark. Es genügt ein einfaches System aus der Physik, um dieses Problem zu verdeutlichen:

Beispiel B.1. Ein Einkaufswagen bewegt sich entlang einer geraden Strecke. Der Zustand des Einkaufswagen zum Zeitpunkt $t \geq 0$ wird durch dessen Position $x(t)$ und Geschwindigkeit $y(t)$ gegeben. Er wird durch die Beschleunigung $u(t)$ kontrolliert, wobei negative Beschleunigung als Bremsen zu verstehen ist. Natürlich ist die Beschleunigung beschränkt. Zur Vereinfachung wird $\mathcal{U} = [-1, 1]$ als Steuerbereich gewählt und Reibungseffekte werden ignoriert. Das System kann durch ein lineares Kontrollsystem modelliert werden:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \qquad -1 \leq u(t) \leq 1$$

Wir wollen den Wagen, ausgehend vom Anfangszustand $\binom{0}{0}$, auf dem Zeitintervall $[0,2]$ maximal beschleunigen und danach im Intervall $]2,3]$ maximal abbremsen. Um die Geschwindigkeit $y(t)$ des Einkaufswagens zum Zeitpunkt $t=3$ zu bestimmen, müssen wir die (eindeutige) Lösung der Differentialgleichung $\dot{y}(t) = u(t)$ mit $y(0) = 0$ und unstetiger rechter Seite

$$u(t) := \begin{cases} +1 & \text{falls } 0 \le t \le 2 \\ -1 & \text{falls } 2 < t \le 3 \end{cases}$$

berechnen, und im Punkt $t = 3$ auswerten. Aus der Integralgleichung (B.2) erhält man die stetige Funktion

$$y(t) = \begin{cases} t & \text{falls } 0 \le t \le 2 \\ 4-t & \text{falls } 2 \le t \le 3 \end{cases} .$$

Dies ist aber keine klassische Lösung, da $y(\cdot)$ an der Stelle $t = 2$ nicht differenzierbar ist.

Unser Ziel ist es daher einen allgemeineren Lösungsbegriff einzuführen, der eindeutige Lösungen für eine möglichst große Klasse von Differentialgleichungen mit unstetiger rechter Seite zulässt. Naheliegend ist es die Funktionen, welche die Integralgleichung (B.2) erfüllen, als Lösung von (B.1) zu deklarieren. Dazu sollte die rechte Seite $f(x,t)$ zumindest die sogenannten „Caratheodory-Bedingungen“ erfüllen:

Bedingung B.2 (Caratheodory-Bedingungen)**. (C1)** $x \mapsto f(x,t)$ ist stetig für alle $t \in I$;

(C2) $t \mapsto f(x,t)$ ist messbar für alle $x \in \mathbb{R}^n$;

(C3) Für alle kompakten $K \subset \mathbb{R}^n$ existiert eine lokal[1] integrierbare Funktion $m : I \to \mathbb{R}_+$, so dass

$$\|f(x,t)\| \le m(t) \qquad \forall t \in I, x \in K .$$

Definition B.3. Sind die Bedingungen (C1)-(C3) erfüllt, so bezeichnen wir die Gleichung $\dot{x} = f(x,t)$ auch als *Caratheodory-Gleichung*.

Im Anhang, Abschnitt A.2, wird der Begriff „absolute Stetigkeit“ von Funktionen definiert, die über einem kompakten Intervall definiert sind. Dieses Konzept lässt sich leicht auf offene, halboffene und unbeschränkte Intervalle übertragen:

[1] d.h. Lebesgue-integrierbar über jedem kompakten Teilintervall von I

Definition B.4. Sei J ein reelles Intervall.
Eine Funktion $x : J \to \mathbb{R}^n$ heißt *lokal absolut stetig*, falls sie absolut stetig auf jedem kompakten Teilintervall von J ist.

Definition und Satz B.5. Eine Funktion $x : J \to \mathbb{R}^n$ heißt *Caratheodory-Lösung* von (B.1) auf $J \subseteq I$, falls sie eine der beiden äquivalenten Bedingungen

(i) $x(\cdot)$ lokal absolut stetig und $\dot{x}(t) = f(x(t), t)$ für fast alle $t \in J$.

(ii) $t \mapsto f(x(t), t)$ ist lokal integrierbar über J und

$$x(t) = x(a) + \int_a^t f(x(s), s)\mathrm{ds} \qquad \forall a, t \in J \ . \tag{B.3}$$

(Dabei sei $\int_a^t = -\int_t^a$ für $a > t$.)

erfüllt.
Ist weiter $x(t_0) = x_0$ für ein $t_0 \in J$, so ist $x(\cdot)$ eine Caratheodory-Lösung des Anfangswertproblems

$$\dot{x}(t) = f(x(t), t), \qquad t \in I, \qquad x(t_0) = x_0 \tag{AWP}$$

auf dem Intervall J.

Beweis. Die Äquivalenz folgt direkt aus dem Hauptsatz der Differential- und Integralrechnung A.14.

Bemerkung B.6. Falls die rechte Seite $f(x, t)$ stetig ist, so ist die Funktion $x(\cdot)$ genau dann eine Caratheodory-Lösung von (B.1), wenn sie eine klassische Lösung ist. Insbesondere ist dann (B.1) eine Caratheodory-Gleichung. In diesem Sinne bildet die Caratheodory-Lösung eine Verallgemeinerung des klassischen Lösungsbegriffs.

Hilfssatz B.7. Erfüllt (B.1) die Caratheodory-Bedingung (C1), so ist $t \mapsto f(x(t), t)$ messbar für jede messbare Funktion $x(\cdot)$.

Beweis. Siehe [23, p.474].

Hilfssatz B.8. Ist $x : [-t_0, -t_0 + d] \to \mathbb{R}^n$ eine (Caratheodory-) Lösung des *zeitumgekehrten* Problems

$$\dot{x}(t) = -f(x(t), -t), \qquad x(-t_0) = x_0 \ , \tag{B.4}$$

so ist $\tilde{x}(t) := x(-t)$ eine Lösung des Ausgangsproblems

$$\dot{x}(t) = f(x(t), t), \qquad x(t_0) = x_0 \ ,$$

auf dem Intervall $[-d + t_0, t_0]$.

Satz B.9 (lokale Existenz)**.** Es sei (B.1) eine Caratheodory-Gleichung. Für alle Anfangswerte $(x_0, t_0) \in \mathbb{R}^n \times I$ existiert ein $d > 0$ und eine Caratheodory-Lösung $x(\cdot)$ von (AWP) auf $J := [t_0 - d, t_0 + d] \cap I$.

Beweis. Wir weisen zunächst die Existenz eines $d > 0$ nach, so dass auf dem Intervall $[t_0, t_0 + d] \cap I$ eine Caratheodory-Lösung existiert. Ohne Einschränkung können wir voraussetzen, dass x_0 kein rechter Randpunkt von I ist (sonst wähle $d > 0$ beliebig). Es gibt also ein $\delta > 0$, so dass $t_0 + \delta \in I$. Wegen Caratheodory-Bedingung (C3) existiert eine lokal integrierbare Funktion $m : I \to \mathbb{R}_+$, so dass

$$\|f(x,t)\| \le m(t) \qquad \forall t \in I, x \in \overline{B_\delta(x_0)} \ .$$

Die Funktion $\rho(t) := \int_{t_0}^t m(s)\mathrm{ds}$ ist dann eine monoton wachsende, absolut stetige Funktion auf $[t_0, \delta]$ mit $\rho(t_0) = 0$. Wähle $0 < d \le \delta$, so dass $\rho(t_0 + d) \le \delta$, und zerlege $[t_0, t_0 + d]$ in $k \ge 1$ Teilintervalle der Länge $h = \frac{d}{k}$. Auf diesen Intervallen $[t_0 + ih, t_0 + (i+1)h]$, $i = 0, 1, \ldots, k-1$, konstruieren wir schrittweise „$i \to i + 1$“ eine approximierende Funktion $x_k(\cdot)$. Setze dazu

$$x_k(t) := \begin{cases} x_0 & \text{für } t \le t_0, \\ x_0 + \int_{t_0}^t f(x_k(s-h), s)\mathrm{ds} & \text{für } t \in [t_0, t_0 + d]. \end{cases} \tag{B.5}$$

Um zu prüfen, ob dies wohldefiniert ist, benötigen wir die Messbarkeit des Integranden gemäß Hilfssatz B.7 und die integrierbare Majorante $m(\cdot)$. Wir erhalten (schrittweise) für alle $t \in [t_0, t_0 + d]$

$$\|x_k(t) - x_0\| \le \int_{t_0}^t \|f(x_k(s-h), s)\|\mathrm{ds} \le \int_{t_0}^t m(s)\mathrm{ds} = \rho(t) \le \rho(t_0 + d) \le \delta \ ,$$

und entsprechend für vorgegebene Zeitpunkte $s \le t$ aus $[t_0, t_0 + d]$

$$\|x_k(t) - x_k(s)\| \le \int_s^t \|f(x_k(s-h), s)\|\mathrm{ds} \le \int_s^t m(s)\mathrm{ds} = |\rho(t) - \rho(s)| \ .$$

Es ist nun leicht zu sehen, dass die Funktionenfolge $(x_k)_{k\in\mathbb{N}}$ gleichmäßig beschränkt und gleichgradig stetig ist (zumal ρ stetig). Nach dem Satz von Arzelà-Ascoli A.10 existiert daher eine gleichmäßig konvergente Teilfolge, die ebenfalls mit $(x_k)_{k\in\mathbb{N}}$ beschrieben werde, mit einem Grenzwert $x(\cdot)$ in $C^0([t_0, t_0 + d], \mathbb{R}^n)$.
Schließlich wird gezeigt, dass $x(\cdot)$ die gesuchte Caratheodory-Lösung auf $[t_0, \delta]$ ist. Zunächst konvergiert $x_k(s-h)$ gegen $x(s)$ für $k \to \infty$ (im $\mathbb{R}^n$), weil $(x_k)_{k\in\mathbb{N}}$ gleichgradig stetig ist und

$$\|x_k(s-h) - x(s)\| \le \|\underbrace{x_k(s-h)}_{x_k(s-\frac{d}{k})} - x_k(s)\| + \underbrace{\|x_k(s) - x(s)\|}_{\to 0} \ .$$

Verwendet man nun $m(\cdot)$ als Majorante und berücksichtigt die Stetigkeit von $f(x,t)$ in x, so ermöglicht uns der Satz von der majorisierten Konvergenz, Grenzwert und Limesbildung in (B.5) zu vertauschen. Man erhält

$$\begin{aligned} x(t) &= \lim_{k\to\infty} x_k(t) = x_0 + \lim_{k\to\infty} \int_{t_0}^{t} f(x_k(s-h),s)\mathrm{ds} \\ &= x_0 + \int_{t_0}^{t} \lim_{k\to\infty} f(x_k(s-h),s)\mathrm{ds} = x_0 + \int_{t_0}^{t} f(\lim_{k\to\infty} x_k(s-h),s)\mathrm{ds} \\ &= x_0 + \int_{t_0}^{t} f(x(s),s)\mathrm{ds}\ . \end{aligned} \tag{B.6}$$

Folglich erfüllt $x(\cdot)$ im Intervall $[t_0, t_0+d]$ die Integralgleichung (B.6), und ist somit dort eine Caratheodory-Lösung.
Beginnt man den Beweis für das zeitumgerkehrte Problem (B.4) unter der Annahme, dass x_0 kein linker Randpunkt I ist, so erhält man eine Lösung $\tilde{x}(t)$ auf dem Intervall $[t_0-d, t_0]\cap I$. (Man beachte, dass auch $-f(x,-s)$ die Caratheodory-Bedinungen erfüllt.) Konkatenation $\tilde{x}\&_{t_0}x$ liefert die beidseitige Lösung. □

Sprechweise B.10. Es sei $[t_0,\infty)\subseteq I$, und $x(\cdot)$ sei eine Lösung von (B.1) auf einem Intervall $[t_0,t_1)$ mit $t_0 < t_1 \le \infty$.
Falls $t_1 < \infty$ ist und es eine weitere Lösung auf einem Intervall $[t_1, t_1+d)$ für ein $d>0$ existiert, so erhalten wir per Konkatenation eine „erweiterte Lösung“ von $x(\cdot)$ auf $[t_0, t_1+d)$. Man sagt, die Lösung $x(\cdot)$ kann auf $[t_0,t_1+d)$ „erweitert“ werden.
Sonst nennt man $x(\cdot)$ „nicht erweiterbar in $[t_0,\infty)$“.
Entsprechende Sprechweisen verwenden wir im Falle $(-\infty,t_0]\subseteq I$ und $I=\mathbb{R}$.

Hilfssatz B.11. Es sei $I=[t_0,\infty)$. Zu jeder Lösung von (B.1) existiert eine erweiterte Lösung $x(\cdot)$ auf einem Intervall $[t_0,\omega)$ mit $t_0<\omega\le\infty$, die nicht erweiterbar in $[t_0,\infty)$ ist. Ist $\omega<\infty$, so heißt ω eine „endliche Fluchtstelle“. Es gilt dann notwendigerweise

$$\|x(t)\|\to\infty \qquad \text{für } t\to\omega\ .$$

Beweis. Siehe [9, p.35f].

Satz B.12 (globale Existenz)**. (a)** Angenommen, es ist $I=[t_0,\infty)$. Die rechte Seite $f(x,t)$ der Caratheodory-Gleichung (B.1) erfülle die Ungleichung

$$\frac{x^* f(x,t)}{\|x\|^2} \le \mu(t) \qquad \forall \|x\|\ge\zeta, t\ge t_0 \tag{B.7}$$

für ein $\zeta > 0$ und eine lokal integrierbare Funktion $\mu(\cdot)$.
Dann kann jede Lösung von (B.1) (per Konkatenation) auf $[t_0, +\infty)$ erweitert werden. Man sagt, die Gleichung (B.1) hat „globale Existenz in die Zukunft".

(b) Angenommen, es ist $I = \mathbb{R}$. Gilt sogar

$$\frac{\|f(x,t)\|}{\|x\|} \le \mu(t) \qquad \forall \|x\| \ge \zeta, t \in \mathbb{R} \tag{B.8}$$

mit ζ und $\mu(\cdot)$ wie in (a), so kann jede Caratheodory-Lösung von (B.1) auf ganz $\mathbb{R}$ erweitert werden. Man sagt, (B.1) hat „globale Existenz in Zukunft und Vergangenheit".

Beweis. 1. Wegen Hilfssatz B.11 genügt zu zeigen, dass keine Caratheodory-Lösung von (B.1) auf einem beschränkten Intervall $[t_0, w)$ existiert mit $\|x(t)\| \to +\infty$ für $t \to w$. Annahme: Dies wäre der Fall.
Die Funktionen $r(t) := \|x(t)\|^2$ bzw. $\log r(t)$ sind laut Satz A.17 absolut stetig und daher fast überall differenzierbar (falls t nahe w). Es gilt fast überall

$$\begin{aligned} \dot r(t) &= \frac{\mathrm{d}}{\mathrm{dt}}\|x(t)\|^2 = \frac{\mathrm{d}}{\mathrm{dt}}\big(\|\cdot\|^2 \circ x(\cdot)\big)\big|_t = (2x_1, \ldots, 2x_n)\big|_{x(t)}\dot x(t) \\ &= 2x(t)^* \dot x(t) = 2x(t)^* f(x(t), t) \end{aligned}$$

und folglich (für t nahe w)

$$\frac{\mathrm{d}}{\mathrm{dt}} \log r(t) = \frac{\dot r(t)}{r(t)} = \frac{2x(t)^* f(x(t),t)}{\|x(t)\|^2} \overset{\text{(B.7)}}{\le} 2\mu(t)\ .$$

Integrieren über $[s,t] \subset [t_0, w)$ liefert für s nahe w fast überall:

$$\begin{aligned} \underbrace{\int_s^t \frac{\dot r(\lambda)}{r(\lambda)}\mathrm{d}\lambda}_{\substack{\log r(t) - \log r(s) \\ = \log(\frac{r(t)}{r(s)})}} &\le 2\int_s^t \mu(\lambda)\mathrm{d}\lambda \\ \overset{\exp(.)}{\Longleftrightarrow}\ r(t) &\le r(s)\exp\big(2\int_s^t \mu(\lambda)\mathrm{d}\lambda\big) \end{aligned}$$

Für $t \to w$ konvergiert die rechte Seite der Ungleichung gegen $r(s)\exp\big(2\int_s^w \mu\big) < \infty$, da μ integrierbar über $[s,w]$ ist, und die linke Seite gegen $+\infty$ gemäß der Annahme. Dies ist ein Widerspruch.

2. Die zweite Aussage folgt aus

$$\frac{x^* f(x,t)}{\|x\|^2} \leq \frac{\|x\| \cdot \|f(x,t)\|}{\|x\|^2} = \frac{\|f(x,t)\|}{\|x\|} \ .$$

Man hat also globale Existenz in die Zukunft. Das zeitumgekehrte System $\dot{x} = -f(x,-t)$ liefert globale Existenz in die Vergangenheit. □

Das nächste Lemma ermöglicht es, die letzte Caratheodory-Bedingung anhand einer schwächeren Bedingung zu überprüfen.

Lemma B.13. Angenommen, $f : \mathbb{R}^n \times I \to \mathbb{R}^n$ erfüllt neben den Caratheodory-Bedingungen (C1) und (C2) auch die „lokale Lipschitzbedingung in x“:

(L1) Für alle $x_0 \in \mathbb{R}^n$ existiert ein $\delta > 0$ und eine lokal integrierbare Funktion $\mu : I \to \mathbb{R}_+$, so dass

$$\|f(x,t) - f(y,t)\| \leq \mu(t)\|x - y\| \quad \forall x, y \in B_\delta(x_0), t \in I \ .$$

Dann ist $\dot{x} = f(x,t)$ genau dann eine Caratheodory-Gleichung, falls die folgende Bedingung gilt:

(C3') Für alle $x \in \mathbb{R}^n$ existiert eine lokal integrierbare Funktion $m : I \to \mathbb{R}_+$, so dass $\|f(x,t)\| \leq m(t)$ für alle $t \in I$.

Beweis. Es genügt zu beweisen, dass (C3') und (L1) die Bedingung (C3) implizieren. Es gelte also (C3') und (L1), und wir geben eine kompakte Menge $K \subset \mathbb{R}^n$ vor. Für alle $x_0 \in K$ folgt die Existenz lokal integrierbarer Funktionen $\mu, m : I \to \mathbb{R}_+$ (die von x_0 abhängen) und einer positiven Konstante δ_{x_0}, so dass für alle $x \in B_{\delta_{x_0}}(x_0)$ und $t \in I$ gilt:

$$\begin{aligned} \|f(x,t)\| &\leq \|f(x_0,t)\| + \|f(x,t) - f(x_0,t)\| \leq m(t) + \mu(t)\|x - x_0\| \\ &\leq m(t) + \mu(t)\delta_{x_0} \end{aligned}$$

Offensichtlich ist die letzte Funktion $\gamma_{x_0}(t) := m(t) + \mu(t)\delta_{x_0}$ lokal integrierbar. Ein Kompaktheitsschluß wird helfen eine Majorante auf ganz K zu bestimmen.
Die Vereinigung $\bigcup_{x_0 \in K} B_{\delta_{x_0}}(x_0)$ bildet eine offene Überdeckung der kompakten Menge K. Daher gibt es eine endliche Teilüberdeckung bestehend aus Kugeln mit Zentren $x_1, x_2, \ldots, x_l$ aus K für ein $l \in \mathbb{N}$. Wir definieren nun die Funktion

$$\gamma(t) := \max\{\gamma_{x_1}(t), \gamma_{x_2}(t), \ldots, \gamma_{x_l}(t)\}$$

auf I. Es sollte bekannt sein [7, p.61], dass das Maximum von endlich vielen (Lebesgue-) integrierbaren Funktion selbst integrierbar ist. Die Funktion $\gamma(t)$ erfüllt also nach Konstruktion die Caratheodory-Bedingung (C3). □

Definition B.14. Eine (Caratheodory-) Lösung $x : J \to \mathbb{R}^n$ des Anfangswertproblems (AWP) auf $J \subseteq I$ heißt *maximale Lösung* von (AWP) bezüglich I auf J, falls die folgende Eigenschaft erfüllt ist:
Ist $\tilde{x} : \tilde{J} \to \mathbb{R}^n$ eine weitere Lösung von (AWP) auf $\tilde{J} \subseteq I$, so folgt $\tilde{J} \subseteq J$ und $x(t) = \tilde{x}(t)$ für alle $t \in \tilde{J}$.

Satz B.15 (Eindeutigkeit)**.** Es sei (B.1) eine Caratheodory-Gleichung und die rechte Seite $f(x,t)$ erfülle die lokale Lipschitzbedingung (L1). Dann existiert eine maximale Lösung $\rho : J \to \mathbb{R}^n$ von (AWP) bezüglich I auf einem nichtleeren Intervall $J \subseteq I$, das relativ offen zu I ist.

Beweis. 1. Es seien $x_1, x_2 : \tilde{J} \to \mathbb{R}^n$ zwei beliebige Lösungen von (AWP) auf $\tilde{J} \subseteq I$. Die Anfangszeit $t_0 \in \tilde{J}$ liege nicht am rechten Rand von $\tilde{J}$. Wir zeigen zunächst, dass es ein $d > 0$, so dass x_1 und x_2 auf $[t_0, t_0 + d]$ wohldefiniert sind und übereinstimmen. Um dies nachzuweisen, genügt es vorauszusetzen, dass anstelle von (L1) die Ungleichung

$$\frac{(x-y)^*(f(x,t)-f(y,t))}{\|x-y\|^2} \leq \mu(t) \quad \forall x, y \in B_\delta(x_0), t \in I \tag{B.9}$$

erfüllt ist. Wegen der Ungleichung von Cauchy-Schwartz ist

$$(x-y)^*(f(x,t)-f(y,t)) \leq \|f(x,t)-f(y,t)\| \cdot \|x-y\| \ ,$$

weshalb (B.9) aus der Lipschitz-Bedingung (L1) folgt. Aufgrund der Stetigkeit von x_1 und x_2 im Punkt t_0 können wir ein $d > 0$ bestimmen, so dass x_1, x_2 auf $[t_0, t_0 + d]$ definiert sind (d.h. $t_0 + d \in \tilde{J}$) und

$$\|x_1(t) - x_0\| < \delta \ , \ \|x_2(t) - x_0\| < \delta \quad \forall t \in [t_0, t_0 + d] \ .$$

Wir führen die Hilfsfunktionen

$$z(t) := x_1(t) - x_2(t), r(t) := \|z(t)\|^2 \text{ und } L(t) := \int_{t_0}^{t} \mu(s)\mathrm{ds}$$

mit Definitionsbereich $[t_0, t_0 + d]$ ein, wobei μ die lokal integrierbare Funktion aus (L1). Die Funktion $t \mapsto r(t)e^{-2L(t)}$ ist absolut stetig (siehe Satz A.17) und es gilt fast überall:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{dt}} r(t) = 2z(t)^* \ \dot{z}(t) &= 2(f(x_1(t),t) - f(x_2(t),t))^* \ (x_1(t) - x_2(t)) \\ &\overset{(B.9)}{\leq} 2\mu(t) r(t) \quad \forall t \in [t_0, t_0 + d] \end{aligned} \tag{B.10}$$

bzw.

$$\frac{\mathrm{d}}{\mathrm{dt}}(r(t)e^{-2L(t)}) = \frac{\mathrm{d}}{\mathrm{dt}}r(t)\; e^{-2L(t)} + r(t)e^{-2L(t)}\left(-2\frac{\mathrm{d}}{\mathrm{dt}}\int_{t_0}^{t}\mu(s)\mathrm{ds}\right) \tag{B.11}$$
$$= \underbrace{e^{-2L(t)}}_{\geq 0}\underbrace{\left(\frac{\mathrm{d}}{\mathrm{dt}}r(t) - 2\mu(t)r(t)\right)}_{\overset{(B.10)}{\leq} 0} \leq 0\ .$$

Die Funktion $r(t)e^{-2L(t)}$ ist also positiv und monoton fallend. Aus $z(t_0) = 0$ folgt somit $z(t) = 0$ für alle $t \in [t_0, t_0 + d]$, was zu zeigen war.

2. Wir wollen nun Eindeutigkeit für alle $t \geq t_0$ beweisen. Angenommen, es gibt ein $t \in \tilde{J}$ mit $t > t_0$ und $x_1(t) \neq x_2(t)$. Dann ist

$$t_1 := \inf\{t \in \tilde{J} \mid t > t_0,\ x_1(t) \neq x_2(t)\}$$

eine reelle Zahl in (t_0, ∞), und es gilt $x_1 = x_2$ auf $[t_0, t_1)$. Da x_1, x_2 stetig sind im Punkt t_1, gilt auch $x_1(t_1) = x_2(t_1)$. Wir beachten, dass die Bedingung (L1) für alle Anfangswerte $x_0 \in \mathbb{R}^n$ erfüllt ist, und dass t_1 kein rechter Randpunkt von $\tilde{J}$ sein kann. Wir dürfen also 1. auf die neue Anfangsbedingung $x(t_1) = \tilde{x}_0$ mit Anfangswert $\tilde{x}_0 := x_1(t_1)$ anwenden, um ein Intervall $[t_1, t_1 + d]$ für ein $d > 0$ zu erhalten, auf dem x_1 und x_2 übereinstimmen. Dies ist ein Widerspruch zur Minimalität von t_1.

3. Um Eindeutigkeit für $t \leq t_0$ zu erhalten, wiederholen wir die Argumentation aus 1. und 2. für das zeitumgekehrte Problem (B.4).

4. Es bleibt zu zeigen, dass es die maximale Lösung $\rho : J \to \mathbb{R}^n$ gibt. Wir setzen

$$\tau_{\min} := \inf\{t \in I \mid \text{es existiert ein Lösung von (AWP) auf } [t, t_0]\} \tag{B.12}$$

und

$$\tau_{\max} := \sup\{t \in I \mid \text{es existiert ein Lösung von (AWP) auf } [t_0, t]\}\ .$$

(Möglicherweise ist $\tau_{\min} = -\infty$ und $\tau_{\max} = \infty$.)
Aus Satz B.9 folgt $\tau_{\min} < \tau_{\max}$. Das Intervall $(\tau_{\min}, \tau_{\max})$ ist also nichtleer. Laut Definition (B.12) gibt es zwei Folgen (s_n) und (t_n) mit $s_n \to \tau_{\min}$ bzw. $t_n \to \tau_{\max}$ für $n \to \infty$, so dass auf jedem Intervall (s_n, t_n) eine Lösung von (AWP) existiert. Wie in 1.-3. festgestellt, stimmen diese Lösungen auf gemeinsamen Gebieten des Definitionsbereichs überein. Es gibt also eine

Lösung ρ auf $J := (\tau_{\min}, \tau_{\max})$.
Um die Maximalität von ρ zu erreichen, müssen wir gegenfalls die Randpunkte von I—falls vorhanden—in den Definitionsbereich J aufnehmen. Falls $\tau_{\min}$ ein linker Randpunkt von I ist und es eine Lösung von (AWP) auf einem $\tau_{\min}$ enthaltenden Intervall existiert, so füge $\tau_{\min}$ dem Intervall J hinzu und setze ρ auf J stetig fort. Ist $\tau_{\max}$ ein rechter Randpunkt von I, so verfahren wir entsprechend.
Per Konstruktion ist J ein nichtleeres Intervall, das offen ist relativ zu I. Angenommen, es ist $\tau_{\min}$ ein innerer Punkt von I und es gibt eine Lösung von (AWP) auf $[\tau_{\min}, t_0]$. Dann könnte diese Lösung mit dem lokalen Existenzsatz B.9 auf ein Intervall $(\tau_{\min} - d, \tau_{\min})$ für ein $d > 0$ erweitert werden, im Widerspruch zur Minimalität von $\tau_{\min}$. Ähnliches gilt für $\tau_{\max} \in \operatorname{int}(I)$. Dies impliziert, dass der Definitionsbereich von Lösungen des Anfangswertproblems stets im konstruiertem J enthalten sind. Aus 1.-3. folgt, dass die Lösungen auf ihrem Definitionsbereich mit ρ übereinstimmen. □

Folgerung B.16. Die rechte Seite $f(x,t)$ der Caratheodory-Gleichung (B.1) erfülle die „globale Lipschitzbedingung in x“:

(L2) Es gibt eine lokal integrierbare Funktion $\mu : I \to \mathbb{R}_+$, so dass

$$\|f(x,t) - f(y,t)\| \leq \mu(t)\|x - y\| \qquad \forall x, y \in \mathbb{R}^n, t \in I \ .$$

Dann existiert eine maximale Lösung $\rho : I \to \mathbb{R}^n$ von (AWP), die auf ganz I definiert ist (d.h. $J = I$ in Satz B.15).

Beweis. Wegen der Dreiecksungleichung ist

$$\|f(x,t)\| - \|f(y,t)\| \leq \|f(x,t) - f(y,t)\| \leq \mu(t)\|x - y\| \ \ \forall x, y \in \mathbb{R}^n, t \in I \ .$$

Wir setzen $y := 0$. Es folgt

$$\frac{\|f(x,t)\|}{\|x\|} \leq \mu(t) + \|f(0,t)\| \ \ \forall \|x\| \geq 1, t \in I \ .$$

Aufgrund der Caratheodory-Bedingungen B.2 ist $t \mapsto \|f(0,t)\|$ lokal integrierbar, und somit auch die gesamte rechte Seite der Ungleichung. $f(x,t)$ erfüllt also die globale Existenzbedingung (B.8). Die maximale Lösung aus Satz B.15 kann folglich auf ganz I erweitert werden. □

Bemerkung B.17. Eine Caratheodory-Gleichung der Form

$$\dot{x}(t) = f(x(t), t), \qquad f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \ , \tag{B.13}$$

welche die lokale Lipschitzbedingung in x erfüllt, definiert ein kontinuierliches System Σ_f ohne Kontrolle:
Man wählt den Zustandsraum $\mathcal{X} = \mathbb{R}^n$, die Zeitmenge $\mathcal{T} = \mathbb{R}$ und den einelementigen Steuerbereich $\mathcal{U} = \{0\}$. Auf der Menge

$$D(\phi) := \big\{(t, t_0, x_0)\ |t, t_0 \in \mathbb{R}, t_0 \le t, x_0 \in \mathbb{R}^n, \text{ es existiert eine Lösung } x(t; t_0, x_0) : [t_0, t] \to \mathbb{R}^n \text{ von (B.13) mit } x(t_0) = x_0\big\}$$

definiert man die Übergangsfunktion $\phi(t, t_0, x_0) := x(t; t_0, x_0)$.
Diese ist wohldefiniert, da nach Voraussetzung $x(t; t_0, x_0)$ eindeutig bestimmt ist, und nichttrivial aufgrund des Satzes B.9 zur lokalen Existenz von Caratheodory-Lösungen.
Es ist leicht einsehbar, dass das Tupel $\Sigma_f := (\mathbb{R}, \mathbb{R}^n, \{0\}, \phi)$ auch die restlichen Axiome der Systemdefinition 1.2 erfüllt. In diesem Sinne dürfen wir „$\dot{x} = f(x, t)$“ als (Caratheodory-) System bezeichnen.

B.2 Lineare Systeme

Homogene Systeme

Wir untersuchen das lineare System ohne Kontrolle

$$\dot{x}(t) = A(t)x(t), \qquad t \in I, \qquad x(t) \in \mathbb{R}^n \tag{B.14}$$

und das „assoziierte Matrixsystem“

$$\dot{X}(t) = A(t)X(t), \qquad t \in I, \qquad X(t) \in \mathbb{R}^{n\times n}\ , \tag{B.15}$$

wobei $A : I \mapsto \mathbb{R}^{n\times n}$ eine messbare und lokal (essentiell) beschränkte Funktion auf einem reellen Intervall I ist. Zur Vereinfachung können wir $I = \mathbb{R}$ voraussetzen, indem wir $A(t) := 0$ für alle $t \in I^c$ setzen.
Die rechte Seite $f(x, t) = A(t)x$ von (B.14) ist linear in x für alle t, und lokal integrabel in t für alle x. Offenbar erfüllt sie die Caratheodory-Bedingungen B.2 (wähle $m(t) := \|A(t)x\|$ in (C3')) und die globale Lipschitzbedingung (L2) in x (wähle $\mu(t) := \|A(t)\|$). Da das assoziierte Matrixsystem äquivalent zu dem Vektorsystem (in $\mathbb{R}^{n^2}$)

$$\dot{z}(t) = \tilde{A}(t)z(t), \qquad z(t) \in \mathbb{R}^{n^2}$$

mit

$$\tilde{A}(t) := \begin{pmatrix} A(t) & & & \\ & A(t) & & \\ & & \ddots & \\ & & & A(t) \end{pmatrix}$$

ist, erfüllt auch (B.15) diese Bedingungen. Es folgt aus Lemma B.16, dass es zu jedem Anfangswertproblem zu (B.14) bzw. (B.15) eine (eindeutige) maximale Lösung gibt.
Von besonderer Bedeutung ist die sogenannte *Fundamentallösung* $\Phi(\cdot\,;t_0)$. Das ist die maximale Lösung des Anfangswertproblems

$$\dot{X} = A(t)X, \qquad X(t_0) = \mathrm{I} \tag{AWP}$$

mit Anfangswert I (Einheitsmatrix) und einer Anfangszeit $t_0 \in \mathbb{R}$. Wie das folgende Lemma zeigt, ist durch die Kenntnis der Fundamentallösung jedes Anfangswertproblem zu (B.14) bzw. (B.15) mit Anfangszeit t_0 gelöst.

Satz B.18. Gegeben seien die Anfangsdaten $t_0 \in I$, $P \in \mathbb{R}^{n\times n}$ und $p \in \mathbb{R}^n$. Dann gilt:

(a) $X(t) := \Phi(t;t_0)\cdot P$ ist die Lösung von (B.15) mit $X(t_0) = P$.

(b) $x(t) := \Phi(t;t_0)\cdot p$ ist die Lösung von (B.14) mit $x(t_0) = p$.

Beweis. Natürlich sind mit $\Phi(\cdot;t_0)$ auch $X(\cdot)$ bzw. $x(\cdot)$ absolut stetig, und erfüllen die Anfangsbedingung. Weiter gilt wegen der Produktregel:

(i) $\dot{X}(t) = (\Phi(t;t_0)P)' = \dot{\Phi}(t;t_0)P + \Phi(t;t_0)\dot{P} = A(t)\Phi(t;t_0)P = A(t)X(t)$ fast überall.

(ii) $\dot{x}(t) = (\Phi(t;t_0)p)' = \dot{\Phi}(t;t_0)p = A(t)\Phi(t;t_0)p = A(t)x(t)$ f.ü. □

Im nächsten Lemma werden Eigenschaften der Fundamentallösung aufgelistet.

Lemma B.19. Es gilt für alle $t, t_0, t_1 \in \mathbb{R}$:

(i) $\Phi(t_0;t_0) = \mathrm{I}$.

(ii) Die Matrix $\Phi(t;t_0)$ hat vollen Rang.

(iii) $\Phi(t;t_0) = \Phi(t;t_1)\cdot\Phi(t_1;t_0)$.

(iv) $\Phi(t_1;t_0)^{-1} = \Phi(t_0;t_1)$.

(v) $\det\Phi(t;t_0) = \exp\left(\int_{t_0}^{t}\mathrm{tr}A(s)\mathrm{d}s\right)$.
(Dabei sei $\int_{t_0}^{t}\mathrm{tr}A(s)\mathrm{d}s = -\int_{t}^{t_0}\mathrm{tr}A(s)\mathrm{d}s$ für $t_0 > t$.)

(vi) $\frac{\mathrm{d}}{\mathrm{d}t_0}\Phi(t;t_0) = -\Phi(t;t_0)A(t_0)$ fast überall.

Beweis. Die Eigenschaften (ii) und (v) folgen direkt aus dem Satz von Liousville B.25. Um (iii) zu zeigen, stellen wir fest, dass beide Seiten der Gleichung die (eindeutige) maximale Lösung des Anfangswertproblems

$$\dot{X} = A(t)X, \qquad X(t_1) = \Phi(t_1; t_0)$$

darstellen. (iii) impliziert (iv) für $t = t_0$. Die letzte Eigenschaft erhält man durch Auflösen in

$$0 = \frac{\mathrm{d}}{\mathrm{d}t_0}(\Phi(t;t_0)\cdot\Phi(t_0;t)) = \frac{\mathrm{d}}{\mathrm{d}t_0}\Phi(t;t_0)\cdot\Phi(t_0;t) + \Phi(t;t_0)\cdot A(t_0)\Phi(t_0;t) \quad \text{f.ü.}$$

□

Im Spezialfall können wir eine explizite Darstellung der Fundamentallösung angeben:

Satz B.20. Es ist $\Phi(t;t_0) = \exp\int_{t_0}^{t} A(s)\mathrm{d}s$, falls

$$A(t_1)A(t_2) = A(t_2)A(t_1) \qquad \forall t_1, t_2 \in I\ . \tag{B.16}$$

Beweisskizze. 1. Die Reihe

$$\exp\int_{t_0}^{t} A(s)\mathrm{d}s = \sum_{k=0}^{\infty}\frac{1}{k!}\left(\int_{t_0}^{t} A(s)\mathrm{d}s\right)^k$$

konvergiert absolut und gleichmäßig. Wir dürfen sie gliedweise ableiten. Weiter ist sie absolut stetig. Wir setzen $X(t) := \exp\int_{t_0}^{t} A(s)\mathrm{d}s$ und zeigen, dass $X(\cdot\,)$ das Anfangswertproblem (AWP) löst.

2. Aus (B.16) folgt für alle $t_1, t_2 \in I$

$$\begin{aligned}\int_{t_0}^{t_1} A(s_1)\mathrm{d}s_1 \int_{t_0}^{t_2} A(s_2)\mathrm{d}s_2 &= \int_{t_0}^{t_1}\int_{t_0}^{t_2} A(s_1)A(s_2)\mathrm{d}s_1\mathrm{d}s_2 \\ &= \int_{t_0}^{t_2}\int_{t_0}^{t_1} A(s_2)A(s_1)\mathrm{d}s_2\mathrm{d}s_1 = \int_{t_0}^{t_2} A(s_2)\mathrm{d}s_2 \int_{t_0}^{t_1} A(s_1)\mathrm{d}s_1\ .\end{aligned}$$

Dies impliziert fast überall

$$\frac{\mathrm{d}}{\mathrm{d}t_1}\left(\int_{t_0}^{t_1} A(s_1)\mathrm{d}s_1 \int_{t_0}^{t_2} A(s_2)\mathrm{d}s_2\right) = \frac{\mathrm{d}}{\mathrm{d}t_1}\left(\int_{t_0}^{t_2} A(s_2)\mathrm{d}s_2 \int_{t_0}^{t_1} A(s_1)\mathrm{d}s_1\right)$$

und somit (Produktregel)

$$A(t_1)\int_{t_0}^{t_2} A(s_2)\mathrm{d}s_2 = \int_{t_0}^{t_2} A(s_2)\mathrm{d}s_2\ A(t_1) \qquad \forall t_1, t_2 \in I\ .$$

3. Das Kommutieren von $A(t)$ mit seinem Integral liefert zusammen mit der Produktregel

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}}\left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^k &= A(t)\left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^{k-1} \\ &+ \int_{t_0}^{t} A(s)\mathrm{ds}\, A(t)\left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^{k-2} + \left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^2 A(t)\left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^{k-3} \\ &+ \ldots + \left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^{k-1} A(t) = kA(t)\left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^{k-1}\end{aligned}$$

für fast alle $t \in I$.

4. Schließlich nutzen wir 1. und 3., um $X(\cdot\,)$ abzuleiten:

$$\begin{aligned}\dot{X}(t) &= \sum_{k=1}^{\infty} \frac{1}{k!} kA(t)\left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^{k-1} = A(t)\sum_{k=0}^{\infty}\frac{1}{k!}\left(\int_{t_0}^{t} A(s)\mathrm{ds}\right)^{k} \\ &= A(t)X(t)\end{aligned}$$

fast überall. Es folgt die Behauptung $X(t) = \Phi(t;t_0)$. □

Beispiel B.21. Ist $A(t) \equiv A$, so ist $\Phi(t;t_0) = e^{A(t-t_0)}$.

Im allgemeinen Fall können wir zumindest eine Potenzreihenentwicklung der Fundamentallösung angegeben:

Satz B.22 (Picard-Lindelöf)**.** Es sei $X(t) := \Phi(t;t_0)$ die Fundamentallösung von (B.15) auf einem Intervall $[t_0, T]$ (für ein $T \geq 0$). Mit dem HDI A.14 lässt sich die folgende Reihenentwicklung der Fundamentallösung zu einem festen Zeitpunkt $t \in [t_0, T]$ herleiten:

$$\begin{aligned}X(t) &= \mathrm{I} + \int_{t_0}^{t} A(s_1)X(s_1)\mathrm{ds}_1 = \mathrm{I} + \int_{t_0}^{t} A(s_1)\left(\mathrm{I} + \int_{t_0}^{s_1} A(s_2)X(s_2)\mathrm{ds}_2\right)\mathrm{ds}_1 \\ &= \mathrm{I} + \int_{t_0}^{t} A(s_1)\mathrm{ds}_1 + \int_{t_0}^{t}\int_{t_0}^{s_1} A(s_1)A(s_2)X(s_2)\mathrm{ds}_2\mathrm{ds}_1 = \ldots \\ &= \mathrm{I} + \sum_{k=1}^{l}\int_{t_0}^{t}\int_{t_0}^{s_1}\int_{t_0}^{s_2}\ldots\int_{t_0}^{s_{k-1}} A(s_1)A(s_2)\ldots A(s_k)\mathrm{ds_k}\ldots\mathrm{ds}_2\mathrm{ds}_1 + \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s_1}\int_{t_0}^{s_2}\ldots\int_{t_0}^{s_l} A(s_1)A(s_2)\ldots A(s_{l+1})X(s_{l+1})\mathrm{ds_{l+1}}\ldots\mathrm{ds}_2\mathrm{ds}_1\end{aligned}$$

Wir beweisen nun, dass man diese Entwicklung unendlich oft fortsetzen darf. Die entstehende Reihe nennt man *Neumann-Reihe*. Sie konvergiert absolut und gleichmäßig auf $[t_0, T]$. Also

$$X(t) = \mathrm{I} + \sum_{k=1}^{\infty} \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \dots \int_{t_0}^{s_{k-1}} A(s_1)A(s_2)\dots A(s_k)\mathrm{ds_k}\dots\mathrm{ds_2ds_1} \tag{B.17}$$

für alle $t \in [t_0, T]$.
Für den Beweis konstruieren wir Hilfsfunktionen—die sog. „Picard-Iterierten“

$$Y_0 \equiv I, \qquad Y_{l+1}(t) = I + \int_{t_0}^{t} A(s)Y_l(s)\mathrm{ds} \qquad (l = 0, 1, 2\dots)\ . \tag{B.18}$$

Die Iterierten entsprechen genau den Partialsummen der Neumann-Reihe, denn

$$\begin{aligned} Y_l(t) &= \mathrm{I} + \sum_{k=1}^{l} (Y_k(t) - Y_{k-1}(t)) \\ &= \mathrm{I} + \sum_{k=1}^{l} \int_{t_0}^{t} A(s_1)(Y_{k-1}(s_1) - Y_{k-2}(s_1))\mathrm{ds_1} \\ &= \mathrm{I} + \sum_{k=1}^{l} \int_{t_0}^{t} \int_{t_0}^{s_1} A(s_1)A(s_2)(Y_{k-2}(s_2) - Y_{k-3}(s_2))\mathrm{ds_2ds_1} \\ &\dots \\ &= \mathrm{I} + \sum_{k=1}^{l} \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \dots \int_{t_0}^{s_{k-1}} A(s_1)A(s_2)\dots A(s_k)\mathrm{ds_k}\dots\mathrm{ds_2ds_1}\ . \end{aligned} \tag{B.19}$$

Es genügt daher zu zeigen:

$$X = \lim_{l\to\infty} Y_l \tag{B.20}$$

bzgl. der Supremumsnorm $\|\cdot\|_0$ im Raum $C^0([t_0, T], \mathbb{R}^{n\times n})$.

Beweis. Wir setzen $\alpha := \operatorname{ess.sup}_{t\in[t_0,T]} \|A(t)\|$.

1. Per Induktion über l folgt, dass der Integrand $A(s)Y_l(s)$ in (B.18) als messbare (essentiell) beschränkte Funktion über $[t_0, T]$ integrierbar ist. Der HDI sichert die Stetigkeit von Y_l $(l = 1, 2, \dots)$.

2. Es gilt

$$\begin{aligned}
&\|Y_k(t) - Y_{k-1}(t)\| \\
&\overset{(B.19)}{=} \left\| \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{0}^{s_2} \dots \int_{t_0}^{s_{k-1}} A(s_1)A(s_2)\dots A(s_k)\mathrm{ds_k}\dots \mathrm{ds_2}\mathrm{ds_1} \right\| \\
&\leq \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \dots \int_{t_0}^{s_{k-1}} \|A(s_1)\| \cdot \|A(s_2)\| \cdots \|A(s_k)\| \mathrm{ds_k}\dots \mathrm{ds_2}\mathrm{ds_1} \\
&\leq \alpha^k \frac{t^k}{k!}
\end{aligned}$$

für alle $t \in [t_0, T]$ und $k \geq 1$. Wir können jetzt die Summanden von (B.19) in der Supremumsnorm abschätzen:

$$\begin{aligned}
\|\mathrm{I}\|_0 &= \sup_{t\in[t_0,T]} \|\mathrm{I}\| = \sup_{t\in[t_0,T]} \sup_{\|x\|=1} \|\mathrm{I}x\| = \sup_{\|x\|=1} \|x\| = 1 \ , \\
\|Y_k - Y_{k-1}\|_0 &= \sup_{t\in[t_0,T]} \|Y_k(t) - Y_{k-1}(t)\| \leq \alpha^k \frac{T^k}{k!} \ .
\end{aligned}$$

3. Es folgt

$$\|\mathrm{I}\|_0 + \sum_{k=1}^{\infty} \|Y_k - Y_{k-1}\|_0 \leq \exp(\alpha T) < \infty \ .$$

Die Folge der Partialsummen $(Y_k)_{k\in\mathbb{N}}$ bildet daher eine Cauchy-Folge im Banachraum $C^0([t_0, T], \mathbb{R}^{n\times n})$, und konvergiert dort gegen eine stetige Funktion Y, die sich durch die Neumann-Reihe (B.17) darstellen lässt. Wir schliessen $X = Y$. □

Bemerkung B.23. $\tilde{\Phi}(\cdot\,; -t_0)$ bezeichne die Fundamentallösung des zeitumgekehrten linearen Systems

$$\dot{X}(t) = -A(-t)X(t)$$

auf dem Intervall $[-t_0, -T]$ (für ein $T < t_0$). Aus Satz B.22 und Hilfssatz B.8 erhalten wir für alle $t \in [T, t_0]$:

$$\begin{aligned}
&\Phi(t; t_0) = \tilde{\Phi}(-t; -t_0) \\
&= \mathrm{I} + \sum_{k=1}^{\infty} \int_{-t_0}^{-t} \int_{-t_0}^{s_1} \int_{-t_0}^{s_2} \dots \int_{-t_0}^{s_{k-1}} (-1)^k A(-s_1)A(-s_2)\dots A(-s_k)\mathrm{ds_k}\dots \mathrm{ds_1} \\
&\overset{\text{Sub.}}{=} \mathrm{I} + \sum_{k=1}^{\infty} \int_{t}^{t_0} \int_{s_1}^{t_0} \int_{s_2}^{t_0} \dots \int_{s_{k-1}}^{t_0} (-1)^k A(s_1)A(s_2)\dots A(s_k)\mathrm{ds_k}\dots \mathrm{ds_2}\mathrm{ds_1}
\end{aligned}$$

Für alle $a > b$ in $\mathbb{R}$ setzen wir

$$\int_a^b f(x,s)\mathrm{ds} := -\int_b^a f(x,s)\mathrm{ds}\ .$$

Auf diese Weise behält die Darstellung (B.17) auch für $t < t_0$ ihre Gültigkeit. Da T beliebig vorgegeben, folgt für **alle** $t \in \mathbb{R}$:

$$\Phi(t;t_0) = \mathrm{I} + \sum_{k=1}^{\infty} \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \dots \int_{t_0}^{s_{k-1}} A(s_1)A(s_2)\dots A(s_k)\mathrm{ds_k}\dots \mathrm{ds_2}\mathrm{ds_1}$$

Definition B.24. Für eine vorgegebene Matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ bezeichne $\mathrm{tr}A := \sum_{i=1}^n a_{ii}$ die *Spur von A*.

Satz B.25 (Liouville)**.** Es sei $X : [t_0, T] \to \mathbb{R}^{n\times n}$ eine Lösung von (B.15). Dann gilt

$$\det X(t) = \exp\left(\int_{t_0}^{t} \mathrm{tr}A(s)\mathrm{ds}\right) \det X(t_0) \qquad \forall t \in [t_0, T]\ . \tag{B.21}$$

Beweis. 1. Betrachte die skalare Differentialgleichung

$$\dot{z}(t) = \mathrm{tr}A(t)\cdot z(t), \qquad t \in [t_0, T]\ . \tag{B.22}$$

Wegen Satz B.20 ist $z(t) = \exp\int_{t_0}^t \mathrm{tr}A(s)\mathrm{ds}\cdot z_0$ die eindeutige Lösung zum Anfangswert $z_0 := \det X(t_0)$. Da $\det X(\cdot\,)$ absolut stetig ist, genügt es zu zeigen, dass $\det X(\cdot\,)$ die Gleichung (B.22) fast überall erfüllt, und daher mit $z(\cdot\,)$ übereinstimmt. Es soll also

$$\frac{\mathrm{d}}{\mathrm{dt}} \det X(t) = \mathrm{tr}A(t)\cdot \det X(t_0) \tag{B.23}$$

für fast alle $t \in [t_0, T]$ gelten.

2. Es sollen a_{ij} und ρ_{ij} die Koeffizienten (abhängig von t) von A bzw. X bezeichnen. Dann wird $\dot{X} = AX$ zu

$$\dot{\rho}_{ij} = \sum_{k=1}^{n} a_{ik}\rho_{kj}, \qquad \forall\ i,j = 1,\dots n\ . \tag{B.24}$$

Weiter gilt

$$\det X = \sum_{\pi\in \mathrm{S}_n} \epsilon(\pi)\rho_{1,\pi(1)}\dots\rho_{n,\pi(n)}\ , \tag{B.25}$$

wobei $\epsilon(\pi)$ das Signum der Permutation π ist. Mit Hilfe der Produktregel folgt fast überall

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}}\det X &= \sum_{\pi\in \mathrm{S}_n}\epsilon(\pi)\dot\rho_{1,\pi(1)}\rho_{2,\pi(2)}\cdots\rho_{n,\pi(n)} + \sum_{\pi\in \mathrm{S}_n}\epsilon(\pi)\rho_{1,\pi(1)}\dot\rho_{2,\pi(2)}\cdots\rho_{n,\pi(n)}\\ &\quad + \ldots + \sum_{\pi\in \mathrm{S}_n}\epsilon(\pi)\rho_{1,\pi(1)}\cdots\rho_{n-1,\pi(n-1)}\dot\rho_{n,\pi(n)}\\ &= \det\begin{pmatrix}\dot\rho_{11} & \dot\rho_{12} & \cdots & \dot\rho_{1n}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n}\\ \vdots & \vdots & \ldots & \vdots\\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}\end{pmatrix} + \det\begin{pmatrix}\rho_{11} & \rho_{12} & \cdots & \rho_{1n}\\ \dot\rho_{21} & \dot\rho_{22} & \cdots & \dot\rho_{2n}\\ \vdots & \vdots & \ldots & \vdots\\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}\end{pmatrix}\\ &\quad + \ldots + \det\begin{pmatrix}\rho_{11} & \rho_{12} & \cdots & \rho_{1n}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n}\\ \vdots & \vdots & \ldots & \vdots\\ \dot\rho_{n1} & \dot\rho_{n2} & \cdots & \dot\rho_{nn}\end{pmatrix}\ .\end{aligned}$$

Wir haben soeben $\det X$ als Summe von n Determinanten dargestellt. Mittels (B.24) erhalten wir für die erste Determinante

$$\det\begin{pmatrix}\sum_{k=1}^n a_{1k}\rho_{k1} & \sum_{k=1}^n a_{1k}\rho_{k2} & \cdots & \sum_{k=1}^n a_{1k}\rho_{kn}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n}\\ \vdots & \vdots & \ldots & \vdots\\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}\end{pmatrix}\ .$$

Diese Determinante ist invariant unter den folgenden (elementaren) Zeilenumformungen. Wir ziehen von der ersten Zeile a_{12} mal die zweite Zeile, a_{13} mal die dritte Zeile, usw., und a_{1n} mal die letzte Zeile ab. Dies ergibt (fast überall)

$$\det\begin{pmatrix}\dot\rho_{11} & \dot\rho_{12} & \cdots & \dot\rho_{1n}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n}\\ \vdots & \vdots & \ldots & \vdots\\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}\end{pmatrix} = \det\begin{pmatrix}a_{11}\rho_{11} & a_{11}\rho_{12} & \cdots & a_{11}\rho_{1n}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n}\\ \vdots & \vdots & \ldots & \vdots\\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}\end{pmatrix} = a_{11}\det X\ .$$

Indem wir bei den restlichen Determinanten analog vorgehen, gewinnen wir (B.23). □

Bemerkung B.26. Aus der Gleichung (B.21) folgt (wegen exp > 0), dass abhängig vom Anfangswert $X(t_0)$ alle Werte $X(t)$ entweder singulär sind oder nicht. Das Vorzeichen der Determinante ändert sich nie.
Insbesondere hat die Matrix $\Phi(t;t_0)$ für alle t vollen Rang.

Inhomogene Systeme

Schließlich berücksichtigen wir noch das inhomogene lineare System

$$\dot{x}(t) = A(t)x(t) + b(t), \qquad t \in I, \qquad x(t) \in \mathbb{R}^n \tag{B.26}$$

mit messbaren, lokal (essentiell) beschränkten Funktionen $A : I \to \mathbb{R}^{n\times n}$ und $b : I \to \mathbb{R}^n$.
Offensichtlich behalten die Existenz- und Eindeutigkeitsbedingungen (Caratheodory-Bedingungen, globale Lipschitzbedingung) beim Übergang zu den inhomogenen Systemen ihre Gültigkeit.

Satz B.27 (Variation der Konstanten)**.** Sei $t_0 \in I$. Die Funktion $x : I \to \mathbb{R}^n$ mit

$$x(t) := \Phi(t;t_0)\left(p + \int_{t_0}^{t} \Phi(t_0;s)b(s)\mathrm{ds}\right) = \Phi(t;t_0)p + \int_{t_0}^{t} \Phi(t;s)b(s)\mathrm{ds} \tag{B.27}$$

ist die Lösung von (B.26) mit $x(t_0) = p$.

Beweis. Wegen der Stetigkeit von $\Phi(t;\cdot)$ und der lokalen Integrierbarkeit von $b(\cdot)$ ist der Integrand lokal integrierbar, und (B.27) ist wohldefiniert.
Da $\Phi(\cdot;t_0)$ invertierbar ist, können wir ohne Einschränkung annehmen, dass die Lösung von (B.26) zur Anfangsbedingung $x(t_0) = p$ von der Form $x(\cdot) = \Phi(\cdot;t_0)z(\cdot)$ ist, wobei $z : I \to \mathbb{R}^n$ eine absolut stetige Funktion ist. Ableiten von x ergibt

$$\begin{aligned} Ax(t) + b(t) \overset{\text{HDI}}{=} \dot{x}(t) &= \dot{\Phi}(t;t_0)z(t) + \Phi(t;t_0)\dot{z}(t) \\ &= A(t)\underbrace{\Phi(t;t_0)z(t)}_{x(t)} + \Phi(t;t_0)\dot{z}(t) \end{aligned}$$

für fast alle $t \in I$. Auflösen nach $\dot{z}$ ergibt $\dot{z}(t) = \Phi(t;t_0)^{-1}b(t) = \Phi(t_0;t)b(t)$ fast überall. Nun verwenden wir den HDI, um

$$z(t) = \underbrace{z(t_0)}_{p} + \int_{t_0}^{t} \Phi(t_0;s)b(s)\mathrm{ds}$$

zu erhalten. □

B.3 Autonome Kontrollsysteme mit kompaktem Steuerbereich

Es werden Kontrollsysteme analysiert, die durch Differentialgleichungen der Form

$$\dot{x}(t) = f(x(t), u(t)), \qquad u(t) \in \mathcal{U}, \qquad x(t) \in \mathbb{R}^n \tag{B.28}$$

gegeben sind, wobei $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ eine stetige Funktion ist und der Steuerbereich $\mathcal{U} \subseteq \mathbb{R}^m$ eine kompakte (nichtleere) Menge.
Es wird insbesondere vorausgesetzt, dass die Eingabegrößen des Systems beschränkt sind, was bei den meisten „realen“ Systemen aus Physik, Biologie, etc. ohnehin der Fall ist. So ist z.B. die Beschleunigung eines Autos durch die Leistungsfähigkeit des Motors beschränkt oder die Passagierzahl in einem Flugzeug durch die Anzahl der Sitzplätze.

Definition B.28. Eine messbare Abbildung $u : I \to \mathcal{U}$ (I Intervall in $\mathbb{R}$) heißt *zulässige* Kontrollfunktion bzgl. des Systems (B.28).

Lemma B.29. Ist $u : I \to \mathcal{U}$ eine zulässige Kontrollfunktion und $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ stetig, so erfüllt die Funktion

$$g(x, t) := f(x, u(t))$$

die Caratheodory-Bedingungen B.2 auf dem Gebiet $\mathbb{R}^n \times I$. In diesem Sinne ist (B.28) eine Caratheodory-Gleichung. Die Theorie aus dem ersten Abschnitt lässt sich übertragen.

Beweisskizze. **(i)** $g(x, t)$ ist stetig in x für alle t, da $f(x, u)$ stetig ist in x für alle $u \in \mathcal{U}$.

(ii) $g(x, t)$ ist messbar in t für jedes x, da

$$t \overset{\text{messbar}}{\mapsto} u(t) \overset{\text{stetig}}{\mapsto} f(x, u(t))$$

messbar ist.

(iii) Es sei ein Kompaktum $K \subset \mathbb{R}^n$ vorgegeben. $f(x, u)$ ist stetig auf der kompakten Menge $K \times \mathcal{U}$ und folglich dort beschränkt. Also existiert ein $M > 0$, so dass

$$\|g(x, t)\| = \|f(x, \underbrace{u(t)}_{\in \mathcal{U}})\| \leq M \qquad \forall t \in I \ .$$

□

Definition B.30. $x : I \to \mathbb{R}^n$ heißt *zulässige* Lösung von (B.28) bezüglich der (zulässigen) Kontrolle $u : I \to \mathcal{U}$, falls $x(\cdot)$ eine Caratheodory-Lösung von (B.28) ist, d.h. $x(\cdot)$ ist lokal absolut stetig mit

$$\dot{x}(t) = f(x(t), u(t)) \qquad \text{für fast alle } t \in I.$$

Insbesondere ist dies genau dann der Fall, wenn $t \mapsto f(x(t), u(t))$ lokal integrierbar ist über I, und der Integralgleichung

$$x(t) = x(s) + \int_a^t f(x(s), u(s))ds \qquad \forall a, t \in I \tag{B.29}$$

(dabei sei $\int_a^t = -\int_t^a$ für $a > t$) genügt.

Wir sagen, $x : J \to \mathbb{R}^n$ ist eine (zulässige) Lösung des Anfangswertproblems

$$\dot{x}(t) = f(x(t), u(t)), \qquad t \in I, \qquad x(t_0) = x_0 \tag{AWP}$$

auf dem Intervall $J \subseteq I$ bezüglich der zulässigen Kontrolle $u : I \to \mathcal{U}$, falls $x(\cdot)$ eine zulässige Lösung von (B.28) bzgl. der Restriktion $u|_J$ ist, welche die Bedingungen $t_0 \in J$ und $x(t_0) = x_0$ erfüllt.
Sie heißt maximal, wenn sie die folgende Eigenschaft besitzt:
Ist $\tilde{x} : \tilde{J} \to \mathbb{R}^n$ eine weitere Lösung von (AWP) auf $\tilde{J} \subseteq I$, so folgt $\tilde{J} \subseteq J$ und $x(t) = \tilde{x}(t)$ für alle $t \in \tilde{J}$.

In Satz B.9 haben wir die lokale Existenz einer zulässigen Lösung $x(\cdot)$ von (AWP) nachgewiesen. Für jede zulässige Kontrollfunktion $u : I \to \mathcal{U}$ und allen Anfangsdaten (x_0, t_0) [2] existiert ein $d > 0$ und eine zulässige Lösung $x(\cdot)$ mit $x(t_0) = x_0$, welche auf einem Intervall $[-t_0 - d, t_0 + d] \cap I$ definiert ist. Im nächsten Satz wird gezeigt, dass sich ein solches $d > 0$ unabhängig von der gewählten Kontrolle bestimmen lässt.

Satz B.31 (gleichmäßige lokale Existenz)**.** Für jede kompakte Menge $C \subset \mathbb{R}^n$ existiert ein $\mu > 0$ und ein $d > 0$, so dass die folgende Aussage gilt:
Ist u eine zulässige Kontrolle auf $[0, d]$ und $x_0 \in C$ ein Anfangswert, so kann jede zugehörige Lösung $x(\cdot)$ von

$$\dot{x}(t) = f(x(t), u(t)), \qquad t \in [0, d], \qquad x(0) = x_0 \tag{B.30}$$

auf das Intervall $[0, d]$ (per Konkatenation) erweitert werden.
Außerdem ist $\|\dot{x}(t)\| < \mu$ fast überall und $\|x(t) - x_0\| < \mu t$ für alle $t \in (0, d]$.

[2] OBdA darf $t_0 := 0$ gesetzt werden

Beweisskizze. Da C,$\mathcal{U}$ kompakt und f stetig, darf man

$$\begin{aligned}\mu :=& 1 + \max\{\|f(x,u)\| : \operatorname{dist}(x,C) \le 1, u \in \mathcal{U}\} \ , \\ d :=& \frac{1}{\mu}\end{aligned}$$

setzen. (Hier ist $\operatorname{dist}(x,C) := \inf_{c\in C} \|x-c\|$.)
Es sei $x(\cdot)$ eine beliebige zulässige Lösung des Anfangswertproblems (B.30) zu einer Kontrolle $u : [0,d] \to \mathcal{U}$ und einem Zustand $x_0 \in C$ auf einem Intervall $I \subseteq [0,d]$. Es ist zu zeigen, dass keine Fluchtstelle $\omega \in I$ existiert mit $\omega \le d$ und $x(t) \to \infty$ für $t \overset{<}{\to} \omega$. Angenommen, dies ist falsch und es gibt eine derartiges ω. Dann lässt sich ein erster Zeitpunkt $t > 0$ mit $\|x(t) - x_0\| = 1$ bestimmen, woraus

$$1 \ge \|x(s) - x_0\| \ge \operatorname{dist}(x(s), C) \qquad \forall 0 \le s \le t$$

folgt. Wegen (B.29) gilt weiter

$$1 = \|x(t) - x_0\| \le \int_0^t \underbrace{\|f(x(s),u(s))\|}_{\le \mu - 1 < \mu} \mathrm{ds} < \mu t < \mu d = 1 \ . \tag{B.31}$$

Dies ist ein Widerspruch zur Definition $d := \frac{1}{\mu}$.
Da $\dot{x}(t) = f(x(t),u(t))$ f.ü. und (B.31) mit Ausnahme der ersten Gleichheit für alle $t \in (0,d]$ gilt, sind auch die Abschätzungen gezeigt. □

Satz B.32 (Existenz- und Eindeutigkeitssatz)**.** Es sei $u : I \to \mathcal{U}$ messbar auf einem Intervall I. Die Anfangsdaten $(x_0,t_0) \in \mathbb{R}^n \times I$ seien vorgegeben. Angenommen, es existiert ein $\delta > 0$ und ein $\lambda \in \mathbb{R}$, so dass

$$\|f(x,u) - f(y,u)\| \le \lambda \|x - y\| \qquad \forall x, y \in B_\delta(x_0), u \in \mathcal{U} \ . \tag{B.32}$$

Dann existiert eine (eindeutige) maximale Lösung $\rho : J \to \mathbb{R}^n$ des Anfangswertproblems (AWP) bzgl. $u(\cdot)$, welche auf einem nichtleeren Intervall $J \subseteq I$ definiert ist, das relativ offen zu I ist.

Gilt weiter

$$\frac{\|f(x,u)\|}{\|x\|} \le \mu \qquad \forall \|x\| \ge 1, u \in \mathcal{U} \tag{B.33}$$

für ein $\mu > 0$, so ist $J = I$ das „maximale“ Existenzintervall.

Beweis. Wir sehen problemlos, dass $g(x,t) := f(x,u(t))$ die Bedingungen $(L1)$ und (B.8) erfüllt. Die Behauptung folgt somit aus den Sätzen B.12 und B.15. □

Definition und Satz B.33. Es sei $\mathcal{U} \subset \mathbb{R}^m$ eine kompakte nichtleere Menge und $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ eine stetige Funktion, welche die Eindeutigkeitsbedingung (B.32) erfüllt.
Wir nennen eine zulässige Eingangsfunktion $u : [t_0, T) \to \mathcal{U}$ „zulässig für x_0“, falls die maximale Lösung $\rho : J \to \mathbb{R}^n$ des Anfangswertproblems

$$\dot{x}(t) = f(x(t), u(t)), \qquad t \in [t_0, T], \qquad x(t_0) = x_0 \tag{B.34}$$

auf $J = [t_0, T]$ definiert ist, und setzen

$$D(\phi) := \left\{(T, t_0, x_0, u) \mid t_0 \leq T, x_0 \in \mathbb{R}^n, u \in \mathcal{U}^{[t_0,T)} \text{ zulässig für } x_0\right\} .$$

Auf $D(\phi)$ definieren wir die Abbildung

$$\phi(T; t_0, x_0, u) := \rho(T) ,$$

wobei ρ die (eindeutige) maximale Lösung von (B.34) auf $[t_0, T]$ ist. Dann ist $\Sigma_f := (\mathbb{R}, \mathbb{R}^n, \mathcal{U}, \phi)$ ein autonomes kontinuierliches System. In diesem Sinne darf man die Gleichung (B.28) als autonomes Kontrollsystem bezeichnen.

Beweisskizze. Die Übergangsfunktion ist natürlich wohldefiniert.
Die Prüfung der Systemaxiome ist dem Leser überlassen.
Die Autonomie des Systems wird im nächsten Satz bewiesen.

Satz B.34. Das System Σ_f aus Satz B.33 ist autonom gemäß Definition 1.9.

Beweis. Es sei $(T, t_0, x_0, u) \in D(\phi)$, und $x : [t_0, T] \to \mathbb{R}^n$ sei die maximale Lösung von (B.34) bzgl. u. Wir betrachten die verschobene Trajektorie $y(t) := x(t - s)$ auf dem Intervall $[t_0 + s, T + s]$ für ein $s > 0$. Es folgt $y(t_0 + s) = x_0$ und

$$\dot{y}(t) = \frac{\mathrm{d}}{\mathrm{d}t} x(t - s) = f\left(x(t - s), u(t - s)\right) = f(y(t), u(t - s))$$

für fast alle $t \in [t_0 + s, T + s]$. Die absolut stetige Funktion $y(\cdot)$ ist also die maximale Lösung des Anfangswertproblems

$$\dot{y}(t) = f(y(t), u^s(t)), \qquad t \in [t_0 + s, T + s], \qquad y(t_0 + s) = x_0$$

bzgl. der verschobenen Kontrolle $u^s(t) := u(t - s)$. Wir erhalten, dass u^s zulässig für x_0 ist und

$$\phi(T; t_0, x, u) = x(T) = y(T + s) = \phi(T + s; t_0 + s, x, u^s) .$$

□

B.4 Erreichbare Mengen von autonomen Kontrollsystemen

In diesem Abschnitt werden die erreichbaren Mengen des autonomen kontinuierlichen Kontrollsystems Σ_f aus Defintion B.33 untersucht. Σ_f ist gegeben durch die Gleichung

$$\dot{x}(t) = f(x(t), u(t)), \qquad u(t) \in \mathcal{U}, \qquad x(t) \in \mathbb{R}^n \tag{B.35}$$

mit stetiger rechter Seite $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ und kompaktem Steuerbereich $\mathcal{U}$. Wir übernehmen die Bezeichnungen aus Abschnitt 1.3. Die erreichbare Menge $\mathcal{A}_T(p)$ von $p \in \mathbb{R}^n$ zur Zeit $T \geq 0$ ist dann die Menge aller Zustände $q \in \mathbb{R}^n$, für welche eine Lösung $x(\cdot\,)$ vom Anfangswertproblem

$$\dot{x} = f(x, u), \qquad t \in [0, T], \qquad x(0) = p$$

bzgl. einer messbaren Kontrolle $u : [0, T] \to \mathcal{U}$ existiert, so dass $x(T) = q$.

Definition und Satz B.35. Ein Punkt $p \in \mathbb{R}^n$ heißt *kritisch*, falls

$$0 \in f(p, \mathcal{U}) := \bigcup_{u \in \mathcal{U}} f(p, u) \ .$$

Die erreichbaren Mengen kritischer Punkte sind *monoton*, d.h.

$$\mathcal{A}_s(p) \subseteq \mathcal{A}_t(p) \qquad \forall 0 \leq s \leq t \ .$$

Beweis. Da p kritisch, existiert ein $u_0 \in \mathcal{U}$ mit $f(p, u_0) = 0$. Die konstante Funktion $x(t) \equiv p$ ist die Lösung zur Kontrolle $u(t) \equiv u_0$ und daher $p \in \mathcal{A}_t(p)$ für alle $t \geq 0$. Sei nun $q \in \mathcal{A}_s(p)$. Wegen $p \in \mathcal{A}_{t-s}(p)$, gilt gemäß (1.2) $q \in \mathcal{A}_t(p)$ für alle $t \geq s$. □

Folgerung B.36 (Lokale Beschränktheit)**.** Es existiert ein $d > 0$, so dass

$$\mathcal{A}_t(x_0) \subset B_1(x_0) \qquad \forall t \in [0, d] \ .$$

Beweis. Setze $C := \{x_0\}$ im Satz zur gleichmäßigen lokalen Existenz B.31. Dann gilt für $\mu := 1 + \max\{\|f(x, u)\| : \|x - x_0\| \leq 1, u \in \mathcal{U}\}$ und $d := \frac{1}{\mu}$ die Abschätzung

$$\|x(t) - x_0\| < \mu t \leq \mu d = 1 \qquad \forall t \in [0, d] \ ,$$

wobei $x(\cdot\,)$ eine beliebige zulässige Lösung von (B.35). □

Beispiel B.37. Betrachte das initilisierte skalare Kontrollsytem

$$\dot{x} = (1 + x^2)u, \qquad |u(t)| \leq 1, \qquad x(0) = 0 \ . \tag{B.36}$$

Hier ist $\mu := 1+\max\{|(1 + x^2)u| : |x| \leq 1, |u| \leq 1\} = 3$. Nach Folgerung B.36 liegt für $t \in [0, \frac{1}{3}]$ die erreichbare Menge $\mathcal{A}_t(0)$ in der offenen Einheitskugel. Wir sehen auch, dass 0 kritisch ist. Die erreichbaren Mengen sind somit monoton. Um genauere Aussagen zu gewinnen, lösen wir (B.36). Die Theorie zu gewöhnlichen Differentialgleichungen mit getrennten Variablen liefert die Lösung

$$x(t) = \tan\left(\int_0^t u(s)ds\right) \ , \tag{B.37}$$

bezüglich einer messbaren Funktion $u : [0, d] \to [-1, 1]$ mit $d < \frac{\pi}{2}$. Man beachte, dass $x(\cdot)$ absolut stetig[3] über $[0, d]$ ist und die Gleichung (B.36) fast überall erfüllt. (Denn $(\tan t)' = \frac{1}{\cos^2 t} = 1 + \tan^2 t$.) Die Funktion $x(\cdot)$ ist folglich eine zulässige Lösung auf $[0, d]$.
Da für alle $x_0 \in \mathbb{R}$ die Abschätzung

$$\begin{aligned} |f(x, u) - f(y, u)| = |(x^2 - y^2)u| &= |(x + y)u| \cdot |x - y| \\ &\leq 2(\delta + |x_0|) \cdot |x - y| \qquad \forall x, y \in B_\delta(x_0), u \in \mathcal{U} \end{aligned}$$

gilt, ist die Eindeutigkeitsbedingung (B.32) erfüllt. Also sind Lösungen von (B.36) auf dem Intervall $[0, d]$, $d < \frac{\pi}{2}$, genau von der Form (B.37).
Aus $|x(t)| \leq \tan t$, $\{\pm \tan t\} \in \mathcal{A}_t(0)$ und der Monotonie erreichbarer Mengen folgt somit

$$\mathcal{A}_t(0) = [-\tan t, \tan t] \qquad \forall t \in [0, \tfrac{\pi}{2}) \ .$$

Für $t = \frac{\pi}{2}$ ist die erreichbare Menge unbeschränkt.

Folgerung B.38 (Beschränktheit)**.** Angenommen es gibt ein $\mu \in \mathbb{R}_+$, so dass

$$\frac{x^* f(x, u)}{\|x\|^2} \leq \mu \qquad \forall \|x\| \geq 1, u \in \mathcal{U} \ . \tag{B.38}$$

Dann besitzt laut Satz B.32 die Gleichung (B.35) globale Existenz in die Zukunft und es gilt:

$$\mathcal{A}_t(x_0) \text{ ist beschränkt und nichtleer für alle } t \geq 0.$$

[3]Da tan Lipschitzstetig über $[-d, d]$ ist, dürfen wir Satz A.17 anwenden.

Beweis. Weil der Steuerbereich $\mathcal{U}$ nichtleer ist und jede konstante Kontrolle $u(t) \equiv u_0$ mit $u_0 \in \mathcal{U}$ zulässig ist, folgt aus Satz B.32, dass es eine zulässige Lösung von (B.35) mit $x(0) = x_0$ (zur Kontrolle $u \equiv u_0$) gibt, die auf dem Intervall $[0, \infty)$ definiert ist. Es folgt $\mathcal{A}_t(x_0) \neq \emptyset$ für alle $t \geq 0$.
Um die Beschränktheit zu zeigen, übertragen wir die Überlegungen im Beweis von Satz B.12 auf unser Kontrollsystem. Sei dazu $x(\cdot)$ eine zulässige Lösung auf $[0, \infty)$ bzgl. einer Kontrolle u und $t \geq 0$ irgendein fester Zeitpunkt. Ohne Einschränkung sei $r(t) := \|x(t)\| > 1$. Wir setzen

$$s := \begin{cases} 0, & \text{falls } \|x(\tau)\| > 1 \ \forall \tau \in [0,t] \\ \max\{\tau \mid 0 \leq \tau \leq t,\ x(\tau) = 1\}, & \text{sonst} \end{cases}$$

Dann gilt für fast alle $\tau \in [s,t]$

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \log r(\tau) = \frac{\dot{r}(\tau)}{r(\tau)} = \frac{2x(\tau)^* f(x(\tau), u(\tau))}{\|x(\tau)\|^2} \overset{\text{(B.38)}}{\leq} 2\mu \ .$$

Integration von $\frac{\mathrm{d}}{\mathrm{d}\tau} \log r(\tau)$ über $[s,t]$ liefert wie im Beweis von Satz B.12 die Abschätzung

$$r(t) \leq r(s) \exp(2\mu(t-s)) \ .$$

Es folgt aus der Definitionen von s die konservative Abschätzung

$$\|x(t)\|^2 \leq \max\{1, \|x_0\|^2\} \cdot \exp(2\mu t) \ ,$$

und daher

$$\mathcal{A}_t(x_0) \subseteq \overline{B\big(0, \max\{1, \|x_0\|\} \cdot \exp(\mu t)\big)} \qquad \forall t \geq 0 \ .$$

□

In der Optimierung ist das Auffinden zeitoptimaler Steuerungen eine wichtige Aufgabe [11]. Im einfachsten Fall besteht das Problem darin, einen Zielpunkt $z \in \mathcal{A}_T(x_0)$, $T \geq 0$, in minimaler Zeit

$$t^{\mathrm{f}} := \inf \{t \geq 0 \mid z \in \mathcal{A}_t(x_0)\} \tag{B.39}$$

von einem Anfangspunkt x_0 zu erreichen. Die Existenz einer zeitoptimalen Kontrolle ist häufig abhängig von der Abgeschlossenheit der Menge $\mathcal{A}_{t^{\mathrm{f}}}(x_0)$. Doch erreichbare Mengen sind nicht immer abgeschlossen, wie das folgende Beispiel zeigt.

Beispiel B.39. Betrachte das Kontrollsystem im $\mathbb{R}^2$

$$\dot{x}_1 = (1 - x_2^2) u^2 \ , \tag{B.40}$$

$$\dot{x}_2 = u \tag{B.41}$$

mit Steuerbereich $\mathcal{U} = [-1, 1] \ni u(t)$ und Anfangsbedingung $x(0) = \binom{0}{0}$.
Man sieht leicht, dass wegen

$$x_1(T) = \int_0^T \left(1 - x_2(s)^2\right) u(s)^2 \mathrm{ds} \leq \int_0^T 1 \cdot 1 \ \mathrm{ds} = T \tag{B.42}$$

die erreichbare Menge $\mathcal{A}_T(0)$ in der Halbebene

$$\left\{(x, y) \in \mathbb{R}^2 : x \leq T\right\}$$

liegt (für eine beliebige Endzeit $T \geq 0$). Wir zeigen nun, dass für $T > 0$ der Punkt $\binom{T}{0}$ zwar der Grenzwert einer Folge aus $\mathcal{A}_T(\binom{0}{0})$ ist, jedoch nicht zu dieser Menge gehört:

Wir zerlegen für jedes $k = 1, 2, \ldots$ das Intervall $[0, T]$ in $2k$ Teilintervalle der Form $[\frac{(j-1)T}{2k}, \frac{jT}{2k}]$, $j = 1, 2, \ldots, 2k$, und bezeichnen das j-te Teilintervall mit I_j. Nun definieren wir für jedes $k = 1, 2, \ldots$ eine zulässige Kontrollfunktion, welche auf diesen Intervallen alterniert, durch

$$u_k(t) := \begin{cases} +1, & \text{falls } t \in I_j \text{ und } j \text{ ungerade} \\ -1, & \text{falls } t \in \operatorname{int}(I_j) \text{ und } j \text{ gerade} \end{cases} .$$

$x(t; u_k)$ bezeichne die zugehörige Lösung auf $[0, T]$ mit Anfangswert $x(0; u_k) = \binom{0}{0}$. Nach Konstruktion gilt

$$\begin{aligned} x_2(T; u_k) &= \int_0^T u_k(s) \mathrm{ds} = \tfrac{1}{2}T - \tfrac{1}{2}T = 0 \ , \\ |x_2(t; u_k)| &\leq \frac{T}{2k} \end{aligned}$$

für alle $k \in \mathbb{N}$ und alle $t \in [0, T]$. Es folgt fast überall

$$1 - \frac{T^2}{4k^2} \leq \underbrace{1 - x_2(t; u_k)^2}_{\dot{x}_1(t; u_k)} \leq 1 \ ,$$

und Integration über $[0, T]$ liefert

$$\left(1 - \frac{T^2}{4k^2}\right) T \leq x_1(T; u_k) \leq T \ . \tag{B.43}$$

Also konvergiert $\binom{x_1(T;u_k)}{x_2(T;u_k)}$ gegen $\binom{T}{0}$ für $k \to \infty$.
Allerdings ist $(T, 0) \notin \mathcal{A}_T(\binom{0}{0})$. Denn sonst würde aus der Gleichheit in (B.42)

$$\left(1 - x_2(t)^2\right) u(t)^2 = 1 \qquad \text{f.ü.}$$

und somit

$$x_2(t) = 0,\ u(t) = \pm 1 \qquad \text{f.ü.}$$

folgen. Dies ist ein Widerspruch, denn $x_2(t) = 0$ f.ü. impliziert $u(t) = 0$ f.ü. wegen (B.41).
Die erreichbare Menge $\mathcal{A}_T(\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right))$ ist zu keinem Zeitpunkt $T > 0$ abgeschlossen!

Insbesondere wird in [11, p.107] bemerkt, dass es keine zeitoptimale Kontrolle gibt, die den Punkt $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ in minimaler Zeit t^{f} von $\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$ erreicht (Idee: $t^{\mathrm{f}} = 1$ und $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right) \notin \mathcal{A}_1(\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right))$).

Wir bereiten nun den Satz von Filippov vor, der ein Kriterium für die Abgeschlossenheit von $\mathcal{A}_t(x_0)$ liefert. Um dieses Kriterium anwenden zu können, ist die Konvexität der Menge $F(x)$ aus der nächsten Definition für alle $x \in \mathbb{R}^n$ erforderlich.

Definition und Lemma B.40. Es wird die mengenwertige Funktion

$$F(x) := f(x, \mathcal{U}) = \bigcup_{u \in \mathcal{U}} f(x, u) \qquad \forall x \in \mathbb{R}^n$$

eingeführt. Da in unserem Fall f stetig und $\mathcal{U}$ kompakt nichtleer ist, ist auch $F(x)$ kompakt und nichtleer für alle $x \in \mathbb{R}^n$, d.h. $F(x) \in \Omega(\mathbb{R}^n)$. Insbesondere ist F als Abbildung der Form $F : \mathbb{R}^n \mapsto \Omega(\mathbb{R}^n)$ stetig (bzgl. des Hausdorffabstands).

Beweis. Aus der Stetigkeit von $x \mapsto f(x, u)$ folgt, dass

$$\forall x \in \mathbb{R}^n\ \forall u \in \mathcal{U}\ \forall \epsilon > 0\ \exists \delta(x, \epsilon, u) > 0 : \\ \|f(x, u) - f(y, u)\| < \epsilon\ \forall \|x - y\| < \delta(x, \epsilon, u)\ .$$

Also liegt jedes Element $f(x, u) \in F(x)$ in einer $\epsilon-$Umgebung von $F(y)$, falls $\|x - y\| < \delta(x, \epsilon, u)$. Und analog liegt jedes Element von $F(y)$ in einer $\epsilon-$Umgebung von $F(x)$, falls $\|x - y\| < \delta(y, \epsilon, u)$. Aus der Minimalitätseigenschaft von $d_H(F(x), F(y))$ gemäß Beispiel A.6 erhalten wir schließlich

$$d_H(F(x), F(y)) \leq \epsilon \qquad \forall \|x - y\| < \min\{\delta(x, \epsilon, u), \delta(y, \epsilon, u)\}\ ,$$

woraus die Stetigkeit von $F(x)$ folgt. □

Beispiel B.41. In Beispiel B.39 ist

$$F(x) = \left\{ \begin{pmatrix} (1 - x_2^2)\, u^2 \\ u \end{pmatrix} : -1 \leq u \leq 1 \right\} .$$

Für $x_2 \neq \pm 1$ ist die Projektion von $F(x)$ auf die erste Koordinate der Graph einer Parabel der Form Cu^2 mit $C \neq 0$. Die Nichtkonvexität dieser Parabel impliziert die Nichtkonvexität von $F(x)$ in diesem Fall.

Lemma B.42. Eine Funktion $x(\cdot)$ ist genau dann eine Lösung von (B.35) bezüglich einer zulässigen Kontrolle u, wenn $x(\cdot)$ eine absolut stetige Funktion ist, für die

$$\dot{x}(t) \in F(x(t)) \text{ f.ü.} \tag{B.44}$$

gilt. Man sagt, $x(\cdot)$ löst die „Differentialinklusion“ (B.44) zum Kontrollproblem (B.35).

Beweis. 1. Ist $x(\cdot)$ eine Lösung von (B.35) zur Kontrolle u, so ist $x(\cdot)$ absolut stetig und $\dot{x}(t) = f(x(t), u(t)) \in F(x(t))$ f.ü . Dies zeigt, dass $x(\cdot)$ eine Lösung von (B.44) ist.

2. Siehe [11, p.106].

Hilfssatz B.43 (Mittelwertsatz)**.** Es sei C eine abgeschlossene konvexe Teilmenge des $\mathbb{R}^n$ und $g : [0, \infty) \to C$ eine lokal integrierbare Funktion. Dann gilt für den Mittelwert

$$\frac{1}{t} \int_0^t g(s)\mathrm{ds} \in C \qquad \forall t \geq 0 .$$

Satz B.44 (Filippov)**.** Angenommen, die rechte Seite des Kontrollsystems (B.35) erfüllt die Existenzbedingung (B.38), und $F(x)$ ist konvex für alle $x \in \mathbb{R}^n$. Dann ist $\mathcal{A}_T(x_0)$ kompakt für alle $T \geq 0$ und stetig in T bezüglich der Hausdorffmetrik.

Beweis. 1. In Folgerung B.38 wurde bereits bewiesen, dass $\mathcal{A}_T(x_0)$ beschränkt ist. Es bleibt, die Abgeschlossenheit der erreichbaren Menge nachzuweisen. Sei dazu $(z_k)_{k\in\mathbb{N}}$ eine beliebige konvergente Folge in $\mathcal{A}_T(x_0)$ mit Grenzwert z^*. Weiter sei $(x_k)_{k\in\mathbb{N}}$ eine Folge zulässiger Lösungen von (B.35) auf $[0, T]$ mit Anfangspunkten $x_k(0) = x_0$, Endpunkten $x_k(T) = z_k$ und zugehörigen Kontrollen u_k $(k = 1, 2, \ldots)$. Es genügt zu zeigen, dass es eine weitere zulässige Lösung x gibt, die $x(T) = z^*$ erfüllt.

2. Aus dem Beweis von Folgerung B.38 folgt, dass alle Lösungstrajektorien x_k in einer kompakten Menge

$$D := \left\{x \in \mathbb{R}^n : \|x\|^2 \leq (1 + \|x_0\|^2) \exp 2\mu T\right\} \quad \text{für ein } \mu > 0$$

verlaufen. Auf der kompakten Menge $D \times \mathcal{U}$ ist die rechte Seite f stetig. Daher gibt es ein $M > 0$, so dass

$$\|f(x,u)\| \leq M \qquad \forall x \in D, u \in \mathcal{U} \ . \tag{B.45}$$

Es folgt für alle $s \leq t$ in $[0,T]$ und $k \in \mathbb{N}$

$$\|x_k(t) - x_k(s)\| = \left\| \int_s^t f(x_k(\tau), u_k(\tau)) \mathrm{d}\tau \right\| \leq M|t-s| \ , \tag{B.46}$$

d.h. alle x_k sind Lipschitzstetig mit Konstante M. Also bildet die Folge $(x_k)_{k\in\mathbb{N}}$ eine gleichgradig stetige und gleichmäßig beschränkte Familie. Nach Arzelà-Ascoli A.10 existiert eine Teilfolge, ohne Einschränkung die ursprüngliche Folge, die gleichmäßig auf $[0,T]$ gegen eine Funktion x konvergiert. Diese ist ebenfalls Lipschitzstetig mit Konstante M ($k \to \infty$ in (B.46)) und ist daher absolut stetig. Offensichtlich gilt weiter

$$x(0) = \lim_{k\to\infty} x_k(0) = x_0, \qquad x(T) = \lim_{k\to\infty} z_k = z^* \ .$$

3. Wir werden gleich sehen, dass $\dot{x}(t) \in F(x(t))$ fast überall.
Sei $t_0 \in [0,T]$ ein beliebiger Zeitpunkt, an welchem $\dot{x}(t_0)$ existiert.
Es gilt

$$\begin{aligned}
\frac{x(t) - x(t_0)}{t - t_0} &= \lim_{k\to\infty} \frac{x_k(t) - x_k(t_0)}{t - t_0} = \lim_{k\to\infty} \frac{1}{t-t_0} \int_{t_0}^t \dot{x_k}(\tau) \mathrm{d}\tau \\
&\overset{\text{Transl.}}{=} \lim_{k\to\infty} \frac{1}{t-t_0} \int_0^{t-t_0} \dot{x_k}(t_0 + \tau) \mathrm{d}\tau \\
&\overset{\text{Subst.}}{=} \lim_{k\to\infty} \int_0^1 \dot{x_k}(t_0 + (t-t_0)\tau) \mathrm{d}\tau \ .
\end{aligned} \tag{B.47}$$

Weil x differenzierbar ist in t_0 und F stetig in $x(t_0)$ ist, können wir für ein vorgegebenes $\epsilon > 0$ ein $\delta > 0$ bestimmen, so dass

$$\left\| \frac{x(t) - x(t_0)}{t - t_0} - \dot{x}(t_0) \right\| < \epsilon \qquad \forall |t - t_0| < \delta \tag{B.48}$$

und

$$d_H(F(x(t_0)), F(y)) < \epsilon \qquad \forall \|x(t_0) - y\| < \delta \ . \tag{B.49}$$

x ist stetig in t_0, d.h. es existiert ein $t > t_0$ mit $|t - t_0| < \delta$, so dass

$$\|x(t_0) - x(\tau)\| < \tfrac{\delta}{2} \qquad \forall \tau \in [t_0, t] \ . \tag{B.50}$$

$(x_k)_{k\in\mathbb{N}}$ konvergiert gleichmäßig gegen x auf $[t_0, t]$, d.h.

$$\exists N \in \mathbb{N} : \|x_k(\tau) - x(\tau)\| < \tfrac{\delta}{2} \qquad \forall \tau \in [t_0, t] \ \forall k \geq N \ . \tag{B.51}$$

Aus der Dreiecksungleichung, (B.50) und (B.51) folgt nun

$$\|x(t_0) - x_k(\tau)\| \leq \|x(t_0) - x_k(\tau)\| + \|x_k(\tau) - x(\tau)\| < \delta \qquad \forall \tau \in [t_0, t] \ \forall k \geq N \ . \tag{B.52}$$

(B.49) und (B.52) ergeben

$$d_H\big(F(x(t_0)), F(x_k(\tau))\big) < \epsilon \qquad \forall \tau \in [t_0, t] \ \forall k \geq N \ . \tag{B.53}$$

Da $\dot{x_k}(\tau) \in F(x_k(\tau))$ fast überall, lässt sich (B.53) wie folgt auswerten:
Ist $k \geq N$, so gehört $\dot{x_k}(\tau)$ zu einer abgeschlossenen ϵ-Umgebung von $F(x(t_0))$ für fast alle $\tau \in [t_0, t]$, d.h.

$$\dot{x_k}(\tau) \in \overline{B_\epsilon(F(t_0))} \qquad \forall' \tau \in [t_0, t], \forall k \geq N \ . \tag{B.54}$$

Dies kann man auch anders schreiben:

$$\dot{x_k}(t_0 + (t - t_0)\tau) \in \overline{B_\epsilon(F(t_0))} \qquad \forall' \tau \in [0, 1], \forall k \geq N \tag{B.55}$$

Mit $F(t_0)$ ist auch $\overline{B_\epsilon(F(t_0))}$ konvex und der Mittelwertsatz B.43 besagt, dass

$$\int_0^1 \dot{x_k}(t_0 + (t - t_0)\tau) \mathrm{d}\tau \in \overline{B_\epsilon(F(t_0))} \qquad \forall k \geq N \ .$$

Schließlich nutzen wir die Abgeschlossenheit von $\overline{B_\epsilon(F(t_0))}$, um

$$\frac{x(t) - x(t_0)}{t - t_0} \overset{\text{(B.47)}}{=} \lim_{k\to\infty} \int_0^1 \dot{x_k}(t_0 + (t - t_0)\tau) \mathrm{d}\tau \in \overline{B_\epsilon(F(t_0))}$$

zu zeigen, woraus mit Hilfe von (B.48)

$$\dot{x}(t_0) \in \overline{B_{2\epsilon}(F(t_0))}$$

folgt. Da ϵ beliebig klein und $F(t_0)$ abgeschlossen, erhalten wir $\dot{x}(t_0) \in F(t_0)$ (für $\epsilon \to 0$). Dies gilt für alle t_0, an denen $\dot{x}(t_0)$ existiert, also für fast alle $t_0 \in [0, T]$.

4. Nach Lemma B.42 ist mit x eine zulässige Lösung von (B.35) auf $[0,T]$ gefunden, die $x(0)=x_0$ und $x(T)=z^*$ erfüllt. Die Kompaktheit von $\mathcal{A}_T(x_0)$ haben wir somit bewiesen.

5. Es bleibt, die Stetigkeit von $t \mapsto \mathcal{A}_t(x_0)$ über dem Intervall $[0,T]$ nachzuweisen.
Es seien $t_1, t_2 \in [0,T]$ und $z_1 \in \mathcal{A}_{t_1}(x_0)$ beliebige Elemente. Dann existiert eine zulässige Lösung $x(\cdot\,)$ (zu einer Kontrolle u) mit $x(0)=x_0$ und $x(t_1)=z_1$. Wir setzen $z_2 := x(t_2)$. Wegen (B.45) gibt es ein $M>0$, so dass

$$\|z_1 - z_2\| = \left\| \int_{t_1}^{t_2} f(x(s),u(s))\mathrm{ds} \right\| \leq M \cdot |t_1 - t_2| \ .$$

Diese Ungleichung erhalten wir aus Symmetriegründen auch, falls wir die Rollen von z_1 und z_2 vertauschen. Es folgt

$$\forall \epsilon > 0 : \mathcal{A}_{t_1}(x_0) \subseteq \overline{B_\epsilon(\mathcal{A}_{t_2}(x_0))} \text{ und } \mathcal{A}_{t_2}(x_0) \subseteq \overline{B_\epsilon(\mathcal{A}_{t_1}(x_0))} \ \forall |t_1 - t_2| < \tfrac{\epsilon}{M} \ .$$

Dies zeigt die Stetigkeit von $t \mapsto \mathcal{A}_t(x_0)$ bezüglich der Hausdorff-Metrik. □

Folgerung B.45 (Existenz zeitoptimaler Kontrollen)**.** Angenommen die Bedingungen aus Satz B.44 sind erfüllt und es gibt ein $T \geq 0$, so dass $z \in \mathcal{A}_T(x_0)$. Dann gibt es eine (zeitoptimale) Steuerung, die den Punkt z von x_0 in minimaler Zeit t^{f} erreicht. Also $t^{\mathrm{f}} \in \mathcal{A}_{t^{\mathrm{f}}}(x_0)$.

Beweis. Nach Definition (B.39) ist $t_f = \inf\{t \geq 0 \mid z \in \mathcal{A}_t(x_0)\}$. Wegen $z \in \mathcal{A}_T(x_0)$ ist $t_f \leq T$. Es gibt daher Zeitpunkte $(t_k)_{k\in\mathbb{N}}$ mit $t_k \overset{>}{\to} t^{\mathrm{f}}$ und zulässige Lösungen $(x_k)_{k\in\mathbb{N}}$, so dass $x_k(0)=x_0$ und $x_k(t_k)=z$. Weiter folgt wie in (B.46)

$$\|x_k(t^{\mathrm{f}}) - z\| = \|x_k(t^{\mathrm{f}}) - x_k(t_k)\| \leq \int_{t^{\mathrm{f}}}^{t_k} M\mathrm{ds}$$

für ein $M>0$. Es konvergiert somit $x_k(t^{\mathrm{f}})$ gegen z für $k \to \infty$. Weil $x_k(t^{\mathrm{f}}) \in \mathcal{A}_{t^{\mathrm{f}}}(x_0)$ und $\mathcal{A}_{t^{\mathrm{f}}}(x_0)$ abgeschlossen ist, schliessen wir $z \in \mathcal{A}_{t_f}(x_0)$. □

Literaturverzeichnis

[1] Hans Wilhelm Alt. *Lineare Funktionalanalysis : eine anwendungsorientierte Einführung*. Springer, Berlin [u.a.], 4., überarb. und erw. aufl. edition, 2002.

[2] R. W. Brockett. System theory on group manifolds and coset spaces. *SIAM J. Control*, 10:265–284, 1972.

[3] R. W. Brockett. Lie theory and control systems defined on spheres. *SIAM J. Appl. Math.*, 25:213–225, 1973. Lie algebras: applications and computational methods (Conf., Drexel Univ., Philadelphia, Pa., 1972).

[4] Roger W. Brockett. Volterra series and geometric control theory. *Automatica—J. IFAC*, 12(2):167–176, 1976.

[5] Carlo Bruni, Gianni DiPillo, and Giorgio Koch. Bilinear systems: an appealing class of "nearly linear" systems in theory and applications. *IEEE Trans. Automatic Control*, AC-19:334–348, 1974.

[6] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Springer, 1988.

[7] Otto Forster. *Analysis. 3*, volume 52 of *Vieweg Studium: Aufbaukurs Mathematik [Vieweg Studies: Mathematics Course]*. Friedr. Vieweg & Sohn, Braunschweig, 3.,verbesserte auflage edition, 1981. Integralrechnung im $\mathbf{R}^n$ mit Anwendungen. [Integral calculus in $\mathbf{R}^n$ with applications].

[8] Otto Forster. *Analysis*. Vieweg-Verlag, Braunschweig/Wiesbaden, 6.,verbesserte auflage edition, 2001.

[9] O. Hájek. *Control theory in the plane*, volume 153 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 1991.

[10] O. Hajek. Bilinear control systems. Special types. *Kibernet. Sistem. Anal.*, (2):173–188, 191, 2002.

[11] Henry Hermes and Joseph P. LaSalle. *Functional analysis and time optimal control*. Academic Press, New York, 1969. Mathematics in Science and Engineering, Vol. 56.

[12] Alberto Isidori. *Nonlinear control systems*. Communications and control engineering series. Springer, Berlin [u.a.], 3. ed., 3. print. edition, 2001. Literaturverz. S. 535 - 544.

[13] Velimir Jurdjevic and Héctor J. Sussmann. Control systems on Lie groups. *J. Differential Equations*, 12:313–329, 1972.

[14] E. N. Khaĭlov. The attainability set of a homogeneous bilinear system with quasi-commuting matrices. *Differ. Uravn.*, 38(12):1620–1626, 1725, 2002.

[15] E. N. Khaĭlov. On the parametrization of the attainable set of a homogeneous bilinear system with quasicommuting matrices. *Vestnik Moskov. Univ. Ser. XV Vychisl. Mat. Kibernet.*, (1):37–42, 58, 2004.

[16] C. Lesiak and A. Krener. The existence and uniqueness of volterra series for nonlinear systems. *IEEE Trans. Automatic Control*, AC-23(6):1090–1095, Dec 1978.

[17] Jürg T. Marti. *Konvexe Analysis*. Birkhäuser Verlag, Basel, 1977. Lehrbücher und Monographien aus dem Gebiet der Exakten Wissenschaften, Mathematische Reihe, Band 54.

[18] R. R. Mohler. *Bilinear Systems and Control*. Encyclopedia of physical science and technology. Acad. Press, San Diego [u. a.], 3. ed. edition, 2002. 660–674.

[19] Mohler R. R. and Shen C. N. *Optimal Control of Nuclear Reactors*. Acad. Press, New York, 1972.

[20] Wulf Rossmann. *Lie groups : an introduction through linear groups*. Oxford graduate texts in mathematics ; 5. Oxford Univ. Press, Oxford [u.a.], 1. publ. in paperback edition, 2006.

[21] Antonio Ruberti, Alberto Isidori, and Paolo d'Alessandro. *Theory of bilinear dynamical systems*. Springer-Verlag, Vienna, 1972. Course held at the Department for Automation and Information, July 1972, International Centre for Mechanical Sciences, Udine. Courses and Lectures, No. 158.

[22] Wilson J. Rugh. *Nonlinear system theory*. Johns Hopkins Series in Information Sciences and Systems. Johns Hopkins University Press, Baltimore, Md., 1981. The Volterra/Wiener approach.

[23] Eduardo D. Sontag. *Mathematical control theory*, volume 6 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 1998. Deterministic finite-dimensional systems.

[24] Héctor J. Sussmann. The "bang-bang" problem for certain control systems in GL(n, R). *SIAM J. Control*, 10:470–476, 1972.

[25] Harry L. Trentelman, Anton Stoorvogel, and Malo L. J. Hautus. *Control theory for linear systems*. Communications and control engineering series. Springer, London [u.a.], 2001. Literaturverz. S. 373 - 384.

[26] Lubin Vulkov, Jerzy Waśniewski, and Jerzy Wsniewski, editors. *Numerical analysis and its applications*. Lecture Notes in Computer Science, Vol.3401. Springer-Verlag, Berlin, 2005. Third International Conference, NAA 2004, Rousse, Bulgaria, June 29 - July 3, 2004, Revised Selected Papers.

[27] Dirk Werner. *Funktionalanalysis*. Springer-Verlag, Berlin, extended edition, 2000.

Stichwortverzeichnis

Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel verfasst zu haben.

Diese Arbeit wurde nicht bereits in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Bayreuth, 28. Juni 2007 .....

Matthias Zerndl