

Skript zur Vorlesung
“Systemtheorie und Regelungstechnik 1”
an der Universität Freiburg

Moritz Diehl

19. August 2014

Inhaltsverzeichnis

Vorwort	7
1 Einführung	9
1.1 Systemtheorie und Regelungstechnik	9
1.2 Aufbau der Vorlesung	11
1.3 Mathematischer Hintergrund und Literatur	12
1.4 Mathematische Notation	14
2 Modellierung dynamischer Systeme	17
2.1 Gewöhnliche Differentialgleichungen	18
2.2 Modell eines gesteuerten Traktors	19
2.3 Physikalische Einheiten und Skalierung	23
2.4 Mechanische Systeme	25
2.4.1 Freie Bewegung eines Körpers im Raum	25
2.4.2 Der harmonische Oszillator	26
2.4.3 Satellitenmodell im dreidimensionalen Raum	27
2.4.4 Rotationsbewegungen und Drehpendel	29
2.5 Lineare und Nichtlineare Systeme	30
2.6 Elektrische Schaltkreise	32
2.6.1 RC-Kreis	33
2.6.2 Elektrischer Oszillator	34
2.7 Zeitinvarianz und Linearität	35
2.7.1 Zeitinvarianz	35
2.7.2 Linearität	37
2.8 Zeitdiskrete Systeme und Simulation	38
2.9 Analytische Lösung für LTI Systeme	40
3 Eingangs-Ausgangsdarstellung	43
3.0.1 MIMO und SISO Systeme	44
3.1 Eingangs-Ausgangsdifferentialgleichung	44
3.1.1 Eingangs-Ausgangsmodell für einen Oszillator	45
3.1.2 Allgemeine Eingangs-Ausgangsform	45
3.2 Blockschaltbilder und Übertragungsglieder	46
3.2.1 Proportional, Integrier, und Differenzierglied	47

3.2.2	Physikalische Realisierung des D-Gliedes als DT_1 -Glied	48
3.2.3	Verzögerungsglied erster Ordnung oder PT_1 -Glied	49
3.2.4	Verzögerungsglied zweiter Ordnung oder PT_2 -Glied	50
3.3	Äquivalenz von Zustands- und E/A-Darstellung	51
3.3.1	Regelungsnormalform	51
3.3.2	Regelungsnormalform für Systeme vom relativen Grad n	52
3.3.3	Zustandstransformationen	53
4	Das charakteristische Polynom	55
4.1	Invarianz unter Zustandstransformationen	55
4.2	Koeffizienten aus E/A-Differentialgleichung	56
4.3	Nullstellen des charakteristischen Polynoms, oder "Polstellen"	57
4.4	Stabilität bei negativen Realteilen der Polstellen	58
5	Dynamisches Verhalten linearer Systeme	63
5.1	Die Sprungantwort	63
5.2	Der Dirac-Impuls und die Impulsantwort	64
5.2.1	Totzeitglieder	65
5.2.2	Approximation des Totzeitglieds durch ein PT_n -Glied	67
5.3	Eingangs-Ausgangs-Stabilität (BIBO-Stability)	69
5.4	Charakteristische Größen von BIBO-Systemen	70
6	Die Übertragungsfunktion	73
6.1	Antwort auf exponentielle Inputs	74
6.2	Laplace-Transformation	75
6.3	Eigenschaften der Laplace-Transformation	76
6.3.1	Überlagerung	76
6.3.2	Zeitableitung	76
6.3.3	Höhere Ableitungen	77
6.3.4	Faltung	77
6.4	Übertragungsfunktion aus der Eingangs-Ausgangs-Differentialgleichung	79
6.5	Übertragungsfunktion aus Zustandsdifferentialgleichung	79
6.6	Gekoppelte und geregelte Systeme	81
6.6.1	Reihenschaltung	81
6.6.2	Parallelschaltung	81
6.6.3	Rückkopplungsschaltung	82
6.6.4	Beispiel: der einfach geregelte Traktor	83
6.6.5	Der PID-Regler	84
6.7	Zeitkonstantenform	84
7	Frequenzgang und Bode-Diagramm	87
7.1	Bode-Diagramm	88
7.1.1	Bode-Diagramm des PT_1 -Glieds	90
7.2	Nyquist-Diagramm, oder Ortskurve	91
7.3	Zusammenhang mit Übertragungsfunktion in der komplexen Ebene	91
7.4	Bode-Diagramm und Verhalten des PT_2 -Gliedes	93

<i>INHALTSVERZEICHNIS</i>	5
7.4.1 Polstellen und Zeitverhalten des PT_2 -Gliedes	94
8 Regelungssysteme	97
8.1 Der Standardregelkreis und Gütekriterien	98
8.2 Gütekriterien im Zeitbereich	100
9 Stabilität von Regelungssystemen	103
9.1 Stabilitätsstest mit Hilfe von Polynomkoeffizienten	104
9.2 Das Nyquist Kriterium	105
9.3 Innere Stabilität	112
10 Der PID Regler	113
10.1 Analyse des PID Gliedes	114
10.2 PID Einstellregeln nach Ziegler-Nichols	115
10.3 Eingangssaturation und Anti-Wind-Up	118
11 Reglerentwurf im Frequenzraum	121
11.1 Approximation des geschlossenen Kreises durch ein PT_2 -Glied	123
11.2 Vom geschlossenen zurück zum offenen Kreis	123
11.3 Korrekturglieder	124
12 Regelung im Zustandsraum	125
12.1 Steuerbarkeit	127
12.2 Polvorgabe	129
12.3 Arbeitspunktwechsel	131
13 Zustandsschätzung	133
13.1 Luenberger Beobachter und Beobachtbarkeit	133
13.2 Vollständiger Zustandsregelkreis	134
13.3 Das Kalman Filter	135
Anhang: Kleines Wörterbuch Deutsch-Englisch	137

Vorwort und Danksagung

Das vorliegende Skript wurde in den Sommersemestern 2013 und 2014 begleitend zur Vorlesung “Systemtheorie und Regelungstechnik 1” an der Universität Freiburg erstellt. Es ist in der Kapitelreihenfolge fast identisch mit den auf Video aufgezeichneten Tafelvorlesungen, aber enthält einige kleine Zusätze und Präzisierungen. Es ist bewusst knapp gehalten und soll - gemeinsam mit den Übungsblättern und den empfohlenen Lehrbüchern – zum Selbststudium und zur Klausurvorbereitung dienen.

Eine erste Version des Skriptes wurde 2013 geschrieben, mit Hilfe dreier Studenten - Moritz Berger, Fritz Koch und Nicolas Moll - die viele der Abbildungen beigetragen haben. Im Sommersemester 2014 wurde das Skript weiter vervollständigt und erweitert, insbesondere unter tatkräftiger Hilfe von Mario Zanon, der als erfahrener Doktorand aus dem Bereich der Regelungstechnik und Koordinator des Übungsgruppenbetriebes wichtige Anregungen, Textteile und Abbildungen beigetragen hat, sowie durch die vier Übungsgruppenleiter des Semesters, Thilo Bronnenmeyer, Heike Dietl, André Blickensdörfer und Lukas Klar, die viele Figuren des Skriptes erstellt und Verbesserungsvorschläge gegeben haben. Dank geht auch an die Studenten der Vorlesung, die durch aufmerksames Lesen und schlaue Fragen geholfen haben, Fehler im Skript zu finden.

Vielen Dank!

Freiburg, im Juli/August 2014

Moritz Diehl

Kapitel 1

Einführung

Die Systemtheorie betrachtet die Dynamik und Regelung technischer und natürlicher Systeme aus einer die Einzeldisziplinen überschreitenden Perspektive. Sie benutzt dazu die Sprache der Mathematik, und aus dieser Perspektive erkennt man starke Ähnlichkeiten zwischen auf den ersten Blick sehr verschieden erscheinenden Systemen. Beispielsweise haben aus Sicht der Systemtheorie ein Kondensator, eine sich erwärmende Herdplatte, ein rotierender Satellit, ein Pumpspeichersee, oder die Salzkonzentration in einer Körperzelle ein sehr ähnliches dynamisches Verhalten, und können mit den gleichen mathematischen Methoden analysiert werden.

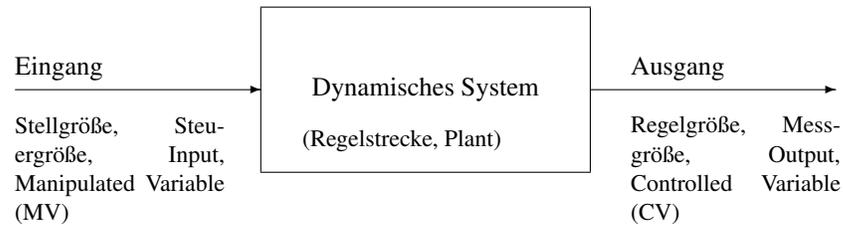
1.1 Systemtheorie und Regelungstechnik

Während sich die Systemtheorie mit der Modellierung, der Simulation und der Analyse der Eigenschaften eines Systems beschäftigt, beginnt die Steuerungs- und Regelungstechnik, sobald wir ein System nicht nur analysieren, sondern auch beeinflussen möchten. Typischerweise haben wir dafür bestimmte Eingangsgrößen zur Verfügung, die wir einstellen können. Diese werden oft “Stellgrößen” genannt, aber auch die Namen “Steuergrößen”, “Eingänge”, “Inputs” oder “Manipulated Variables (MV)” sind üblich. Man denke hierbei z.B. an das Steuerrad oder Gaspedal eines Autos, die Eingangsspannung eines elektrischen Motors oder eines Piezoelementes, oder die Insulinzufuhr bei einem Diabetes-Patienten.

Oft kann man durch Messungen überprüfen, ob der gewünschte Effekt auch tatsächlich eingetreten ist. Man kann z.B. die Position oder Geschwindigkeit eines Autos, die Drehzahl eines elektrischen Motors, die Verformung eines Piezoelementes, oder den Blutzuckerspiegel eines Menschen messen. Im Kontext der Regelung werden diese Größen als “Regelgrößen”, oder auch als “Messgrößen”, “Ausgänge”, “Outputs” oder “Controlled Variables (CV)” bezeichnet.

Das System selbst wird in dieser Eingangs-Ausgangs-Perspektive oft mit dem Namen “Regelstrecke” oder “Prozess” bezeichnet, oder im Englischen mit “Plant”. In dieser Vorlesung werden wir jedoch meist einfach nur mit dem Namen “System” verwenden. In der folgenden Abbildung ist ein solches ungeregeltes System illustriert, das

oft auch “offener Kreis” oder “open-loop system” genannt wird.



Wenn man den Zeitverlauf der Eingänge wählt, ohne diesen von den gemessenen Ausgängen abhängig zu machen, spricht man von “Steuerung”, oder auch “open-loop control”. Diese Art der Einflussnahme wird in vielen Bereichen des Lebens wie auch der Technik erfolgreich angewendet, z.B. beim Eierkochen, wenn man das Ei für eine im Vorhinein bestimmte Anzahl von Minuten ins kochende Wasser legt und es danach wieder herausnimmt. In einem Zelt auf dem Base-Camp des Mount-Everest würde das gekochte Ei mit dieser Methode jedoch überraschend weich sein, da der Luftdruck dort geringer ist und das Wasser bei niedrigerer Temperatur kocht. Dieses Ergebnis wäre nicht eingetreten, wenn wir beim Kochen den Zustand des Eies hätten testen könnten, um die Kochdauer gegebenenfalls zu verlängern - so wie wir es meist beim Kuchenbacken tun, wenn wir einen Stab in den gebackenen Kuchen stecken und überprüfen, ob noch Teig daran kleben bleibt.

Ein entscheidender Schritt zum Erreichen des gewünschten Ergebnisses auch unter geänderten Umständen ist das sogenannte “Feedback”, das im gesprochenen Deutsch meist mit dem englischen Wort bezeichnet wird, aber für das auch die deutschen Worte “Rückführung” oder “Rückkopplung” bestehen. Feedback besteht aus dem Vergleich der gemessenen Ausgangsgröße mit dem gewünschten Referenzwert, und eine dadurch ausgelöste Veränderung der Eingangsgröße durch den sogenannten “Regler”. Diese Rückkopplung mit Hilfe des Reglers hat meist das Ziel, die “Regelabweichung”, also die Differenz zwischen dem gemessenen Ausgang und dem Referenzwert, so klein wie möglich zu machen. Im Deutschen spricht man im Gegensatz zur Steuerung im rückgekoppelten Fall von “Regelung”. Im Englischen gibt es kein eigenes Wort und man spricht von “closed-loop control”. Ein geregeltes System ist in folgender Abbildung illustriert.



Interessanterweise kann das Verhalten des zurückgekoppelten Gesamtsystems, also das Übertragungsverhalten vom Referenzwert auf den Ausgangswert, wieder als ein neues System interpretiert werden, das oft als “geschlossener Regelkreis” oder “closed-loop system” bezeichnet wird. Dieses Gesamtsystem kann mit denselben mathematischen Methoden modelliert und analysiert werden wie das unregelte System aus der vorhergehenden Abbildung, auch wenn seine spezifischen Eigenschaften sich durch den Regeleingriff stark verändern können. Vielleicht die wichtigste dieser Veränderungen ist die Möglichkeit, mit Hilfe von Feedback ein instabiles System zu stabilisieren, wie wir es z.B. beim Gehen oder Fahrradfahren mit Hilfe unseres Gleichgewichtsorgans und unserer Augen tun, oder wie es das automatische Regelsystem eines Hubschraubers oder Flugzeuges tut: hier gibt der Pilot mit Hilfe des Steuerknüppels Signale, die als Eingänge in den Regler gegeben werden, der das System mit Hilfe eines Inertialsensors (einem Gyroskop) stabilisiert und den Vorgaben des Piloten folgen lässt. Da man das geregelte System selbst wieder als System auffassen kann, bildet die Systemtheorie die Grundlage für die Regelungstechnik, und wird auch in dieser Vorlesung zuerst behandelt. Unser letztendliches Ziel ist jedoch der Entwurf von Regelsystemen, die in der zweiten Hälfte der Vorlesung behandelt werden. Beispiele für einfache geregelte Systeme in unserem Haushalt sind der mechanisch zurückgekoppelte Wasserstandsregler im Toiletten-Spülkasten, das Thermostat im Heizkörper, oder der elektronische Raumtemperaturregler. Regler sind heute in fast jedem technischen System zu finden: jeder Fahrstuhl, jeder Motor, jedes Auto, jedes Flugzeug, jedes Mobiltelefon und jeder Computer sind mit vielerlei Regelkreisen ausgestattet und könnten ohne diese nicht betrieben werden. Auch wenn man diese Regelkreise, die meist in Form von eingebetteten Mikrocontrollern auftreten, oft gar nicht wahrnimmt: die Regelungstechnik ist eine der versteckten Technologien, die heute zum Betrieb und zur Entwicklung fast jeden technischen Gerätes unerlässlich sind.

1.2 Aufbau der Vorlesung

Die Vorlesung startet in Kapitel 2 mit der Modellierung dynamischer Systeme, um den Übergang von der wirklichen Welt in die mathematische Welt für einige ausgewählte technische Systeme im Detail zu vollziehen. Dabei werden wir von physikalischen Prinzipien ausgehend Zustandsraummodelle erstellen. Im Kapitel 3 werden wir eine andere Systemdarstellung kennenlernen, die sich auf die Eingänge und Ausgänge fokussiert. Wir spezialisieren uns dann in den folgenden Kapiteln auf den wichtigen Spezialfall linearer zeitinvarianter Systeme mit einem Eingang und einem Ausgang. Diese

Systeme heißen auf Englisch *linear-time-invariant single-input-single-output systems*, oder kurz *LTI-SISO Systeme*, und sind der Fokus dieses gesamten ersten Teils der Vorlesung Systemtheorie und Regelungstechnik. In Kapitel 4 zeigen wir, dass die Stabilitätseigenschaften von LTI-SISO Systemen durch das sogenannte “charakteristische Polynom” bestimmt werden. In Kapitel 5 lernen wir das dynamische Übertragungsverhalten von wichtigen LTI-SISO Systemen kennen. Wir führen dann in Kapitel 6 das mächtige mathematische Hilfsmittel der *Laplace-Transformation* ein, die es erlaubt, jedes LTI-SISO System vollständig durch eine einzige und einfach zu beschreibende Funktion zu beschreiben, die wir die *Übertragungsfunktion* nennen. In Kapitel 7 wird dann das wichtige und sehr praktische Konzept der Frequenzantwort eingeführt, die in einem engen Zusammenhang zur Übertragungsfunktion steht, und die durch Anregung des Systemeingangs mit sinusförmigen Signalen und durch Messen der Amplitude und Phase des Ausgangssignals experimentell erhalten werden kann.

In Kapitel 8 schließen wir die Regelschleife zum ersten Mal, und diskutieren das daraus resultierende Verhalten des geschlossenen Regelkreises anhand einiger Beispiele. Leider kann man durch schlechten Entwurf eines Regelkreises selbst ursprünglich stabile Systeme destabilisieren oder in unnötig große Oszillationen versetzen; deshalb ist die Stabilität des geschlossenen Regelkreises ein zentrales Problem der Regelungstechnik, das wir in Kapitel 9 adressieren. In Kapitel 10 diskutieren wir im Detail den in der regelungstechnischen Praxis meistverwendeten Regler, dem sogenannten PID-Regler, dessen Name sich vom Englischen *proportional-integral-derivative control* ableitet. Dieser Regler besteht aus drei zueinander parallelen Elementen, für deren optimale Einstellung viele in der Praxis entwickelte Rezepte bestehen, von denen wir einige kennenlernen. Im Kapitel 11 diskutieren wir den systematischen Entwurf von Reglern im Frequenzraum, mit Hilfe des Bode-Diagramms der offenen Kette. Das Kapitel 12 stellt eine gänzlich andere, modernere Methode des Reglerentwurfs dar, die nicht auf dem Ein-/Ausgangsverhalten des Systems beruht, sondern den gesamten inneren Systemzustand erst schätzt und dann mit Hilfe des Feedbacks verändert. Ein wichtiges Teilproblem dabei ist die Zustandsschätzung, die wir im letzten Kapitel, Kapitel 13, behandeln. Das Skript endet mit einem kleinen Wörterbuch Deutsch-Englisch.

1.3 Mathematischer Hintergrund und Literatur

Die Systemtheorie nutzt die Mathematik als Sprache, deren solide Kenntnis deshalb unerlässlich zum Verständnis dieser Vorlesung ist. Der Grund für die wichtige Rolle der Mathematik ist, dass man durch Übersetzung eines realen Problems in die mathematische Welt dieses oft viel schneller und besser lösen kann. In der mathematischen Welt stehen einem sehr mächtige Hilfsmittel zur Verfügung, derer man sich sowohl mit Papier und Bleistift, als auch mit dem Computer bedienen kann. Der Prozess der Hin- und Rückübersetzung von der realen physikalischen Welt in die mathematische Welt ist für die praktische Verwendbarkeit der Methoden sehr wichtig (siehe Abb. 1.1), und spielt deshalb in der Vorlesung eine große Rolle.

Der Kurs ist so aufgebaut, dass er genau mit den mathematischen Grundbegriffen auskommt, die in der Vorlesung und dem Skript zur “Mathematik I+II” (von Ernst Kuwert) vermittelt wurden. Zudem nutzen wir viele Resultate aus der Vorlesung “Dif-

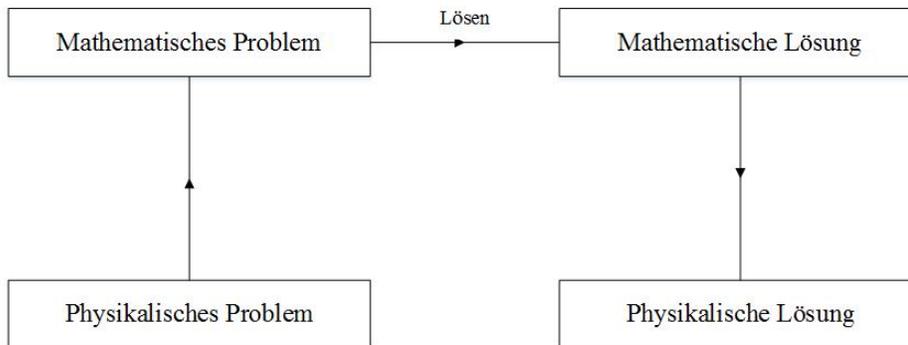


Abbildung 1.1: Die mathematische Modellierung von Problemen ist ein wichtiger Schritt auf dem Weg zu ihrer Lösung.

ferentialgleichungen für Studierende der Mikrosystemtechnik“, in der insbesondere bereits die für uns wichtige Laplace-Transformation eingeführt worden ist. Es wird den Teilnehmern dieser Vorlesung empfohlen, die Skripte dieser beiden vorhergehenden Vorlesungen noch einmal genau durchzusehen und etwaige Lücken durch das Gespräch mit Kommilitonen, durch Selbststudium, oder notfalls durch eine Anfrage zur kurzen Stoffwiederholung in der aktuellen Vorlesung zu schließen.

Dieses Vorlesungsskript ist als ein selbstständiger Text konzipiert, der jedoch sehr knapp ist und kaum mehr Inhalt enthält als der Tafelanschrieb der Vorlesung. Deshalb empfehlen wir den Studenten, sich mindestens eines der folgenden drei Lehrbücher zusätzlich zum Vorlesungsskript zu besorgen und zum Selbststudium zu verwenden.

Wir empfehlen zum Ersten das folgende, sehr ausführlich und schön geschriebene englischsprachige Lehrbuch:

G.F. Franklin, J.D. Powell, A. Emami-Naeini: Feedback Control of Dynamic Systems, Pearson (ISBN-13: 978-0-13-601969-5)

Zudem empfehlen wir zwei deutschsprachige Bücher, die ebenso dem vertiefenden Selbststudium dienen können. Dies ist zum einen der deutschsprachige Klassiker

O. Föllinger: Regelungstechnik, Hüthig, Heidelberg,

und zum anderen das folgende neuere Lehrbuch, das eine ähnliche Notation wie die Vorlesung verwendet, und damit zum Vertiefen des Lehrstoffes sehr gut geeignet ist:

J. Lunze: Regelungstechnik 1 - Systemtheoretische Grundlagen, Analyse und Entwurf einschleifiger Regelungen, Springer

Alle drei Bücher sind in Papierform in der Universitätsbibliothek vorhanden. Das letztgenannte Buch von Jan Lunze ist dort auch als online Version erhältlich.

1.4 Mathematische Notation

In diesem Skript benennen wir mit \mathbb{R} die Menge der reellen Zahlen, mit \mathbb{R}_+ die nicht-negativen, also inklusive der Null, und mit \mathbb{R}_{++} die positiven reellen Zahlen. Geschlossene Intervalle zwischen zwei reellen Zahlen a und b , die die Zahlen enthalten, bezeichnen wir mit eckigen Klammern, wir definieren $[a, b] := \{x | a \leq x \leq b\}$. Offene Intervalle, in denen die Zahlen nicht enthalten sind, bezeichnen wir mit runden Klammern, also $(a, b) = \{x | a < x < b\}$, und auch Mischungen wie $[a, b) = \{x | a \leq x < b\}$ sind möglich. Das Symbol ∞ bezeichnet “Unendlich” und kann auch als Intervallgrenze auftreten, z.B. bezeichnet das Intervall $[0, \infty)$ dasselbe wie \mathbb{R}_+ .

Das Symbol \mathbb{Z} bezeichnet die Menge der ganzen Zahlen, und \mathbb{N} die Menge der natürlichen Zahlen inklusive der Null, also $\mathbb{N} = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$. Die Menge der reellwertigen Vektoren der Dimension $n \in \mathbb{N}$ bezeichnen wir mit \mathbb{R}^n , und mit $\mathbb{R}^{n \times m}$ die Menge der Matrizen mit n Zeilen und m Spalten. Vektoren im \mathbb{R}^n sind in dieser Vorlesung immer “stehende”, oder Spaltenvektoren, d.h. wir identifizieren \mathbb{R}^n mit $\mathbb{R}^{n \times 1}$ und schreiben ein $x \in \mathbb{R}^n$ detailliert als

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Vektoren können durch Transponieren zu “liegenden” Zeilenvektoren werden, z.B. gilt

$$x^\top = [x_1 | x_2 | \dots | x_n].$$

Es gilt die normale Matrizenmultiplikation. Wenn wir die Größe eines Vektors $x \in \mathbb{R}^n$ messen wollen, nehmen wir meist die Euklidische Norm, die wir mit $\|x\|_2$ bezeichnen, und die durch $\|x\|_2 = \sqrt{x^\top x}$ definiert ist.

Kleinbuchstaben wie z.B. “ x ” bezeichnen meist Vektoren oder Skalare, und Grossbuchstaben wie z.B. “ A ” meist Matrizen. Griechische Buchstaben wie $\alpha, \beta, \gamma, \delta, \lambda, \sigma$ sind fast immer Skalare. Wir wenden diese Konventionen nicht dogmatisch an, und definieren die Matrixdimension der jeweiligen Größen explizit, z.B. durch $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$ oder $\alpha \in \mathbb{R}$, und erlauben uns bei Bedarf auch von der Konvention abweichende Definitionen wie z.B. $T \in \mathbb{R}$.

In diesem Skript werden wir viel mit Funktionen der Zeit zu tun haben, die wir z.B. mit $f(t)$ bezeichnen, oder genauer mit

$$f : \mathbb{R} \rightarrow \mathbb{R}, t \mapsto f(t).$$

Da Ableitungen nach der Zeit sehr häufig vorkommen, nutzen wir eine in den Natur- und Ingenieurwissenschaften bestehende Konvention und schreiben oft $\dot{f}(t)$ statt $f'(t)$, um explizit auszudrücken, dass es sich um eine Ableitung nach der Zeit handelt, d.h. wir definieren

$$\dot{f}(t) := \frac{\partial f}{\partial t}(t).$$

Diese Definition gilt auch für höhere Ableitungen, also z.B.

$$\ddot{f}(t) := \frac{\partial^2 f}{\partial t^2}(t).$$

Manche der Funktionen werden nicht nur skalare Werte haben, sondern vektorwertig sein, z.B.

$$x : \mathbb{R} \rightarrow \mathbb{R}^n, t \mapsto x(t),$$

und die Zeitableitung $\dot{x}(t)$ ist dann entsprechend komponentenweise definiert.

Ab Kapitel 6 werden wir neben einer zeitabhängigen Funktion $f(t)$ auch deren Laplacetransformierte benutzen, die wir dann mit $F(s)$ bezeichnen, und die durch

$$F(s) = \int_0^{\infty} f(t) e^{st} dt$$

definiert ist. Dabei benutzen wir den gleichen Buchstaben wie für die Zeitfunktion, aber als Grossbuchstaben, und als Argument immer (s) statt (t) . Die Laplacetransformierte ist eine komplexwertige Funktion $F : \mathbb{C} \rightarrow \mathbb{C}, s \mapsto F(s)$. Hierbei bezeichnet \mathbb{C} die Menge der komplexen Zahlen, $\mathbb{C} = \{x + jy | x, y \in \mathbb{R}\}$ (siehe Abb. 1.2). In diesem Skript ist j der Name für die imaginäre Einheit, $j = \sqrt{-1}$, wie in der Elektro- und Regelungstechnik üblich. Die Zahl e ist die Eulersche Zahl, $e = 2,718\dots$, und e^z bezeichnet dementsprechend die komplexe Exponentialfunktion, also $e^z = \exp(z)$, für die z.B. die Eulersche Formel $e^{jz} = \cos z + j \sin z$ gilt.

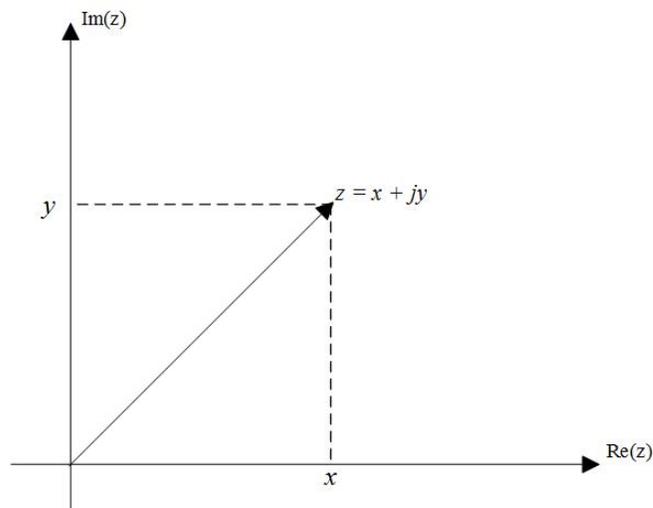


Abbildung 1.2: Darstellung von komplexen Zahlen.

Kapitel 2

Modellierung dynamischer Systeme

In der Systemtheorie betrachten wir gesteuerte Systeme, also dynamische Systeme, deren Eingänge wir frei wählen können. Diese zeitabhängigen Eingangsgrößen nennen wir in der gesamten Vorlesung $u(t)$. Diese sind die “Stellgrößen”, die wir auch manchmal mit den Namen “Steuerungen” oder “Kontrollen” bezeichnen werden. Typische Steuerungen $u(t)$ sind z.B. der Lenkwinkel oder die Gaspedalstellung eines Autos. Ein gesteuertes System kann eine oder mehrere Steuerungen haben. Falls es $m \in \mathbb{N}$ Steuerungen gibt, kann man $u(t)$ als vektorwertige Funktion auffassen, also als $u : \mathbb{R} \rightarrow \mathbb{R}^m, t \mapsto u(t)$ mit

$$u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{bmatrix}$$

Ein Beispiel für ein System mit $m = 2$, also zwei Steuerungen, könnte ein Auto sein, dessen Lenkwinkel wir mit $u_1(t)$ bezeichnen, und dessen Gaspedalstellung mit $u_2(t)$. Wir werden in dieser Vorlesung jedoch fast ausschließlich Systeme mit einem einzigen Eingang behandeln, so dass im folgenden fast immer angenommen wird, dass $u(t)$ skalar ist, also $m = 1$. Um beim obigen Beispiel des Autos zu bleiben, könnte man zur Vereinfachung annehmen, dass die Geschwindigkeit konstant ist, wie z.B. bei einem Traktor mit konstantem Gang auf einem ebenen Feld, und dass nur der Lenkwinkel als Steuerung zur Verfügung steht.

Es ist erstaunlich und faszinierend, welche Vielfalt von Systemverhalten man mit einer einzigen Steuerung erreichen kann. Um im Beispiel des Traktors mit konstanter Geschwindigkeit zu bleiben, kann man allein durch Wahl des Lenkwinkels entscheiden, ob man auf dem Feld weiter seine Runden dreht, ob man zum Freiburger Münsterplatz fährt und dort eine Runde dreht, oder ob man den Traktor in einen benachbarten Fluss oder gegen das nächstgelegene Bauernhaus fährt und damit die Fahrt beendet.

Dieser letzte Fall tritt fast unausweichlich ein, wenn man mit geschlossenen Augen fährt. Er kann und sollte natürlich durch Regelung, also Rückkopplung mit Hilfe von Sensorinformationen, vermieden werden. Diese ist jedoch noch nicht Thema dieses ersten Teils der Vorlesung, und wird erst ab Kapitel 8 wieder behandelt werden. Im Folgenden betrachten wir zunächst also nur open-loop gesteuerte Systeme, und untersuchen, wie sie sich verhalten, wenn man bestimmte Steuerungsinputs auf sie anwendet.

2.1 Gewöhnliche Differentialgleichungen

Die meisten dynamischen Systeme in Technik und Naturwissenschaften lassen sich sehr gut durch sogenannte “gewöhnliche Differentialgleichungen” beschreiben, die auf Englisch “Ordinary Differential Equations (ODE)” heissen. Die wichtigste Größe in einer solchen Differentialgleichung ist der sogenannte Systemzustand, den wir mit $x(t)$ bezeichnen. Dieser besteht fast immer aus mehreren Komponenten und ist somit ein Vektor, $x(t) \in \mathbb{R}^n$, wobei wir $n \in \mathbb{N}$ die “Zustandsdimension” nennen. Der Zustand ist das Gedächtnis des Systems, und fasst alle Informationen aus der Systemvergangenheit zusammen, die nötig sind, um das zukünftige Systemverhalten vorherzusagen.

Ein Beispiel aus der Physik wäre das Sonnensystem, das ein dynamisches System ohne Steuerungen ist. Wenn man die genauen Orte und Geschwindigkeiten aller acht Planeten (Merkur, Venus, Erde, Mars, Jupiter, Saturn, Uranus, Neptun) zu einem Zeitpunkt kennt, dann kann man mit Hilfe der Newtonschen Gesetze die zukünftige Planetenbewegung vorhersagen. Die Zustandsdimension n des Planetensystems ist wegen der drei Ortskoordinaten und drei Geschwindigkeitskomponenten für jeden der acht Planeten $n = 8 \cdot (3 + 3) = 48$. Man kann im Prinzip aus diesen 48 Zahlen die Planetenpositionen für jede Sekunde der folgenden Jahrzehnte berechnen.

Ein Beispiel für ein gesteuertes System im Zusammenhang mit der Astronomie wäre ein Satellit, dessen Bewegung man durch drei zueinander orthogonale Antriebsdüsen beeinflussen kann. Die Bewegung des Satelliten in seinem Orbit rund um die Erde ist wieder durch seine dreidimensionale Position und Geschwindigkeit, also durch sechs Zahlen, eindeutig bestimmt, aber seine Geschwindigkeit kann durch die drei Düsen verändert werden. Für den Systemzustand gilt hier also $x(t) \in \mathbb{R}^6$ und für die Steuerung $u(t) \in \mathbb{R}^3$.

Ein gesteuertes dynamisches System kann generell durch die folgende Form von Differentialgleichung auf einem Intervall $[0, t_e]$ beschrieben werden.

$$\dot{x}(t) = f(x(t), u(t)), \quad \text{für } t \in [0, t_e], \quad (2.1)$$

$$\text{mit } x(0) = x_0, \quad (2.2)$$

wobei wir $x_0 \in \mathbb{R}^n$ als Anfangswert bezeichnen. Man beachte, dass die Funktion f aus n Komponenten besteht, und im Detail auch geschrieben werden kann als $f :$

$\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, mit

$$f(x, u) = \begin{bmatrix} f_1(x, u) \\ f_2(x, u) \\ \vdots \\ f_n(x, u) \end{bmatrix}.$$

Wir werden die obige Differentialgleichung zur Vereinfachung manchmal auch ohne das Argument t und ohne den Anfangswert schreiben, also in der Form

$$\dot{x} = f(x, u).$$

Aus der Vorlesung über Differentialgleichungen wissen wir, dass die Zustandstrajektorie $x : [0, t_e] \rightarrow \mathbb{R}^n$ eindeutig bestimmt ist, wenn der Anfangswert $x_0 \in \mathbb{R}^n$ sowie die Inputtrajektorie $u : [0, t_e] \rightarrow \mathbb{R}^m, t \mapsto u(t)$ bekannt sind. Existenz und Eindeutigkeit dieser Lösung und ihre Existenz auf dem Intervall $[0, t_e]$ gelten unter einigen technischen Annahmen, die im Satz von Picard und Lindelöf (1894) beschrieben sind. Diese sind für gut formulierte Modelle realer Systeme mit realistischen Eingangssignalen fast immer erfüllt. Aber anstatt uns in mathematischen Details zu verlieren, wollen wir jetzt ein erstes Beispielsystem genauer betrachten, das uns noch eine Weile begleiten wird.

2.2 Modell eines gesteuerten Traktors

Wir betrachten einen Traktor, der mit konstanter Geschwindigkeit V auf einem ebenen Feld fährt, und dessen Lenkwinkel α wir einstellen können (siehe Abb. 2.1). Da der Traktor langsam fährt und wir starke Arme haben, nehmen wir an, dass wir den Lenkwinkel $\alpha(t)$ in jedem Moment t frei wählen können. Wir möchten nun vorhersagen, wohin der Traktor fährt, wenn wir einen bestimmten Verlauf für die Steuerung $\alpha(t)$ anlegen.

Die erste Frage ist: was ist der Systemzustand? Sicher müssen wir wissen, wo sich der Traktor auf dem Feld befindet. Wir wählen dafür die Position des Achsmittelpunktes der Hinterachse relativ zu einer Ecke des Feldes, und können dann die X- und Y-Koordinate dieses Punktes angeben, die wir X und Y nennen. In welche Richtung verändert sich diese Position? Dafür müssen wir die Orientierung des Traktors auf dem Feld kennen, die wir mit dem Winkel β bezeichnen. Dies ist der Winkel, den die Längsachse des Traktors mit der X-Koordinatenachse, z.B. der südlichen Kante des Feldes, bildet. Mit Hilfe der (konstanten) Geschwindigkeit V erhalten wir die zwei Differentialgleichungen

$$\dot{X} = V \cos \beta, \quad (2.3)$$

$$\dot{Y} = V \sin \beta, \quad (2.4)$$

die die Bewegung des hinteren Achsmittelpunktes auf der Ebene des Feldes beschreiben. Der Traktor mit seinen zwei Achsen und den vorkommenden Winkeln ist in Abbildung 2.1 skizziert.

Ein dritter Zustand ist durch die Orientierung β des Traktors selbst geben, denn der Traktor kann nicht willkürlich auf der Stelle gedreht werden, sondern seine Orientierung verändert sich langsam beim Fahren. Der Orientierungswinkel β erhöht sich,

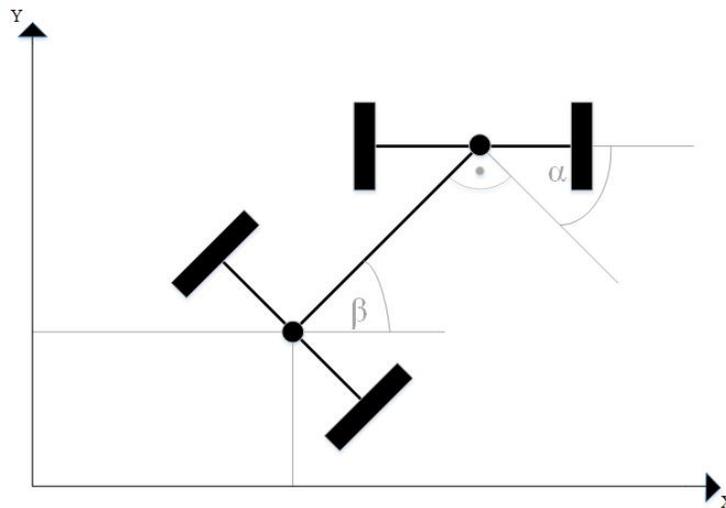


Abbildung 2.1: Visualisierung der Zustände des kinematischen Traktormodells.

wenn wir nach links steuern, und verringert sich, wenn wir nach rechts steuern. Wir wählen die Definition des Lenkwinkels α so, dass er positiv für eine Linkskurve ist. Eine trigonometrische Analyse zeigt, dass der Kurvenradius, oder Wendekreis, durch $R = L / \tan \alpha$ gegeben ist, wobei L der Abstand von der Vorder- zur Hinterachse ist. Da der Traktor bei gegebenem Lenkwinkel α mit der Geschwindigkeit V auf einer Kreiskurve mit Radius R fährt, wächst der Orientierungswinkel β gemäß der Formel $\dot{\beta} = V/R$ bzw.

$$\dot{\beta} = \frac{V}{L} \tan \alpha. \quad (2.5)$$

Damit haben wir bereits alle Zustände und ihre Differentialgleichungen gefunden. Die Zustandsdimension n ist drei, und der Zustandsvektor x ist gegeben durch

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ \beta \end{bmatrix}.$$

Die Steuerung ist gegeben durch

$$u = [\alpha].$$

Das gesamte Differentialgleichungssystem

$$\dot{X}(t) = V \cos \beta(t), \quad (2.6)$$

$$\dot{Y}(t) = V \sin \beta(t), \quad (2.7)$$

$$\dot{\beta}(t) = \frac{V}{L} \tan \alpha(t). \quad (2.8)$$

kann nun durch Definition der Funktion

$$f(x, u) = \begin{bmatrix} V \cos x_3 \\ V \sin x_3 \\ \frac{V}{L} \tan u \end{bmatrix}$$

kompakt geschrieben werden als

$$\dot{x} = f(x, u),$$

bzw. detailliert als

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} V \cos x_3 \\ V \sin x_3 \\ \frac{V}{L} \tan u \end{bmatrix}.$$

Der Anfangswert x_0 ist gegeben durch die drei Zahlen X_0, Y_0, β_0 , die die Position und Orientierung des Traktors zum Zeitpunkt Null beschreiben:

$$x_0 = \begin{bmatrix} X_0 \\ Y_0 \\ \beta_0 \end{bmatrix}.$$

Wir könnten das obige Differentialgleichungssystem nun mit Hilfe numerischer Simulationsroutinen lösen, z.B. durch einen ODE Löser in MATLAB wie `ode45` oder `ode15s`, aber wir wollen es für dieses Beispiel und den Spezialfall einer konstanten Steuerung einmal analytisch berechnen.

Wenn wir nämlich den Lenkwinkel α , also den Eingang u , konstant auf einem fixen Wert α_0 halten, ergibt sich aus Gleichung (2.8)

$$\dot{\beta} = \frac{V}{L} \tan \alpha_0 = \frac{V}{R_0} = \text{const},$$

wobei wir den zum Winkel α_0 gehörenden Wendekreisradius $R_0 = L / \tan \alpha_0$ benutzt haben. Es ergibt sich durch Integration

$$\beta(t) = \beta_0 + \int_0^t \frac{V}{R_0} d\tau,$$

also

$$\beta(t) = \beta_0 + \frac{V}{R_0} t.$$

Der Orientierungswinkel des Traktors wächst also linear mit der Zeit. Für die X-Position $X(t)$, die bei X_0 startet, gilt nach Gleichung (2.6)

$$\dot{X}(t) = V \cos \left(\beta_0 + \frac{V}{R_0} t \right).$$

Dies kann man analytisch integrieren, mit dem Ergebnis

$$X(t) = X_0 + R_0 \sin \left(\beta_0 + \frac{V}{R_0} t \right).$$

Zur Überprüfung kann man die untere Zeile ableiten und sieht, dass die obere dabei herauskommt. Zudem haben wir durch Addition von X_0 erreicht, dass $X(0) = X_0$ gilt. Für $Y(t)$, das der Gleichung

$$\dot{Y}(t) = V \sin\left(\beta_0 + \frac{V}{R_0} t\right).$$

folgt, kann man analog berechnen, dass

$$Y(t) = Y_0 - R_0 \cos\left(\beta_0 + \frac{V}{R_0} t\right).$$

Der Traktor fährt tatsächlich einen Kreis mit dem Radius R_0 , startend mit der am Anfang gegebenen Position und Orientierung. Er tut dies mit einer Kreisfrequenz von $\frac{V}{R_0}$, braucht also für eine Rundfahrt die Zeit $R_0/(2\pi V)$.

Bei anderen, zeitveränderlichen Signalen $\alpha(t)$ kann der Traktor ganz andere Wege fahren. Als Beispiel für ein zeitabhängiges Steuersignal $\alpha(t)$ könnte folgende Anweisung stehen: halte α für drei Sekunden auf 5 Grad, dann zwei Minuten auf 0 Grad, danach 4 Sekunden auf -10 Grad, etc. Der Traktor wird dann, wie zuvor schon erwähnt, sehr wahrscheinlich schnell im nächsten Bach landen, da man $\alpha(t)$ quasi mit geschlossenen Augen, also ohne Feedback, einstellt. Dieses Fahrverhalten ist in Abbildung 2.2 dargestellt.

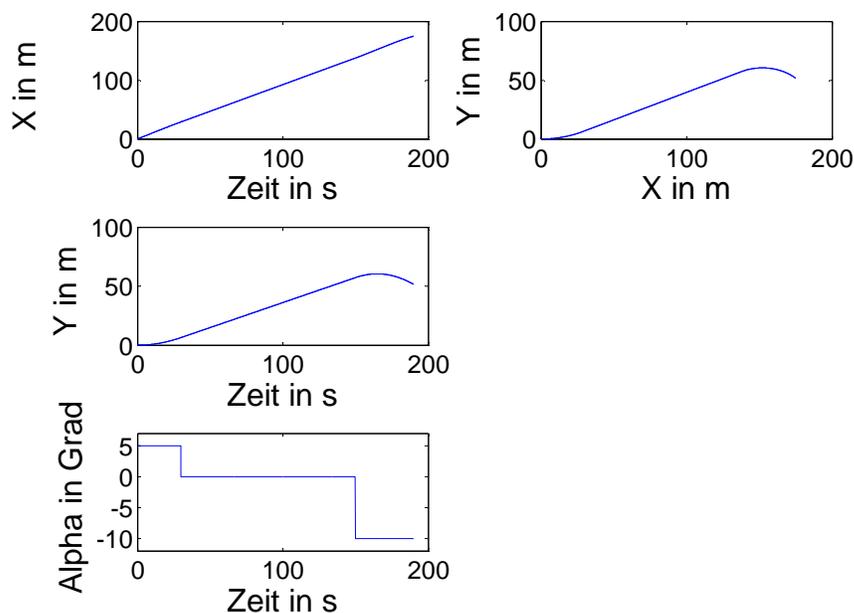


Abbildung 2.2: Simulation des kinematischen Traktormodells.

2.3 Physikalische Einheiten und Skalierung

Im vorigen Abschnitt haben wir noch nicht definiert, welche Einheiten die Zustände haben. Das ist eigentlich auch kein Problem, da die verwendeten und hergeleiteten Formeln unabhängig von den gewählten Einheiten gültig sind. Einsteins Formel $E = mc^2$ kann sowohl in Amerika als auch in Europa auf genau die gleiche Weise geschrieben werden, obwohl wir verschiedene Einheiten für die Lichtgeschwindigkeit verwenden. Alle durch analytische Umformungen - z.B. durch Ableiten und Integrieren - erhaltenen Gleichungen sind unabhängig von den Einheiten gültig.

Sobald wir jedoch etwas konkret per Hand oder auf dem Computer berechnen wollen, müssen wir uns für bestimmte Einheiten entscheiden. Für die Wahl der Einheiten ist es bei der Modellierung extrem hilfreich, mit den sogenannten SI-Einheiten (aus dem Französischen *ystème international*), also mit Meter [m], Sekunde [s], Kilogramm [kg], Newton [N], Joule [J], Watt [W], Pascal [Pa], Ampere [A], Volt [V], etc. zu arbeiten, da diese Einheiten so aufeinander abgestimmt sind, dass beim Rechnen mit reinen Zahlen keine Korrekturfaktoren eingeführt werden müssen, z.B. gilt $1 \text{ W} = 1 \text{ A} \cdot 1 \text{ V}$ oder $1 \text{ J} = 1 \text{ W} \cdot 1 \text{ s}$ oder $1 \text{ Pa} = 1 \text{ N} / 1 \text{ m}^2$, etc. Damit kann man mit den einheitenbehafteten physikalischen Formeln so rechnen, als ob es sich um reine Zahlen handelte. Es ist jedoch extrem wichtig, sich bei jeder Gleichung bewusst zu sein, in welchen Einheiten sie geschrieben ist. Durch nachträgliche Einheitenkontrolle kann man leicht Fehler finden. Hätten wir z.B. im Traktormodell als letzte Gleichung durch einen Rechen- oder Modellierungsfehler statt (2.8) die Gleichung

$$\dot{\beta} = V \tan \alpha$$

erhalten, könnten wir direkt sehen, dass die Einheiten der linken und rechten Seite nicht zusammenpassen, denn auf der linken Seite haben wir Winkel durch Zeit, z.B. rad/s, und auf der rechten Seite Länge durch Zeit, z.B. m/s.

Für unser Traktormodell könnten wir mit den folgenden SI-Einheiten rechnen: X und Y werden in Meter [m] gemessen, und α und β in der mathematischen Einheit Radian [rad], die einheitenfrei ist und für die glücklicherweise gilt, dass $1 \text{ rad} = 1$. Wir könnten nun einheitenfreie Bewegungsgleichungen einführen, indem wir für jede Zustandsgröße x_i eine einheitenfreie reelle Zahl $x_i^z \in \mathbb{R}$ einführen, ebenso für den Input u eine reelle Zahl u^z .

Für das Traktormodell ergeben sich dann mit den Zuordnungen $X = X^z \text{m}$, $Y = Y^z \text{m}$, aus den Gleichungen (2.6)-(2.8) die folgenden, fast einheitenfreien Gleichungen

$$\dot{X}^z(t) = \frac{V}{(\text{ms}^{-1})} \cos(\beta(t)) \frac{1}{\text{s}}, \quad (2.9)$$

$$\dot{Y}^z(t) = \frac{V}{(\text{ms}^{-1})} \sin(\beta(t)) \frac{1}{\text{s}}, \quad (2.10)$$

$$\dot{\beta}(t) = \frac{V}{(\text{ms}^{-1})} \frac{\text{m}}{L} \tan(\alpha(t)) \frac{1}{\text{s}}. \quad (2.11)$$

Hierbei haben wir für die Winkel die Tatsache ausgenutzt, dass diese bereits einheitenfreie Größen sind, die man ohne weitere Anpassung in trigonometrische Funktionen

einsetzen kann. Man beachte, dass die Zeiteinheit noch vorhanden ist, die obenstehenden Gleichungen also noch keine reinen Zahlengleichungen sind. Wenn wir jetzt noch beschliessen, die Zeit in Sekunden zu messen, also $t = t^z$ s setzen, und wenn wir die reellen Zahlen $V^z = \frac{V}{(\text{ms}^{-1})}$ und $L^z = \frac{L}{\text{m}}$ definieren, dann bekommen wir eine gänzlich einheitenfreie Differentialgleichung der Form:

$$\frac{dX^z}{dt^z}(t^z) = V^z \cos(\beta(t^z)), \quad (2.12)$$

$$\frac{dY^z}{dt^z}(t^z) = V^z \sin(\beta(t^z)), \quad (2.13)$$

$$\frac{d\beta}{dt^z}(t^z) = \frac{V^z}{L^z} \tan(\alpha(t^z)). \quad (2.14)$$

Diese einheitenfreien mathematischen Gleichungen sehen wieder genauso aus, wie die ursprünglichen physikalischen Gleichungen (2.6)-(2.8), was aber nur daran liegt, dass die SI-Einheiten so gut aufeinander abgestimmt sind. Der Vorteil, der dadurch entsteht, dass wir jetzt einheitenfreie Gleichungen haben, ist, dass wir das erhaltene Modell nun mit den mathematischen Methoden der Systemtheorie behandeln können. Man kann eigentlich erst jetzt eine wirklich einheitenfreie mathematische Funktion f^z definieren, $f^z : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$, die gegeben ist durch

$$f^z(x^z, u^z) = \begin{bmatrix} V^z \cos x_3^z \\ V^z \sin x_3^z \\ \frac{V^z}{L^z} \tan u^z \end{bmatrix}$$

mit dem Zustandsvektor $x^z \in \mathbb{R}^3$

$$x^z = \begin{bmatrix} X^z \\ Y^z \\ \beta \end{bmatrix},$$

und der Steuerung $u^z \in \mathbb{R}$ mit $u^z = \alpha$. Die Differentialgleichung kann dann als

$$\frac{dx^z}{dt^z} = f^z(x^z, u^z)$$

geschrieben werden. Wir werden in diesem Kurs die Notation beim Übergang vom physikalischen Modell zum mathematischen Modell manchmal etwas vereinfachen, und einfach x und x^z etc. miteinander identifizieren, wenn wir zu konkreten Zahlen übergehen. Als Zeiteinheit werden wir durchgehend die Sekunde nehmen.

Noch eine Bemerkung: aus numerischen Gründen ist es manchmal hilfreich, dass man das Modell nicht mit SI-Einheiten, sondern mit dem Problem besser angepassten Einheiten formuliert. Man nennt dies ‘‘Skalierung’’ oder ‘‘Normierung’’. Betrachtet man z.B. die Schwingung eines Quarzkristalls in einer Uhr, so führt die Verwendung der Sekunde als Zeiteinheit oder des Meters als Längeneinheit zu sehr kleinen bzw. sehr großen Zahlen in den Modellgleichungen, die bei numerischen Berechnungen Schwierigkeiten verursachen können. Misst man hingegen die Zeit in Mikrosekunden und die Länge in Mikrometern, kommt man auf weniger extreme Zehnerpotenzen in den Modellgleichungen, was nicht nur der Lesbarkeit hilft, sondern vor allem auch den numerischen Berechnungen auf dem Computer, der einen beschränkten Zahlenbereich hat (in Fließkomma-Arithmetik mit `double` Zahlen z.B. nur 16 gültige Dezimalstellen).

2.4 Mechanische Systeme

In der Mechanik bezeichnet man Modelle, die nur die Bewegungen, aber nicht die wirkenden Kräfte enthalten, als “kinematische Modelle”. Das im vorigen Abschnitt 2.2 eingeführte Traktormodell ist eigentlich ein “kinematisches Traktormodell”, da wir die wirkenden Kräfte nicht betrachtet haben und einfach angenommen haben, dass wir die Geschwindigkeit vorgeben können. Will man aber z.B. einen auf glattem Eis fahrenden Traktor modellieren, muss man die vom Boden auf den Reifen wirkende Kraft und die Trägheit des Traktors mitmodellieren, um die schlitternde Bewegung des Traktors realistisch vorhersagen zu können. Man spricht in der Mechanik von “dynamischen Modellen”, wenn Kräfte und die daraus resultierenden Beschleunigungen mitmodelliert werden. Die Modellierung komplexer mechanischer Systeme ist ein grosses Fachgebiet mit einer langen Geschichte und vielen aktuellen Forschungsaktivitäten, aber hier werden wir nur einige einfache Modelle besprechen. Gleich zu Beginn müssen wir eine für uns sehr wichtige Tatsache besprechen: nämlich dass der Systemzustand bei dynamischen Modellen nicht nur aus der Position X besteht, sondern auch aus der Geschwindigkeit \dot{X} . Dies liegt daran, dass ein Körper ohne Krafteinwirkung im Zustand der gleichförmigen Bewegung verbleibt, und dass eine Kraft F nach Newtons berühmter Formel $F = ma$ (mit $a = \ddot{X}$) nur die Beschleunigung $\ddot{X} = F/m$ eines Körpers mit der Masse m direkt beeinflussen kann, während die Geschwindigkeit daraus erst durch Integration erhalten wird.

2.4.1 Freie Bewegung eines Körpers im Raum

Will man die Bewegung eines sich frei im Raum bewegenden Körpers durch eine Kraft F beeinflussen, die wir in diesem Unterabschnitt als Systemeingang u betrachten, so kann man das zukünftige Verhalten nur dann vorhersagen, wenn man neben der Anfangsposition X auch die Anfangsgeschwindigkeit \dot{X} kennt. Man denke z.B. an einen auf Eis gleitenden Traktor oder an einen Satelliten mit Antriebsdüsen.

Die Differentialgleichung $\ddot{X} = F/m$ nennt man “eine Differentialgleichung von zweiter Ordnung”, da zweite Zeitableitungen in ihr vorkommen. Um das Modell in unsere Standardform $\dot{x} = f(x, u)$ zu bekommen, die eine Differentialgleichung von erster Ordnung beschreibt, müssen wir *zwei* Zustände definieren, nämlich neben der Position X auch die Geschwindigkeit \dot{X} die wir mit dem eigenen Symbol V benennen. Zusammen gehorchen diese zwei Zustände den folgenden zwei Gleichungen erster Ordnung:

$$\dot{X} = V, \quad (2.15)$$

$$\dot{V} = F/m. \quad (2.16)$$

Man kann diese in die Standardform $\dot{x} = f(x, u)$ bringen, wenn man die Zustände x , Steuerung u , und die Systemfunktion f wie folgt definiert:

$$x = \begin{bmatrix} X \\ V \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad u = F, \quad f(x, u) = \begin{bmatrix} V \\ F/m \end{bmatrix} = \begin{bmatrix} x_2 \\ u/m \end{bmatrix}.$$

Der entscheidende Schritt war die Einführung der Geschwindigkeit $V = \dot{X}$ als eigener Zustand. Diesen Trick kann man für alle Differentialgleichungen höherer Ordnung

anwenden: man muss alle Zeitableitungen bis auf die höchste in den Gleichungen vorkommende als Zustände wählen.

Um das Verhalten dieses Systems für gegebene Anfangswerte X_0 und V_0 und gegebenen Steuerungsinput $F(t)$ zu simulieren, kann man die beiden Gleichungen einfach integrieren: aus $\dot{V}(t) = F(t)/m$ erhält man

$$V(t) = V_0 + \frac{1}{m} \int_0^t F(\tau) d\tau$$

und aus $\dot{X}(t) = V(t)$ erhält man

$$X(t) = X_0 + \int_0^t V(\tau) d\tau.$$

Bei konstanter Kraft F erhält man den expliziten Ausdruck

$$X(t) = X_0 + V_0 t + \frac{F}{2m} t^2.$$

2.4.2 Der harmonische Oszillator

Oft können wir die auf einen Körper wirkende Kraft F nicht direkt vorgeben, sondern nur durch ein elastisches Element, z.B. eine Feder auf den Körper übertragen. Wenn wir z.B. die Ruheposition u der Feder vorgeben können, ist die durch eine lineare Feder erzeugte Kraft gegeben durch

$$F = k(u - X)$$

wobei k die sogenannte Federkonstante ist, die die SI-Einheit [N/m] hat. Das System wird dann beschrieben durch die Differentialgleichung

$$\ddot{X} = \frac{k}{m}(u - X), \quad (2.17)$$

oder, falls wir das System wieder in unsere Zustands-Standardform bringen, durch die zwei Gleichungen

$$\dot{X} = V, \quad \text{und} \quad \dot{V} = \frac{k}{m}(u - X).$$

Wir werden im folgenden Kapitel lernen, wie wir das dynamische Verhalten dieses und anderer linearer Systeme analytisch berechnen können. Wir erwähnen hier jedoch bereits, dass das System für $u = 0$ Lösungen der Form $X(t) = A \sin \omega_0 t$ zulässt. Durch Ableiten erhalten wir

$$V(t) = \dot{X}(t) = A\omega_0 \cos \omega_0 t$$

und durch nochmaliges Ableiten

$$\dot{V}(t) = \ddot{X}(t) = -A\omega_0^2 \sin \omega_0 t.$$

Durch Einsetzen dieses Lösungsansatzes in (2.17) erhalten wir, mit $u = 0$,

$$-A\omega_0^2 \sin \omega_0 t = \frac{k}{m} (-A \sin \omega_0 t)$$

und sehen, dass dies für beliebige Amplituden A erfüllt werden kann, sobald für die Kreisfrequenz ω_0 gilt:

$$\omega_0^2 = \frac{k}{m}.$$

Wir haben also ein schwingungsfähiges System erhalten, das umso schneller oszilliert, je stärker die Feder ist und je kleiner die Masse ist. Dieses System wird uns noch in ganz anderen physikalischen Zusammenhängen begegnen, und wird oft als "harmonischer Oszillator" bezeichnet. Die Frequenz $\omega_0 = \sqrt{\frac{k}{m}}$ nennt man seine "Eigenfrequenz".

2.4.3 Satellitenmodell im dreidimensionalen Raum

Interessanterweise kann man für dreidimensionale Bewegungen den Vektor X direkt als dreidimensionale Größe definieren, und die Geschwindigkeit und Kraft ebenso. Die Zustandsdimension ist dann sechs, und der Eingang hat die Dimension drei, die Kraft F^D der Düse. Es gilt also

$$x = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ V_1 \\ V_2 \\ V_3 \end{bmatrix} \quad \text{und} \quad u = \begin{bmatrix} F_1^D \\ F_2^D \\ F_3^D \end{bmatrix}.$$

Abbildung 2.3 zeigt die Bewegung eines Satelliten mit Masse m im Gravitationspotential der Erde.

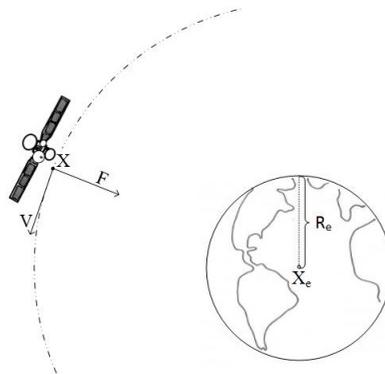


Abbildung 2.3: Visualisierung der Satellitenbewegung um die Erde: Ort X , Geschwindigkeit V und die auf den Satelliten wirkende Gesamtkraft F .

Der Positionsvektor des Satelliten im dreidimensionalen Raum sei X_e und neben der durch die Düsen erzeugten Kraft u wirke auch die Erdanziehungskraft, dann gilt folgende Formel für die Beschleunigung des Satelliten

$$\dot{V} = \frac{F}{m} = \frac{1}{m} \left(u - \frac{mgR_e^2}{\|X - X_e\|^3} (X - X_e) \right) \quad (2.18)$$

wobei $g \approx 9,81 \text{ ms}^{-2}$ die Erdbeschleunigung und $R_e \approx 6367 \text{ km}$ der Erdradius ist. Die Funktion $f : \mathbb{R}^6 \times \mathbb{R}^3 \rightarrow \mathbb{R}^6$ wird damit zu

$$f(x, u) = \begin{bmatrix} V \\ \frac{u}{m} - \frac{gR_e^2}{\|X - X_e\|^3} (X - X_e) \end{bmatrix}.$$

Man beachte, dass der Vektor f aus zwei Untervektoren mit jeweils drei Komponenten besteht. Die Bewegung des Satelliten könnte man nun mit Hilfe numerischer Simulationsroutinen berechnen.

Im Fall der Bewegung in einer Ebene, und wenn $u = 0$, kann man die Gleichung durch elliptische Bahnen lösen, und es war eines der überzeugendsten Resultate von Newton, dass er mit seinem Gravitationsgesetz die zuvor von Kepler genau beschriebenen Ellipsen als Umlaufbahnen der Planeten um die Sonne erhielt. Wir wollen dies hier nicht wiederholen, aber im allereinfachsten Fall können wir eine kreisförmige ebene Bewegung des Satelliten annehmen, die die Form

$$X(t) = X_e + R \begin{bmatrix} \cos \omega_0 t \\ \sin \omega_0 t \\ 0 \end{bmatrix}$$

hat, und dann sehen, ob sie für bestimmte Werte von R und ω_0 die Differentialgleichung erfüllt. Es gilt (wie zuvor beim Oszillator) nach zweimaligem Ableiten

$$\dot{V} = -R\omega_0^2 \begin{bmatrix} \cos \omega_0 t \\ \sin \omega_0 t \\ 0 \end{bmatrix}.$$

Einsetzen in Gleichung 2.18 mit $u = 0$ ergibt

$$-R\omega_0^2 \begin{bmatrix} \cos \omega_0 t \\ \sin \omega_0 t \\ 0 \end{bmatrix} = -\frac{gR_e^2}{R^3} R \begin{bmatrix} \cos \omega_0 t \\ \sin \omega_0 t \\ 0 \end{bmatrix}.$$

Dies kann tatsächlich erfüllt werden, wenn

$$\omega_0^2 = \frac{gR_e^2}{R^3}$$

gilt. Ein auf einer Kreisbahn fliegender Satellit muss also umso schneller um die Erde fliegen, je tiefer die Kreisbahn gelegen ist. Eine interessante Frage in diesem Zusammenhang wäre, wie gross der Radius R sein müsste, wenn der Satellit sich an einem

Tag genau einmal um die Erde drehen soll, also auf einem geostationären Orbit fliegt. Für $\omega_0 = \frac{2\pi}{24 \cdot 3600\text{s}}$ ergibt sich

$$R = \left(\frac{gR_e^2}{\omega_0^2} \right)^{\frac{1}{3}} \approx 42 \cdot 10^3 \text{ km.}$$

was einer Höhe über dem Grund von $R - R_e \approx 35 \cdot 10^3 \text{ km}$ entspricht.

Für den allgemeinen Fall einer gesteuerten Bewegung im Gravitationsfeld der Erde kann man das Systemverhalten für gegebene Anfangswertvektoren X_0 und V_0 und für gegebenen Steuerungsinpult $u(t)$ am einfachsten numerisch bestimmen. Falls noch ein zweites Gravitationsfeld, z.B. das des Mondes, mitberücksichtigt werden muss, ist die numerische Berechnung die einzige Möglichkeit, das Systemverhalten zu simulieren. Typische Fragen, mit denen Ingenieure beim Planen und Durchführen der Apollo Mission zum Mond in den sechziger Jahren des 20. Jahrhunderts beschäftigt waren, sind "An welchem Ort ist das Raumschiff jetzt gerade?" oder "Mit welcher Geschwindigkeit wird das Raumschiff in die Mondumlaufbahn eintreten?" Die Raumfahrt war eine wichtige Triebfeder für die Weiterentwicklung der Systemtheorie und Regelungstechnik in der Mitte des vergangenen Jahrhunderts, und arbeitet mit Modellen wie dem in diesem Unterabschnitt beschrieben.

2.4.4 Rotationsbewegungen und Drehpendel

Ebenso wie bei Translationsbewegungen kann man auch bei Rotationen eine Geschwindigkeit definieren, die wir die Drehgeschwindigkeit nennen. Die Rolle der Masse übernimmt in diesem Fall das Trägheitsmoment, das wir oft mit I bezeichnen, und die Rolle der Kraft übernimmt das Drehmoment, das wir oft mit T bezeichnen, und dass die Einheit [Nm] hat, was zufällig die gleiche SI-Einheit wie die Energie ist.

Ein einfaches Drehpendel lässt sich z.B. wie folgt beschreiben: wir bezeichnen den Rotationswinkel mit $\phi(t)$, und die momentane Winkelgeschwindigkeit $\dot{\phi}(t)$ mit $\omega(t)$. Hat das Drehpendel das Trägheitsmoment I um seine Drehachse, und wirkt das Drehmoment $T(t)$, dann wird die Bewegung durch die triviale Differentialgleichung $\dot{\phi}(t) = \omega(t)$ sowie die Gleichung

$$\dot{\omega}(t) = \frac{T(t)}{I}$$

beschrieben. Wenn wir jetzt noch annehmen, dass der Schwerpunkt S des Drehpendels um eine Distanz L tiefer liegt als die horizontale Drehachse, dann ergibt sich ein rückstellendes Drehmoment, wie in Abbildung 2.4 zu sehen, das vom Auslenkungswinkel ϕ abhängt.

Wir definieren diesen Winkel ϕ so, dass er in der Ruhelage gleich Null ist, und es ergibt sich bei einer Masse m des Drehpendels das Drehmoment

$$T(t) = -mgL \sin \phi(t) + u(t)$$

wobei wir mit $u(t)$ das von außen als Steuerung hinzugefügte Drehmoment bezeichnen, das wir z.B. mit einem elektrischen Motor erzeugen können. Die zwei Differenti-

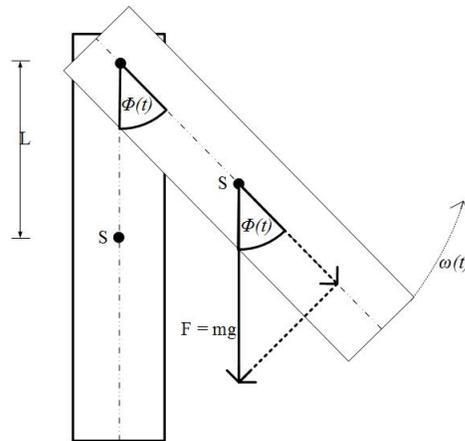


Abbildung 2.4: Drehpendel mit winkelabhängigem rückstellendem Drehmoment.

gleichungen des gesteuerten Drehpendels ergeben sich damit zu

$$\dot{\phi}(t) = \omega(t) \quad (2.19)$$

$$\dot{\omega}(t) = \frac{mgL}{I} \sin \phi(t) + \frac{1}{I} u(t). \quad (2.20)$$

Die Bewegung des Drehpendels bei bekannten Anfangswerten und Steuerungen können wir nicht sehr einfach analytisch berechnen. Wir können die Systemgleichungen für kleine Auslenkungswinkel ϕ jedoch vereinfachen, indem wir $\sin \phi \approx \phi$ setzen. Dadurch ergibt sich das folgende Differentialgleichungssystem

$$\dot{\phi}(t) = \omega(t), \quad (2.21)$$

$$\dot{\omega}(t) = \frac{mgL}{I} \phi(t) + \frac{1}{I} u(t), \quad (2.22)$$

das der Gleichung des harmonischen Oszillators stark ähnelt und für $u(t) = 0$ mit demselben Ansatz wie zuvor gelöst werden kann. Es handelt sich also wieder um ein schwingungsfähiges System, und die Eigenfrequenz ω_0 ist gegeben durch

$$\omega_0 = \sqrt{\frac{mgL}{I}}.$$

2.5 Lineare und Nichtlineare Systeme

Wir haben bereits jetzt gesehen, dass die meisten der vorgestellten Systemmodelle nichtlinear sind, also Terme enthalten, die nichtlinear von den Zuständen oder Steuerungen abhängen, wie z.B. das kinematische Traktormodell, das Satellitenmodell, oder das Drehpendel. Andere Modelle, wie z.B. die freie Bewegung im Raum oder der harmonische Oszillator waren linear, das heißt, dass die in der Gleichung vorkommenden

Terme alle linear von x und u abhängen. Eine lineare gewöhnliche Differentialgleichung mit n Zuständen und m Kontrollen lässt sich immer in der Form

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (2.23)$$

schreiben, mit Matrizen $A \in \mathbb{R}^{n \times n}$ und $B \in \mathbb{R}^{n \times m}$. Systeme dieser Form nennen wir lineare, zeitinvariante Systeme, oder kurz LTI-Systeme, nach dem englischen *linear time invariant*. LTI-Systeme sind der Fokus dieser Vorlesung, und dies aus zwei Gründen. Der erste ist, dass man ihr Verhalten viel leichter verstehen und vorhersagen kann als das von nichtlinearen Systemen. Der zweite Grund ist, dass man fast alle nichtlinearen Systeme in der Nähe einer Gleichgewichtslage sehr gut durch LTI-Systeme beschreiben kann.

Das zu einem nichtlinearen System $\dot{x} = f(x, u)$ gehörende LTI-System ergibt sich aus einer Taylorentwicklung erster Ordnung, wenn man annimmt, dass der Systemzustand $x(t)$ und die Steuerung $u(t)$ nur wenig von einem stationären Zustand, also einer Ruhelage abweichen. Bezeichnen wir diese Ruhelage, die im Englischen *steady state* genannt wird, mit x_{ss} und u_{ss} , dann betrachten wir

$$x(t) = x_{ss} + \delta x(t) \quad \text{und} \quad u(t) = u_{ss} + \delta u(t),$$

wobei $\delta x(t)$ und $\delta u(t)$ als klein angenommen werden dürfen. Die Linearisierung von $f(x, u)$ ergibt sich dann zu

$$f(x_{ss} + \delta x, u_{ss} + \delta u) \approx \underbrace{f(x_{ss}, u_{ss})}_{=0} + \frac{\partial f}{\partial x}(x_{ss}, u_{ss}) \delta x + \frac{\partial f}{\partial u}(x_{ss}, u_{ss}) \delta u.$$

Hierbei ist der erste Term Null, da wir angenommen haben, dass es sich bei x_{ss} und u_{ss} um einen stationären Zustand handelt, für den $\dot{x} = 0$ gelten muss, also $0 = f(x_{ss}, u_{ss})$. Da zudem

$$\frac{d(x_{ss} + \delta x(t))}{dt} = \underbrace{\frac{dx_{ss}}{dt}}_{=0} + \frac{d(\delta x(t))}{dt} = \delta \dot{x}(t)$$

ergibt sich aus der Linearisierung von $\dot{x}(t) = f(x(t), u(t))$ die linear zeitinvariante Differentialgleichung

$$\delta \dot{x}(t) = \frac{\partial f}{\partial x}(x_{ss}, u_{ss}) \delta x(t) + \frac{\partial f}{\partial u}(x_{ss}, u_{ss}) \delta u(t). \quad (2.24)$$

Die genau die Form von Gleichung (2.23) hat, wenn wir

$$A = \frac{\partial f}{\partial x}(x_{ss}, u_{ss}) \quad \text{und} \quad B = \frac{\partial f}{\partial u}(x_{ss}, u_{ss})$$

setzen. Als ein Beispiel haben wir bereits das Drehpendel gesehen, dessen Ruhelage durch geschickte Koordinatenwahl genau bei den Zuständen $x_1 = x_2 = 0$ und der Steuerung $u = 0$ lag, und dessen Systemgleichung durch

$$f(x, u) = \begin{bmatrix} x_2 \\ -\frac{mgL}{I} \sin x_1 + \frac{1}{I}u \end{bmatrix}$$

gegeben war. Durch Linearisierung an der Stelle $x_{ss} = [0 \mid 0]^T$ und $u_{ss} = 0$ erhält man

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{mgL}{I} & 0 \end{bmatrix} \quad \text{und} \quad B = \begin{bmatrix} 0 \\ \frac{1}{I} \end{bmatrix}.$$

Dieses Vorgehen kann man auf jedes nichtlineare System in der Nähe einer Gleichgewichtslage anwenden, und da wir es in der Regelungstechnik sehr oft mit kleinen Abweichungen zu tun haben, ist eine linearisierte Systembetrachtung oft ausreichend. In der Tat ist es meist das Ziel der Regelung, das System genau in einem stationären Zustand zu halten, so dass die Annahme linearer Systemmodelle umso besser erfüllt ist, je besser die Regelung funktioniert. Eine in der Technik extrem wichtige Beispiellasse für mechanische Systeme, die aufgrund der kleinen Abweichungen als linear aufgefasst werden können, sind die in jeder mechanischen Konstruktion auftretenden Schwingungen und Resonanzen, sei es in schwingenden Brücken oder Gebäuden, in vibrierenden Flugzeugen, Autos, oder Transformatoren, oder in oszillierenden Piezoelementen. In fast allen Fällen ist die Auslenkung so klein, dass eine lineare Systemanalyse das System extrem genau beschreiben kann. Viele andere technische Systeme sind von Beginn an beinahe linear, wie z.B. die Hauptelemente elektrischer Schaltkreise - Widerstand, Kondensator und Spule - oder die Wärmeleitung in festen Körpern.

2.6 Elektrische Schaltkreise

Beim Modellieren elektrischer Schaltkreise kann man neben den Kirchhoffschen Gesetzen von einfachen Grundgleichungen für die drei Elemente *Widerstand*, *Kapazität* und *Induktivität* ausgehen. Für jedes dieser Elemente kann man jeweils einen Zusammenhang zwischen dem durch das Element fließenden Strom $i(t)$ und der am Element anliegenden Spannung $v(t)$ angeben. Hierbei kann man sich willkürlich eine Eingangs- und eine Ausgangsseite definieren, wobei der Strom als positiv definiert wird, wenn er vom Eingang zum Ausgang fließt, und die Spannungsdifferenz als positiv definiert wird, wenn die Spannung vor dem Element höher ist als hinter dem Element, also $v(t) = v_{\text{ein}}(t) - v_{\text{aus}}(t)$ mit Eingangsspannung $v_{\text{ein}}(t)$ und Ausgangsspannung $v_{\text{aus}}(t)$. Mit diesen Konventionen gelten nun die folgenden drei elementaren Gleichungen.

Zum ersten die **Widerstandsgleichung** für einen Widerstand R , der in Ohm $[\Omega]$ gemessen wird:

$$v(t) = Ri(t). \quad (2.25)$$

Zum zweiten die **Kapazitätsgleichung** für einen Kondensator (oder ein ähnliches Element) mit Kapazität C , die in Farad [F] gemessen wird:

$$\frac{dv}{dt}(t) = i(t)/C. \quad (2.26)$$

Zum dritten die **Induktionsgleichung** für eine Spule (oder ein ähnliches Element) mit Induktivität L , die in Henry [H] gemessen wird:

$$\frac{di}{dt}(t) = v(t)/L. \quad (2.27)$$

Da die Kondensatorgleichung und Induktionsgleichung jeweils eine Zeitableitung enthalten, kann man vermuten, dass jedes dieser beiden Elemente einen Zustand enthält bzw. in einen elektrischen Schaltkreis miteinbringt. Dies ist tatsächlich meist der Fall, wenn nicht einige dieser "Zustände" sich direkt aus anderen Gleichungen ergeben. Wir geben zwei Beispiele.

2.6.1 RC-Kreis

Einen Stromkreis, wie in Abbildung 2.5, mit Eingangsspannung $v(t)$, wobei der Strom $i(t)$ durch einen in Serie geschalteten Widerstand R und einen Kondensator mit Kapazität C fließt. Die Spannung $v_C(t)$ über den Kondensator folgt der Gleichung

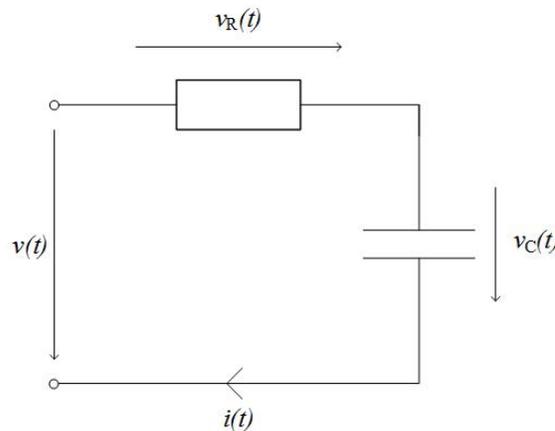


Abbildung 2.5: RC-Glied.

$$\dot{v}_C(t) = i(t)/C \quad (2.28)$$

und die Spannung über den Widerstand ist durch

$$v_R(t) = Ri(t) \quad (2.29)$$

beschrieben. Der Strom durch beide Elemente ist der gleiche, $i(t)$, während sich die Spannungen aufaddieren zu

$$v(t) = v_R(t) + v_C(t). \quad (2.30)$$

Da im Kondensator eine Zeitableitung der Spannung auftritt, können wir $v_C(t)$ als einen Zustand definieren. Wenn $v(t)$ der vorgegebene Input ist, dann ergibt sich aus der letzten Gleichung (2.30) $v_R(t) = v(t) - v_C(t)$ und mit der Widerstandsgleichung (2.29) $i(t) = (v(t) - v_C(t))/R$, und mit der Kondensatorgleichung (2.28) schließlich die Differentialgleichung in Standardform

$$\dot{v}_C(t) = \frac{1}{RC}(v(t) - v_C(t)). \quad (2.31)$$

Dieses System hat nur einen Zustand, $v_C(t)$, und sein Verhalten bei konstantem Eingangssignal $v(t) = v_0$ und bei Anfangswert $v_C(0) = 0V$ kann man analytisch angeben. Es ist gegeben durch

$$v_C(t) = v_0 \left(1 - \exp\left(-\frac{t}{RC}\right) \right), \quad (2.32)$$

wie man leicht durch Ableiten dieses Ausdrucks,

$$\dot{v}_C(t) = \frac{v_0}{RC} \exp\left(-\frac{t}{RC}\right), \quad (2.33)$$

und Einsetzen in (2.31) nachprüft.

2.6.2 Elektrischer Oszillator

Wenn wir zu dem vorigen System noch eine in Reihe geschaltete Induktivität L hinzufügen, wird auch aus dem Strom $i(t)$ ein Zustand, und zu den beiden bisherigen Gleichungen

$$\dot{v}_C(t) = i(t)/C \quad (2.34)$$

und

$$v_R(t) = Ri(t) \quad (2.35)$$

kommt noch die Induktivitätsgleichung

$$\dot{i}(t) = v_L(t)/L \quad (2.36)$$

hinzu. Zudem ist die Gleichung, die den Gesamtspannungsabfall beschreibt, um ein weiteres Glied erweitert:

$$v(t) = v_R(t) + v_C(t) + v_L(t). \quad (2.37)$$

Zu der Kondensatorspannung $v_C(t)$ kommt jetzt noch, wie in Abbildung 2.6 zu sehen, ein weiterer Zustand, der Strom $i(t)$ durch die Induktivität. Wenn diese zwei Größen sowie die Eingangsspannung $v(t)$ gegeben sind, dann erhält man aus (2.35) und (2.37)

$$v_L(t) = v(t) - Ri(t) - v_C(t) \quad (2.38)$$

und mit (2.36) die Differentialgleichung für den Strom durch die Spule:

$$\dot{i}(t) = \frac{1}{L} (v(t) - Ri(t) - v_C(t)) \quad (2.39)$$

Die andere Differentialgleichung beschreibt die Spannung über den Kondensator und ist gegeben durch (2.34), also

$$\dot{v}_C(t) = i(t)/C. \quad (2.40)$$

Definiert man den Zustandsvektor $x(t)$ und die Steuerung $u(t)$ als

$$x(t) = \begin{bmatrix} v_C(t) \\ i(t) \end{bmatrix} \quad \text{und} \quad u(t) = v(t),$$

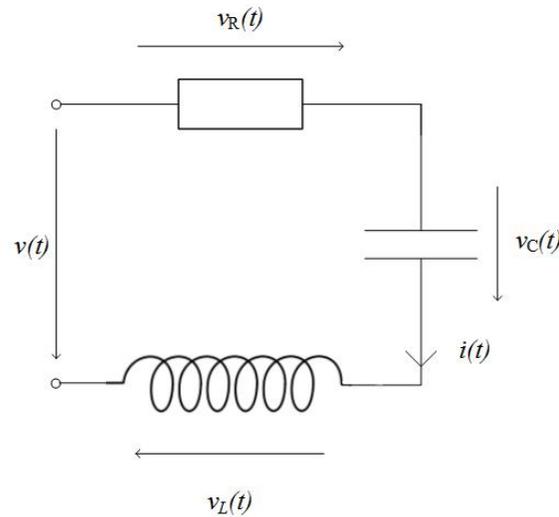


Abbildung 2.6: RLC-Glied.

dann kann das System wieder durch eine gewöhnliche Differentialgleichung $\dot{x} = f(x, u)$ beschrieben werden, mit

$$f(x, u) = \left[\begin{array}{c} \frac{1}{C} i(t) \\ \frac{1}{L} (v(t) - Ri(t) - v_C(t)) \end{array} \right].$$

Dies ist eine lineare Funktion der Form $Ax + Bu$, mit

$$A = \begin{bmatrix} 0 & \frac{1}{C} \\ -\frac{1}{L} & -\frac{R}{L} \end{bmatrix} \quad \text{und} \quad B = \begin{bmatrix} 0 \\ \frac{1}{L} \end{bmatrix}.$$

Wir werden im nächsten Kapitel sehen, dass dieses System als ein gedämpfter harmonischer Oszillator aufgefasst werden kann.

2.7 Zeitinvarianz und Linearität

Wir wollen kurz noch die zwei wichtigen Systemeigenschaften diskutieren, die wir bereits angesprochen haben: *Zeitinvarianz* und *Linearität*.

2.7.1 Zeitinvarianz

Wenn die Systemgleichungen nicht von der Zeit abhängen, wie wir das gleich zu Beginn dieses Kapitels angenommen haben, dann sprechen wir von einem *zeitinvarianten System*, sonst von einem *zeitvarianten* oder *zeitabhängigen System*.

Ein zeitinvariantes System hat die gleichen Zustandstrajektorien, wenn man es zu verschiedenen Zeitpunkten mit den gleichen Anfangsbedingungen und den gleichen Steuerungen startet. Mathematisch formulieren wir das so: falls $x_A(t)$ eine Trajektorie zu den Steuerungen $u_A(t)$ auf dem Intervall $t \in [0, T]$ und dem Anfangswert $x_A(0) = x_0$ ist, dann ist für jeden "Zeitshift" $\delta \in \mathbb{R}$ auch die "geschiftete" Trajektorie $x_B(t) := x_A(t - \delta)$ auf dem Intervall $t \in [\delta, T + \delta]$ eine Lösung zu der geschifteten Steuerung $u_B(t) := u_A(t - \delta)$ und dem Anfangswert $x_B(\delta) = x_0$. Fast alle technischen Systemmodelle, die wir betrachten, sind zeitinvariant. Wenn wir zum Beispiel Wasser kochen, dann braucht das Wasser heute ebenso lang bis es siedet wie morgen oder übermorgen. Um dies zu illustrieren wurde die Differentialgleichung $\ddot{x} = -x^3 - 0.2\dot{x} + u - x$ simuliert, siehe Abbildung 2.7.

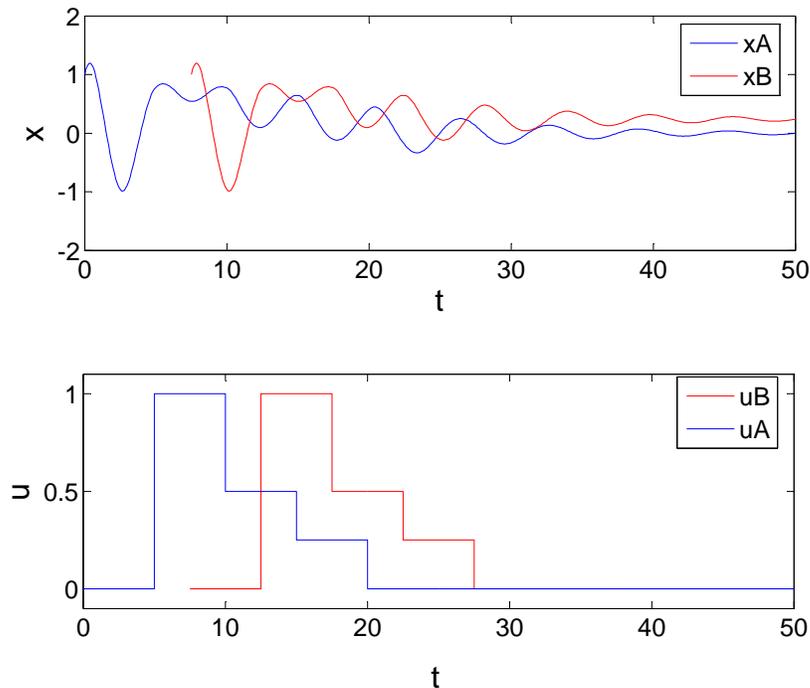


Abbildung 2.7: Zeitinvarianz: zeitlich verschobene Steuerungen und Anfangswerte führen bei einem zeitinvarianten System zu zeitlich verschobenen Zustandstrajektorien.

Auf der anderen Seite gibt es auch einige zeitabhängige Systeme in unserem täglichen Leben, z.B. solche, die durch den Tageslichteinfall beeinflusst sind, wie der menschliche Schlafrhythmus oder die Temperatur eines Gartenteiches. Generell kann man ein zeitabhängiges System durch eine zeitabhängige Differentialgleichung be-

schreiben:

$$\dot{x}(t) = f(t, x(t), u(t)).$$

und ein zeitinvariantes System durch eine zeitunabhängige Differentialgleichung, so wie bisher alle in diesem Kapitel betrachteten Systeme:

$$\dot{x}(t) = f(x(t), u(t)).$$

Wir werden uns in dieser Vorlesung auch weiterhin ausnahmslos mit zeitinvarianten System beschäftigen.

2.7.2 Linearität

Ein gesteuertes System ist genau dann ein lineares System, wenn der Systemzustand $x(t)$ auf eine lineare Weise von dem Anfangswert $x(0)$ und der Steuerungstrajektorie $u(t)$ abhängt. Wenn wir also zu einem Anfangszustand $x_A(0) \in \mathbb{R}^n$ und einer Steuerungstrajektorie $u_A : [0, T] \rightarrow \mathbb{R}^m$ die Zustandstrajektorie $x_A : [0, T] \rightarrow \mathbb{R}^n$ bekommen, und wir den Anfangswert und die Steuerung mit einem Faktor α multiplizieren, dann wird die daraus resultierende Zustandstrajektorie auch einfach die ursprüngliche sein, mit dem gleichen Faktor α multipliziert. Wenn wir zudem zu einem anderen Anfangswert $x_B(0)$ und einer anderen Steuerung $u_B : [0, T] \rightarrow \mathbb{R}^m$ die Zustandstrajektorie $x_B : [0, T] \rightarrow \mathbb{R}^n$ erhalten, dann würden wir für jede Linearkombination der Anfangswerte und Inputs auch eine Linearkombination der Zustandstrajektorie bekommen. Wenn wir die Linearkombination mit $x_C(0)$ und $u_C(t)$ benennen, mit Faktoren $\alpha, \beta \in \mathbb{R}$ so dass

$$x_C(0) = \alpha x_A(0) + \beta x_B(0)$$

und

$$u_C(t) = \alpha u_A(t) + \beta u_B(t) \quad \text{für alle } t \in [0, T],$$

dann ist die resultierende Zustandstrajektorie gegeben durch die Linearkombination:

$$x_C(t) = \alpha x_A(t) + \beta x_B(t) \quad \text{für alle } t \in [0, T].$$

Dies ist in Abbildung 2.8 dargestellt.

Linearität kann man für alle dynamischen Systeme der folgenden Form leicht zeigen:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad t \in [0, T], \quad x(0) = x_0. \quad (2.41)$$

Man beachte, dass wir hier im Gegensatz zu unser LTI-Standardform (2.41) sogar erlauben können, dass die Matrizen selbst von der Zeit t abhängen, wir also ein zeitabhängiges System haben; die Linearität wird dadurch nicht zerstört. Der Beweis ist so einfach, dass wir ihn als Übungsaufgabe lassen. Linear-Zeitvariante Systeme kommen in der Technik vor allem als Linearisierung nichtlinearer Systeme entlang stark variierender Trajektorien vor, z.B. bei der Regelung von Roboterarmbewegungen oder bei periodischen Prozessen wie den Umdrehungen eines Dieselmotors. Der Fokus dieser Vorlesung sind jedoch, wie bereits gesagt, lineare zeitinvariante Systeme, die auf

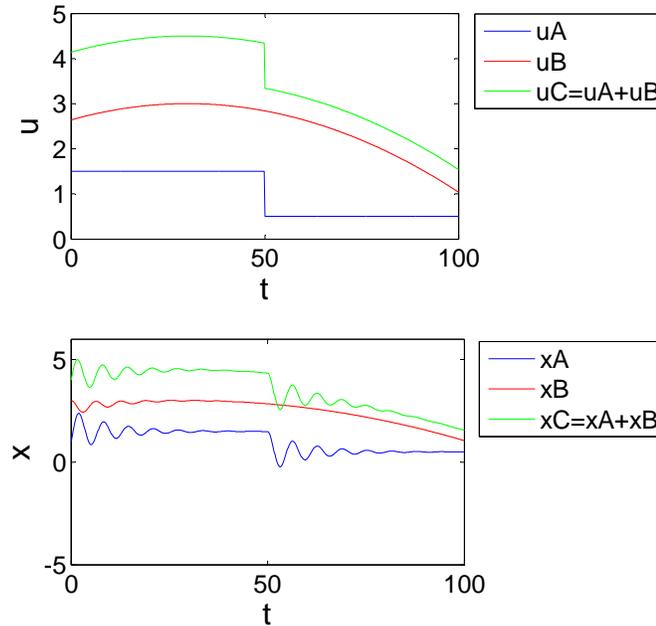


Abbildung 2.8: Linearität: die Systemantwort auf die Summe zweier Eingangstrajektorien ist bei einem linearen System die Summe der Zustandstrajektorien.

Englisch *linear time invariant* heißen und die wir deshalb kurz *LTI-Systeme* nennen, und die durch die lineare zeitinvariante Differentialgleichung

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (2.42)$$

mit zeitlich konstanten Matrizen A und B beschrieben werden.

2.8 Zeitdiskrete Systeme und Simulation

Viele technische Systeme, insbesondere in der Computer- und Informationstechnik, sind zeitlich getaktet, und anstatt in der kontinuierlich fließenden Zeit verändern sie sich nur zu gegebenen Zeitpunkten auf einem regelmäßigen Zeitraster. Wenn der Zeitschritt zwischen zwei Abtastzeitpunkten durch $\Delta t \in \mathbb{R}$ gegeben ist, dann kann man ein gesteuertes System in diskreter Zeit beschreiben durch eine Abbildung $\Phi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ die den Zustand zum Zeitpunkt $t = k\Delta t$, den wir mit $x_k \in \mathbb{R}^n$ benennen, und die diskrete Steuerung $u_k \in \mathbb{R}^m$, auf den neuen Zustand x_{k+1} abbilden:

$$x_{k+1} = \Phi(x_k, u_k), \quad k = 0, 1, 2, \dots, N$$

wobei $N \in \mathbb{N}$ die Länge des diskreten Zeithorizonts ist. Durch diese rekursive Abbildung wird für ein gegebenes x_0 und gegebene Steuerungen u_0, u_1, \dots, u_{N-1} die

gesamte Zustandstrajektorie x_0, x_1, \dots, x_N eindeutig bestimmt (siehe Abb. 2.9).

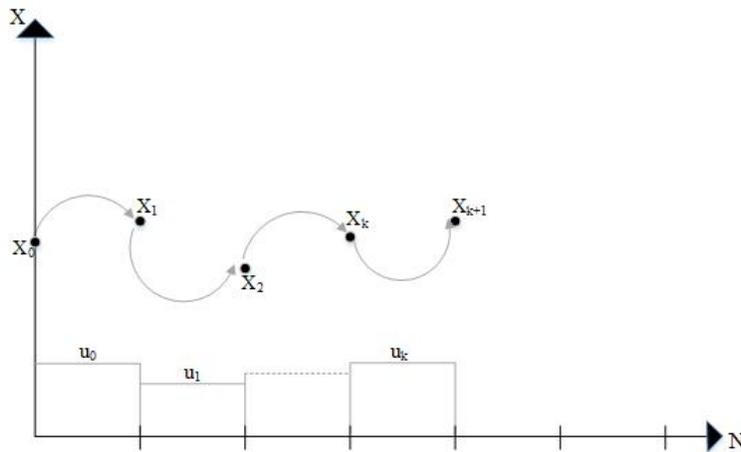


Abbildung 2.9: Darstellung eines zeitdiskreten Systems.

Wie gesagt, liegt der Fokus dieser Vorlesung nicht auf zeitdiskreten Systemen, sondern auf Systemen in kontinuierlicher Zeit, die beschrieben sind durch

$$\dot{x} = f(x, u).$$

Ein wichtiger Anwendungsfall diskreter Zeitsysteme, den wir hier erwähnen wollen, ist jedoch die numerische Simulation eines solchen kontinuierlichen Systems mit Hilfe des Computers: dafür müssen wir notwendigerweise einen meist kleinen Zeitschritt Δt einführen, um dann den kontinuierlichen Systemzustand $x(t)$ zu den Zeitpunkten $t_k = k\Delta t$ durch ein Zeitschrittverfahren zu approximieren. Das heißt, dass wir eine Abbildung Φ einführen, die eine Approximation der Zeit-Evolution auf dem kleinen Intervall ist, die also so gewählt ist, dass für die wirkliche kontinuierliche Trajektorie $x(t)$ gilt, dass

$$x(t + \Delta t) \approx \Phi(x(t), u(t)).$$

Hierbei wird u oft als konstant auf dem Zeitintervall angenommen. Eine einfache, aber in der Praxis nicht empfehlenswerte Abbildung ist der sogenannte *Euler-Integrator*: hier wird die Abbildung Φ definiert durch

$$\Phi_{\text{Euler}}(x, u) = x + \Delta t f(x, u).$$

Man kann sich diese Formel einfach herleiten, wenn man die resultierende Evolutionsgleichung

$$x_{x+1} = \Phi_{\text{Euler}}(x_k, u_k) = x_k + \Delta t f(x_k, u_k)$$

umordnet zu

$$\frac{x_{k+1} - x_k}{\Delta t} = f(x_k, u_k),$$

was eine Differenzenapproximation der Differentialgleichung $\dot{x} = f(x, u)$ ist. Leider muss man in der Praxis mit dem Euler-Integrator meist sehr kleine Schritte machen, um eine ausreichend genaue Simulation zu bekommen (siehe Abb. 2.10).

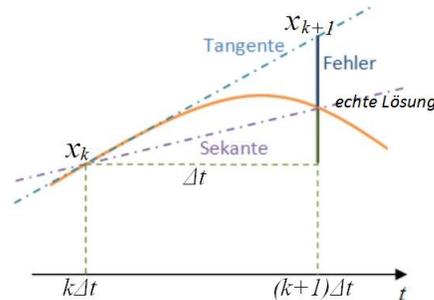


Abbildung 2.10: Visualisierung des auftretenden Fehlers beim Euler-Integrator.

Ein deutlich effizienteres Verfahren geht auf Runge und Kutta zurück, und nutzt die mehrfache Differenzierbarkeit der Funktion f aus, die in der Praxis fast immer gegeben ist. Die Herleitung ist kompliziert, aber die Implementation des meistverwendeten Runge-Kutta Integrators der Ordnung 4, den wir mit RK4 abkürzen, extrem einfach: die Funktion $\Phi_{\text{RK4}}(x, u)$ ist hier einfach durch folgende fünf Anweisungen gegeben:

$$\begin{aligned} k_1 &= f(x, u) \\ k_2 &= f\left(x + \frac{\Delta t}{2} k_1, u\right) \\ k_3 &= f\left(x + \frac{\Delta t}{2} k_2, u\right) \\ k_4 &= f\left(x + \Delta t k_3, u\right) \\ \Phi_{\text{RK4}}(x, u) &= x + \frac{\Delta t}{6} (k_1 + 2k_2 + 2k_3 + k_4). \end{aligned}$$

Mit Hilfe dieser Funktion in einer Programmschleife der Form

$$x_{k+1} = \Phi_{\text{RK4}}(x_k, u_k), \quad k = 0, 1, 2, \dots, N$$

können Sie für gegebenen Anfangswert x_0 leicht eine gewöhnliche Differentialgleichung mit treppenförmig variierenden Inputs auf dem Intervall $[0, N\Delta t]$ simulieren, also für eine Serie von diskreten Steuerungsinputs u_0, u_1, \dots, u_{N-1} , die auf jedem der kleinen Zeitintervalle als konstant angenommen werden.

2.9 Analytische Lösung für LTI Systeme

Im Spezialfall linear zeitinvarianter Systeme kennen wir aus der Vorlesung “Differentialgleichungen” bereits eine exakte Darstellung der Lösung des Simulationsproblems,

die wir hier kurz wiederholen möchten: für gegebenen Anfangswert x_0 und gegebene Steuerungstrajektorie $u(t)$, $t \in [0, T]$, hat das LTI System

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

für $t \in [0, T]$ die analytische Lösung

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau. \quad (2.43)$$

Man beachte, dass $x(t)$ und x_0 Vektoren im \mathbb{R}^n sind, und e^{At} und $e^{A(t-\tau)}$ Matrizen im $\mathbb{R}^{n \times n}$, und dass $B \in \mathbb{R}^{n \times m}$ und $u(\tau) \in \mathbb{R}^m$.

Der Beweis der Gültigkeit dieser Formel ist einfach, wenn man die Eigenschaften der Matrixexponentialfunktion nutzt: zum ersten gilt für $t = 0$ dass $e^{A0} = e^{A0}$ die Einheitsmatrix ist, so dass gilt

$$x(0) = Ix_0 + \int_0^0 \dots d\tau = x_0.$$

Zum zweiten gilt für die Ableitung des Ausdrucks (2.43)

$$\begin{aligned} \dot{x}(t) &= Ae^{At}x_0 + \int_0^t Ae^{A(t-\tau)}Bu(\tau)d\tau + e^{A(t-t)}Bu(t) \\ &= A \left(e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \right) + Bu(t) \\ &= Ax(t) + Bu(t), \end{aligned}$$

d.h. dass $x(t)$ die Differentialgleichung erfüllt.

Beispiel-System mit einem Zustand

Wir wollen die die folgende skalare ODE lösen:

$$\dot{x}(t) = ax(t) + bu(t)$$

wobei $a \in \mathbb{R}$. Als Anfangswert setzen wir $x_0 = 0$ und als Input die konstante Steuerung $u(t) = 1$ für alle Zeitpunkte $t \in [0, \infty)$. Die Lösung ist gegeben durch

$$\begin{aligned} x(t) &= e^{at}x_0 + \int_0^t e^{a(t-\tau)}bu(\tau)d\tau \\ &= 0 + \int_0^t e^{a(t-\tau)}bd\tau \\ &= e^{at} \int_0^t e^{-a\tau}bd\tau \\ &= e^{at} \left[\frac{b}{(-a)} e^{-a\tau} \right]_0^t \\ &= e^{at} \frac{b}{(-a)} (e^{-at} - 1) \\ &= \frac{b}{(-a)} (1 - e^{at}). \end{aligned}$$

Ein System dieser Art wird uns noch häufig begegnen. Man beachte, dass der Term e^{at} für $a < 0$ mit wachsender Zeit t gegen Null konvergiert, das System also *stabil* ist, während er für $a > 0$ unbegrenzt wächst, das System also *instabil* ist. Das Verhalten dieses Systems ist in Abbildung 2.11 illustriert.

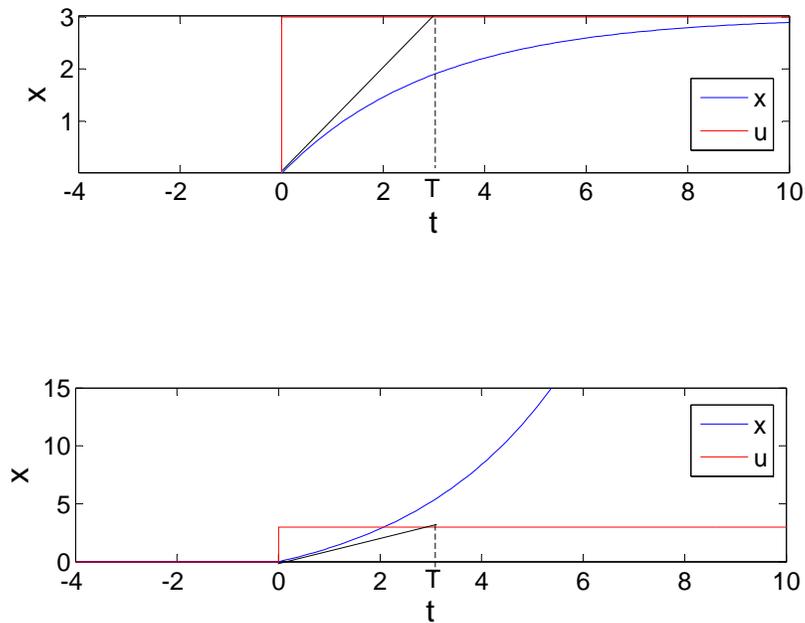


Abbildung 2.11: Simulation der Sprungantwort des Systems $\dot{x} = ax + bu$ mit der Anfangsbedingung $x(0) = 0$ und der Steuerung $u(t) = 3$. Oben: für $a = -\frac{1}{T}$ (und $b = \frac{1}{T}$ mit $T = 3$) ist das System stabil. Unten: für $a = \frac{1}{T}$ (und $b = \frac{1}{T}$) ist das System instabil. Man beachte, dass die Zeitkonstante T eine anschauliche Bedeutung hat.

Kapitel 3

Eingangs-Ausgangsdarstellung

Bei vielen dynamischen Systemen kann man den Zustand nicht direkt messen, oder man ist gar nicht an den Werten der Zustände interessiert. Stattdessen möchte man evtl. nur einige Größen $y(t) \in \mathbb{R}^p$ messen, wobei $p \in \mathbb{N}$ die Dimension des Ausgangs ist. Als Beispiele kann man sich vorstellen, dass man beim elektrischen Schwingkreis aus dem letzten Abschnitt nur an einer einzigen Ausgangsspannung interessiert ist, z.B. der Spannung über dem Widerstand und dem Kondensator, die gegeben ist als eine Funktion des Systemzustands $x(t)$, durch die Formel $y(t) = R i(t) + v_C(t)$. Ebenso könnte man beim Drehpendel nur an der Auslenkung $\phi(t)$ interessiert sein, die einer der beiden Zustände des Systems ist.

Ganz allgemein kann man ein gesteuertes dynamisches System mit Ausgängen darstellen als

$$\dot{x}(t) = f(x(t), u(t)), \quad (3.1)$$

$$y(t) = g(x(t), u(t)), \quad (3.2)$$

mit der "Ausgangsfunktion" $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^p$ (siehe Abbildung 3.1).

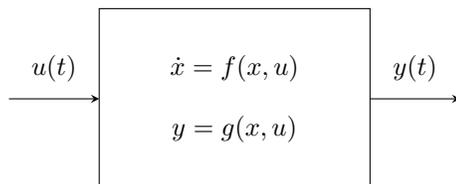


Abbildung 3.1: Blockdiagramm eines nichtlinearen zeitinvarianten Systems mit Eingang $u(t)$ und Ausgang $y(t)$.

Auch für Eingangs-Ausgangssysteme kann man einen Unterschied zwischen linearen und nichtlinearen Systemen machen und gegebenenfalls ein nichtlineares Modell durch Linearisierung in ein verwandtes lineares System umwandeln. Ein allgemeines

lineares zeitinvariantes Modell mit m Eingängen $u(t)$ und p Ausgängen $y(t)$ und n internen Zuständen $x(t)$ kann man beschreiben als

$$\dot{x}(t) = Ax(t) + Bu(t), \quad (3.3)$$

$$y(t) = Cx(t) + Du(t), \quad (3.4)$$

mit Matrizen $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$ und $D \in \mathbb{R}^{p \times m}$. Wir sprechen in diesem Fall auch von einem LTI-System.

3.0.1 MIMO und SISO Systeme

Falls es mehrere Eingänge und mehrere Ausgänge gibt, spricht man von einem MIMO-System nach dem Englischen *multiple-input-multiple-output*. Falls es jeweils nur einen Input und einen Output gibt, also $m = p = 1$, spricht man von einem SISO-System, nach *single-input-single-output*. Auch die Mischungen MISO und SIMO kommen vor. Die vier verschiedenen Arten von Eingangs-Ausgangs-Systemen sind in Abbildung 3.2 illustriert. In dieser Vorlesung werden wir uns hauptsächlich mit LTI-SISO-Systemen beschäftigen, also linearen Systemen mit einem Eingang und einem Ausgang. Sie haben ein überraschend vielseitiges Verhalten.

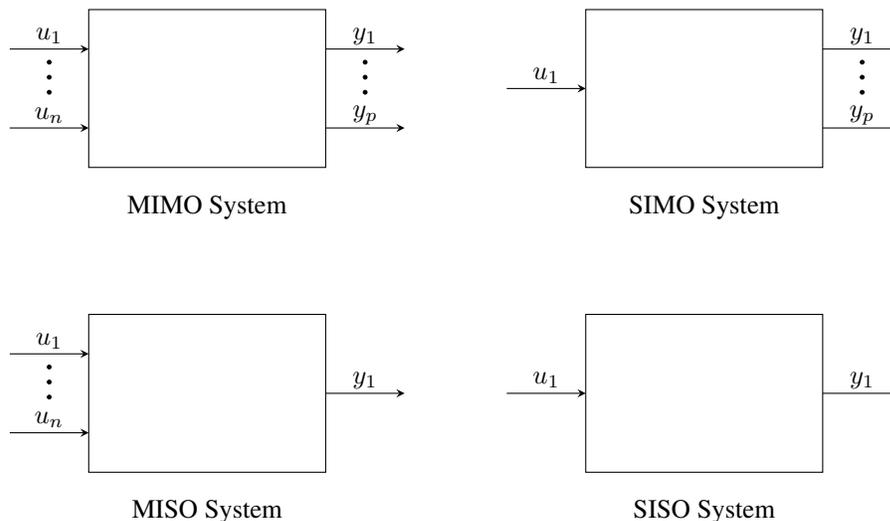


Abbildung 3.2: Blockschaltbilder von MIMO, SIMO, MISO und SISO Systemen.

3.1 Eingangs-Ausgangsdifferentialgleichung

Aus der Eingangs-Ausgangsperspektive ist der Systemzustand $x(t)$, der beim Modellieren eine so fundamentale Größe war, nichts anderes als ein mathematischer Zwischenschritt auf dem Weg von den Eingängen zu den Ausgängen. In vielen Fällen,

insbesondere für LTI Systeme, kann man für die Eingänge und Ausgänge direkt eine Differentialgleichung aufstellen, die dann allerdings oft höhere Zeitableitungen von $y(t)$ oder $u(t)$ enthält.

3.1.1 Eingangs-Ausgangsmodell für einen Oszillator

Als ein Beispiel nehmen wir den elektrischen Schwingkreis aus dem vorigen Kapitel, mit dem Ausgang $y(t) = Ri(t) + v_C(t)$. Leitet man diesen einmal nach der Zeit ab, erhält man

$$\dot{y}(t) = R\dot{i}(t) + \dot{v}_C(t),$$

was man durch Einsetzen von (2.36) und (2.40) zu

$$\dot{y}(t) = \frac{R}{L}(v(t) - Ri(t) - v_C(t)) + \frac{i(t)}{C}$$

umwandeln kann, bzw. durch Einsetzen der Definition von $u(t)$ und $y(t)$ zu

$$\dot{y}(t) = \frac{R}{L}(u(t) - y(t)) + \frac{i(t)}{C}.$$

Nochmaliges Ableiten liefert

$$\ddot{y}(t) = \frac{R}{L}(\dot{u}(t) - \dot{y}(t)) + \frac{\dot{i}(t)}{C}$$

was durch erneutes Einsetzen von (2.36) den Ausdruck

$$\ddot{y}(t) = \frac{R}{L}(\dot{u}(t) - \dot{y}(t)) + \frac{1}{LC}(v(t) - Ri(t) - v_C(t)) \quad (3.5)$$

$$\ddot{y}(t) = \frac{R}{L}(\dot{u}(t) - \dot{y}(t)) + \frac{1}{LC}(u(t) - y(t)) \quad (3.6)$$

liefert. Dies ist eine Differentialgleichung höherer Ordnung, die allein eine Relation zwischen $y(t)$ und $u(t)$ sowie ihren Ableitungen darstellt. Durch Sammeln aller y und u Terme auf der linken bzw. rechten Seite erhält man

$$\ddot{y}(t) + \frac{R}{L}\dot{y}(t) + \frac{1}{LC}y(t) = \frac{R}{L}\dot{u}(t) + \frac{1}{LC}u(t).$$

3.1.2 Allgemeine Eingangs-Ausgangsform

Man kann zeigen, dass man jedes LTI-SISO System mit n Zuständen durch eine lineare Differentialgleichung n -ter Ordnung darstellen kann, die die folgende Form hat:

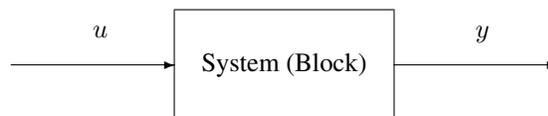
$$\frac{d^n y}{dt^n} + a_{n-1} \frac{d^{(n-1)} y}{dt^{n-1}} + \dots + a_1 \frac{dy}{dt} + a_0 y = b_n \frac{d^n u}{dt^n} + \dots + b_1 \frac{du}{dt} + b_0 u. \quad (3.7)$$

Man beachte, dass vor dem ersten Term kein Koeffizient a_n steht, da man die gesamte Gleichung durch den Wert a_n teilen könnte, wenn dieser von Null verschieden ist (und anderenfalls wäre die Gleichung von niedrigerer Ordnung d.h. n wäre kleiner). Man

beachte auch, dass b_n durchaus Null sein kann, und es häufig auch ist. Wenn $q \leq n$ der größte Index ist, für den b_q ungleich von Null ist, dann nennt man $n - q$ den relativen Grad eines Systems. Um die Bedeutung des relativen Grads zu verstehen, kann man die Antwort des Systems auf einen Input-Sprung betrachten. Man kann zeigen, dass ein Sprung in u zu einem Sprung in der $(n - q)$ -ten Zeitableitung von y führt, während alle kleineren Zeitableitungen von y stetig bleiben. Auf eine Weise “verzögert” das System die Sprungantwort also umso mehr, je höher der relative Grad ist. Der Fall $n = q$, also $b_n \neq 0$, tritt genau dann auf, wenn $D \neq 0$, wenn man also einen direkten “Durchgriff” des Inputs u auf den Output y hat. Der relative Grad des Systems ist dann Null.

3.2 Blockschaltbilder und Übertragungsglieder

Wenn man sich einmal entschieden hat, sich bei einem System nur noch für die Inputs und Outputs zu interessieren, dann kann man es durch eine “Black-Box” charakterisieren, die aus gegebenen Inputsignalen $u(t)$ die Outputsignale $y(t)$ generiert. Die Boxen, oder “Blöcke” können ganz beliebige z.B. auch nichtlineare Systeme repräsentieren.



Interessant wird es nun, wenn man mehrere Systeme zusammenschaltet, also die Outputs eines Systems als Inputs eines anderen nimmt. Auf diese Weise kann man aus mehreren einfachen Blöcken sehr komplexe Systemmodelle erstellen. Die einzigen Elemente sind:

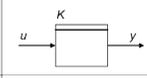
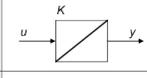
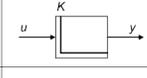
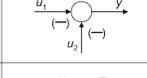
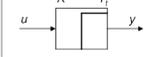
- Signale, die Outputs mit Inputs verbinden. Signale haben eine Wirkungsrichtung und wirken immer auf das System, auf das der Pfeil zeigt.
- Blöcke, die Einzelsysteme repräsentieren, die aus den jeweiligen Inputs die jeweiligen Outputs generieren. Wir nennen die Blöcke manchmal auch “Übertragungsglieder”.

Es ist wichtig, im Kopf zu behalten, dass die Signale immer Zeitverläufe darstellen, und die Blöcke aus einem Input-Zeitverlauf $u : \mathbb{R} \rightarrow \mathbb{R}$ einen Output Zeitverlauf $y : \mathbb{R} \rightarrow \mathbb{R}$ generieren. Signalpfeile können Abzweigungen haben, wenn das entsprechende Signal an verschiedenen Stellen benötigt wird. In Abbildung 3.3 zeigen wir eine Liste wichtiger Übertragungsglieder, die wir im folgenden Kapitel im Detail besprechen werden.

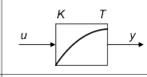
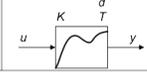
Man beachte, dass in manchen Blöcken mehrere Zustände “versteckt” sind, und in anderen keine. In der Blockdarstellung kann man große Systeme mit beliebig vielen

Häufig verwendete Übertragungsglieder

I. Elementare, lineare Übertragungsglieder

a)	Proportionalglied (P-Glied)	$y(t) = K \cdot u(t)$	
b)	Integrierglied (I-Glied)	$y(t) = K \cdot \int_0^t u(\tau) d\tau$	
c)	Differenzierglied (D-Glied)	$y(t) = K \cdot \frac{d}{dt} u(t)$	
d)	Summierglied (S-Glied)	$y(t) = \pm u_1(t) \pm u_2(t)$	
e)	Totzeitglied (Tt-Glied)	$y(t) = K \cdot u(t - T_t)$	

II. Zusammengesetzte, lineare Übertragungsglieder

a)	Verzögerungsglied 1. Ordnung (PT1-Glied)	$T\dot{y}(t) + y(t) = K \cdot u(t)$	
b)	Verzögerungsglied 2. Ordnung (PT2-Glied)	$T^2\ddot{y}(t) + 2dT\dot{y}(t) + y(t) = K \cdot u(t)$	

III. Nichtlineare Übertragungsglieder

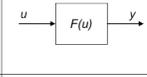
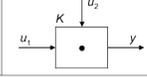
a)	Kennlinienglied (KL-Glied)	$y(t) = F(u(t))$	
b)	Multiplizierglied (M-Glied)	$y(t) = K \cdot u_1(t) \cdot u_2(t)$	

Abbildung 3.3: Wichtige Übertragungsglieder (aus: Ch. Ament, Skript Systemtheorie, Univ. Freiburg, 2007)

internen Zuständen als einen einzigen Block darstellen, der nur durch Eingänge und Ausgänge mit der Außenwelt kommuniziert. Als Normalfall werden SISO-Systeme betrachtet und jedes Signal, das durch einen einfachen Pfeil dargestellt wird, entspricht einer reellen Zahl. Werden MIMO-Systeme behandelt, dann nimmt man manchmal Pfeile mit doppelter Linie.

3.2.1 Proportional, Integrier, und Differenzierglied

Drei wichtige Glieder, die in der Regelungstechnik oft eingesetzt werden, sind die drei folgenden, die man gut mit den elektrischen Schaltelementen Widerstand, Induktivität, und Kondensator in Zusammenhang bringen kann:

- Das **Proportionalglied (P-Glied)** ist durch die Formel

$$y(t) = Ku(t)$$

beschrieben, und hat keinen internen Zustand. Man kann es z.B. mit einem elektrischen Widerstand, dessen Eingang die Spannung ist, und dessen Ausgang der Strom ist, realisieren.

- Das **Integrier-Glied (I-Glied)** ist durch die Formel

$$y(t) = \int_0^t u(t)dt + y_0$$

beziehungsweise durch die Differentialgleichung

$$\dot{y}(t) = u(t), \quad y(0) = y_0,$$

definiert, und es enthält einen internen Zustand, nämlich y , dessen Anfangswert y_0 man für eine korrekte Systemsimulation kennen muss. Oft nimmt man an, dass $y_0 = 0$. Man kann ein I-Glied z.B. durch eine Induktivität realisieren, deren Eingang die Spannung ist und deren Ausgang der Strom, oder durch einen Kondensator, dessen Eingang der Strom und dessen Ausgang die Spannung ist.

- Das **Differenzier-Glied (D-Glied)** ist durch die Formel

$$y(t) = \frac{du}{dt}(t)$$

definiert und kann nicht in Form einer Zustandsdifferentialgleichung geschrieben werden. Es kann somit in seiner reinen Form auch nicht physikalisch realisiert werden, sondern nur approximiert. Man kann sich das D-Glied als einen Kondensator vorstellen, dessen Eingang die Spannung ist, und dessen Ausgang der Strom, oder eine Induktivität, deren Eingang der Strom und Ausgang die Spannung ist. Fügt man diesen Schaltelementen einen Widerstand hinzu, was in der Realität immer der Fall wäre, kommt man direkt auf eine physikalische Realisierung des D-Gliedes.

3.2.2 Physikalische Realisierung des D-Gliedes als DT₁-Glied

Wir betrachten den RC-Kreis aus Abschnitt 2.6 mit der Eingangsspannung $u(t)$ als Eingang und der Kondensatorspannung $x(t)$ als Zustand. Der RC-Kreis war durch die Zustandsdifferentialgleichung

$$\dot{x}(t) = \frac{1}{RC}(u(t) - x(t)). \quad (3.8)$$

beschrieben. Nehmen wir jetzt aber, anders als in Abschnitt 2.6, den Strom durch den RC-Kreis als Ausgang y , also

$$y(t) = (u(t) - x(t))/R, \quad (3.9)$$

dann erhalten wir eine Approximation des D-Gliedes, die umso besser ist, je kleiner der Widerstand R ist. Um dies besser zu verstehen, leiten wir hier zunächst die Eingangs-Ausgangsdifferentialgleichung her: es ergibt sich durch Ableiten der Ausgangsgleichung (3.9)

$$\dot{y} = (\dot{u} - \dot{x})/R,$$

und durch Einsetzen der Zustandsgleichung (3.8)

$$\dot{y} = \frac{1}{R} \left(\dot{u} - \frac{1}{RC}(u - x) \right),$$

sowie durch Verwenden der Identität (3.9), also $x = u - Ry$, die Eingangs-Ausgangsform:

$$\dot{y} = \frac{1}{R} \left(\dot{u} - \frac{y}{C} \right).$$

Das Verhalten dieses Systems für konstantes $\dot{u}(t) = \dot{u}_{ss}$ kennen wir bereits: es konvergiert exponentiell mit einer Zeitkonstanten R gegen den Wert $y_{ss} = C\dot{u}_{ss}$ wobei z.B. bei einem Anfangswert $y(0) = 0$ der genaue Ausdruck

$$y(t) = y_{ss}(1 - e^{-t/R})$$

gilt. Wenn \dot{u} nicht konstant ist, sondern langsam variiert, aber die Zeitkonstante R klein ist im Vergleich zur Zeitkonstante der Variationen in \dot{u} , dann können wir approximativ annehmen, dass

$$y(t) = C\dot{u}(t),$$

was einer physikalischen Realisierung des D-Gliedes entspräche. Generell wird ein System mit der Differentialgleichung

$$\dot{y} = \frac{1}{T}(\dot{u} - y)$$

oder, äquivalent,

$$T\dot{y} + y = \dot{u}$$

“DT₁-Glieder” oder “verzögertes D-Glieder” genannt, wobei T die Zeitkonstante darstellt. Ein DT₁-Glieder hat einen internen Zustand.

3.2.3 Verzögerungsglied erster Ordnung oder PT₁-Glieder

Ein dem DT₁-Glieder verwandtes und sehr häufig vorkommendes Glied hatten wir bereits im RC-Kreis in Abschnitt 2.6 kennengelernt, dessen Eingang die Eingangsspannung war und dessen Ausgang die Spannung über dem Kondensator, so dass wir direkt Gleichung (3.8) verwenden können, mit $y(t) = x(t)$, also

$$\dot{y}(t) = \frac{1}{RC}(u(t) - y(t)). \quad (3.10)$$

Hier stellt sich bei konstantem Input $u(t) = u_{ss}$ der Gleichgewichtswert $y_{ss} = u_{ss}$ ein, mit exponentiell abklingendem Fehler, wobei die Zeitkonstante des Abklingens durch $T = RC$ gegeben ist (siehe Abbildung 3.4).

Ganz allgemein nennt man ein Glied, das durch die Differentialgleichung

$$T\dot{y} + y = Ku$$

beschrieben wird, ein “PT₁-Glieder” oder “einfach verzögertes Proportionalglied”. Ein PT₁-Glieder hat einen internen Zustand.

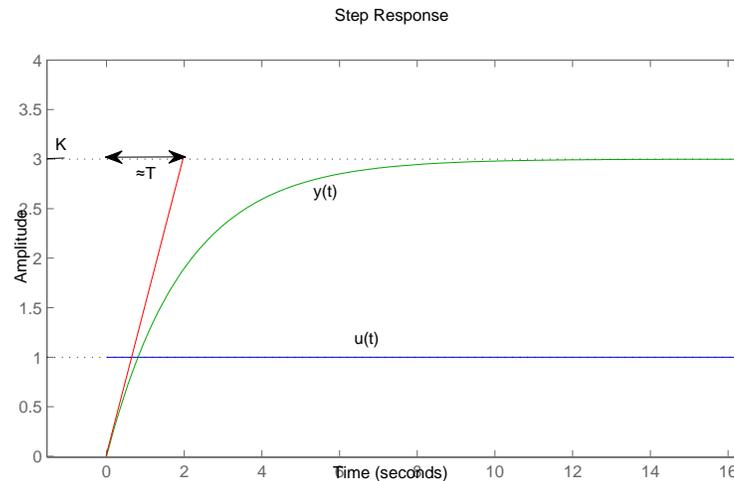


Abbildung 3.4: Sprungantwort eines PT₁-Gliederes mit Eingang $u(t) = 1$ und Ausgang $y(t)$. Verstärkung $K = 3$ und Zeitkonstante $T = 2$.

3.2.4 Verzögerungsglied zweiter Ordnung oder PT₂-Glieder

Mitglieder einer sehr interessanten weiteren Systemklasse haben wir bereits als “harmonischen Oszillator” oder “elektrischen Schwingkreis” kennengelernt. In Abschnitt 3.1 haben wir bereits die Eingangs-Ausgangsdifferentialgleichung für einen elektrischen Schwingkreis hergeleitet, wenn man die Gesamtspannung als Eingang u und der Spannung über Widerstand und Kondensator als Ausgang y betrachtet, die sich ergab als:

$$\ddot{y} + \frac{R}{L}\dot{y} + \frac{1}{LC}y = \frac{R}{L}\dot{u} + \frac{1}{LC}u.$$

Hätten wir als Ausgang y nur die Spannung über dem Kondensator betrachtet, wäre die folgende, der obigen recht ähnliche Gleichung herausgekommen:

$$\ddot{y} + \frac{R}{L}\dot{y} + \frac{1}{LC}y = \frac{1}{LC}u. \quad (3.11)$$

Dies ist ein Beispiel für das sogenannte PT₂-Glieder oder “Verzögerungsglied zweiter Ordnung”, das das erste Glied in diesem Abschnitt ist, das schwingungsfähig ist. Es

kann ganz allgemein in der Form

$$T^2 \ddot{y} + 2dT\dot{y} + y = Ku$$

geschrieben werden. Man nennt T wieder die Zeitkonstante, und d die Dämpfung. Für den elektrischen Schwingkreis erhält man z.B. durch Multiplikation von (3.11) mit LC die äquivalente Gleichung

$$LC\ddot{y} + RC\dot{y} + y = u$$

und man hat es in die PT₂-Standardform gebracht mit $T = \sqrt{LC}$ und $d = \frac{R\sqrt{C}}{2\sqrt{L}}$ und $K = 1$.

Falls T und d positiv sind, ist das PT₂-Glied stabil, und konvergiert für konstanten Input $u(t) = u_{ss}$ gegen den Wert $y_{ss} = Ku_{ss}$, deshalb heißt es manchmal auch "zweifach verzögertes Proportionalglied". Allerdings oszilliert es beim Übergang von einem Gleichgewichtszustand zu einem anderen, wenn die Dämpfung d kleiner als eins ist. Beim elektrischen Schwingkreis ist dies z.B. genau dann der Fall, wenn $R^2 < 4\frac{L}{C}$. Wir werden dieses Übertragungsglied noch oft und sehr viel genauer betrachten. Es hat zwei interne Zustände.

3.3 Äquivalenz von Zustands- und E/A-Darstellung

Der Weg von einem Zustandsmodell zu einer äquivalenten Eingangs-Ausgangsdifferentialgleichung ist nicht trivial, aber immer möglich, wenn zwei technische Bedingungen erfüllt sind, die wir hier nicht diskutieren wollen (Beobachtbarkeit und Steuerbarkeit). Die Koeffizienten in der Eingangs-Ausgangs-Differentialgleichung (3.7), die wir hier der Übersicht halber noch einmal wiederholen,

$$\frac{d^n y}{dt^n} + a_{n-1} \frac{d^{(n-1)} y}{dt^{n-1}} + \dots + a_1 \frac{dy}{dt} + a_0 y = b_n \frac{d^n u}{dt^n} + \dots + b_1 \frac{du}{dt} + b_0 u. \quad (3.12)$$

sind dann eindeutig.

Umgekehrt kann man auch jedes Modell in der Form (3.12) in ein äquivalentes LTI-System in Zustandsform (3.3)-(3.4) umformulieren. Aber während die Input-Output Differentialgleichung für ein gegebenes System eindeutig bestimmte Koeffizienten hat, ist umgekehrt die Wahl der Zustandsform-Matrizen für eine gegebene Input-Output Differentialgleichung nicht eindeutig. Wir diskutieren die Möglichkeiten der Realisierung in Zustandsform in den folgenden Abschnitten.

3.3.1 Regelungsnormalform

Eine Möglichkeit ist, die Zustandsform bei bekannten Koeffizienten a_0, \dots, a_{n-1} und b_0, \dots, b_n in (3.7) durch die folgenden Matrizen darzustellen:

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{n-2} & -a_{n-1} \end{bmatrix} \quad \text{und} \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (3.13)$$

sowie

$$C = [b_0 - b_n a_0 \mid b_1 - b_n a_1 \mid \cdots \mid b_{n-1} - b_n a_{n-1}] \quad \text{und} \quad D = [b_n]. \quad (3.14)$$

Die hier getroffene Wahl für die Matrizen A, B, C, D zu einem System (3.7) nennt man “Regelungsnormalform”.

3.3.2 Regelungsnormalform für Systeme vom relativen Grad n

Wir wollen die Regelungsnormalform zur Illustration für einen wichtigen Spezialfall selbst herleiten, nämlich für den Fall, dass $b_0 \neq 0$, aber $b_1 = 0, b_2 = 0, \dots, b_n = 0$, also für den Fall einer Differentialgleichung vom relativen Grad n , die gegeben ist durch:

$$\frac{d^n y}{dt^n} + a_{n-1} \frac{d^{(n-1)}y}{dt^{n-1}} + \dots + a_1 \frac{dy}{dt} + a_0 y = b_0 u. \quad (3.15)$$

Diese Differentialgleichung können wir umordnen zu

$$\frac{d^n y}{dt^n} = -a_{n-1} \frac{d^{(n-1)}y}{dt^{n-1}} - \dots - a_1 \frac{dy}{dt} - a_0 y + b_0 u.$$

Falls wir nun den Zustandsvektor $x \in \mathbb{R}^n$ wie folgt wählen,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y \\ \dot{y} \\ \vdots \\ \frac{d^{(n-1)}y}{dt^{n-1}} \end{bmatrix},$$

dann kann die obige Differentialgleichung auch geschrieben werden als

$$\dot{x}_n = -a_{n-1}x_n - \dots - a_1x_2 - a_0x_1 + b_0u.$$

Zusammen mit den trivialen Differentialgleichungen

$$\dot{x}_k = x_{k+1}, \quad \text{für} \quad k = 1, 2, \dots, n-1$$

ergibt sich die Differentialgleichung

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b_0 \end{bmatrix} u$$

Durch Dividieren der gesamten Gleichung durch den Faktor b_0 kann man das b_0 aus der letzten Matrix entfernen und erhält umskalierte Zustände $x_k^{\text{neu}} = x_k/b_0, k = 1, \dots, n$.

Man kann dann den Output bekommen als $y = b_0 x_1^{\text{neu}}$. Zusammengefasst ergibt sich in diesen neuen Zuständen also:

$$\dot{x}^{\text{neu}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} x^{\text{neu}} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u$$

und

$$y = [b_0, 0, \dots, 0] x^{\text{neu}}.$$

Dies ist ein Modell in der Regelungsnormalform (3.13)-(3.14).

3.3.3 Zustandstransformationen

Aus der Eingangs-Ausgangsperspektive ist nicht nur der Systemzustand zu einer bloßen Hilfsgröße geworden, sondern werden auch die Zustandsraummatrizen A, B, C uneindeutig. Man kann den Zustand und die Matrizen immer auch anders wählen, indem man eine lineare Koordinatentransformation mit einer invertierbaren Matrix $T \in \mathbb{R}^{n \times n}$ durchführt. Setzen wir in der Gleichung (3.3)-(3.4) für den Zustand $x(t)$ den Ausdruck $x(t) = Tz(t)$ ein, so erhalten wir

$$\begin{aligned} T\dot{z}(t) &= ATz(t) + Bu(t), \\ y(t) &= CTz(t) + Du(t), \end{aligned}$$

und durch multiplizieren der ersten Gleichung mit T^{-1} erhalten wir

$$\dot{z}(t) = \underbrace{T^{-1}AT}_{=:A_z} z(t) + \underbrace{T^{-1}B}_{=:B_z} u(t), \quad (3.16)$$

$$y(t) = \underbrace{CT}_{=:C_z} z(t) + Du(t). \quad (3.17)$$

Mit den neu definierten Matrizen können wir jetzt ein aus der Eingangs-Ausgangsperspektive 100% äquivalentes Zustandsraummodell formulieren:

$$\dot{z}(t) = A_z z(t) + B_z u(t), \quad (3.18)$$

$$y(t) = C_z z(t) + Du(t). \quad (3.19)$$

Wir können diese Nichteindeutigkeit der Zustandsdarstellung zum besseren Verständnis des Systems nutzen, indem wir z.B. die Matrix A auf Diagonalform bringen, oder in die Regelungsnormalform (3.13) (was immer geht, wenn das Zustandsraummodell die technische Bedingung der "Steuerbarkeit" erfüllt¹).

¹Ein System $\dot{x} = Ax + Bu$ mit $A \in \mathbb{R}^{n \times n}$ ist "steuerbar" wenn die Matrix $[B, AB, A^2B, \dots, A^{n-1}B]$ den Rang n hat.

Kapitel 4

Das charakteristische Polynom

Wir wissen aus der Matrixalgebra, dass man zu jeder quadratischen Matrix $A \in \mathbb{R}^{n \times n}$ das sogenannte "charakteristische Polynom" p_A definieren kann, das definiert ist durch den folgenden Ausdruck unter Verwendung der Determinante:

$$p_A(\lambda) = \det(\lambda I - A). \quad (4.1)$$

Hierbei ist $I \in \mathbb{R}^{n \times n}$ die Einheitsmatrix und $\lambda \in \mathbb{R}$ eine reelle Zahl, die das Argument des Polynoms darstellt. Ausgeschrieben ergibt sich also

$$p_A(\lambda) = \det \left(\begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{bmatrix} - \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \right).$$

Das charakteristische Polynom der Systemmatrix A eines LTI-Systems hat besondere Eigenschaften, die wir kurz diskutieren werden.

4.1 Invarianz unter Zustandstransformationen

Das charakteristische Polynom ist invariant unter Zustandstransformationen mit invertierbaren Matrizen $T \in \mathbb{R}^{n \times n}$. Dies ergibt sich aus dem Determinantenproduktsatz:

$$\det(T^{-1}(\lambda I - A)T) = \underbrace{\det(T^{-1})}_{=\det(T)^{-1}} \det(\lambda I - A) \det(T) = \det(\lambda I - A)$$

Im Gegensatz zur Matrix A ist das charakteristische Polynom damit eine invariante Größe, die für ein gegebenes Eingangs-Ausgangsverhalten eindeutig bestimmt ist.

Das charakteristische Polynom hat also tatsächlich genau die Koeffizienten, die in der Eingangs-Ausgangsdifferentialgleichung (3.7) vor den entsprechenden Zeitableitungen von y stehen.

4.3 Nullstellen des charakteristischen Polynoms, oder "Polstellen"

Die Nullstellen des charakteristischen Polynoms p_A sind genau die Eigenwerte der Matrix A . Dies ergibt sich durch Transformation der Matrix A auf Jordan-Normalform $J = T^{-1}AT$. Wir wollen diesen wichtigen Sachverhalt etwas genauer diskutieren.

Zuerst zum Polynom: nach dem Fundamentalsatz der Algebra kann jedes reelle Polynom, also auch $p_A(\lambda) = a_0 + a_1\lambda + \dots + a_{n-1}\lambda^{n-1} + \lambda^n$, als das Produkt von Linearfaktoren geschrieben werden, die zu seinen komplexen Nullstellen $\lambda_k \in \mathbb{C}$ gehören:

$$p_A(\lambda) = \prod_{k=1}^r (\lambda - \lambda_k)^{n_k}. \quad (4.3)$$

Dabei können einige Nullstellen mehrfach vorkommen, und wir benennen deshalb die $r \leq n$ paarweise verschiedenen Werte der Nullstellen mit $\lambda_k \in \mathbb{C}$, $k = 1, \dots, r$, und benennen die Vielfachheit jedes dieser Nullstellenwerte mit n_k . Die Gesamtzahl der mit ihrer Mehrfachheit gezählten Nullstellen ist n , es gilt also $\sum_{k=1}^r n_k = n$.

Jetzt zur Matrix A : Für jede Matrix $A \in \mathbb{R}^{n \times n}$ kann man eine invertierbare Matrix $T \in \mathbb{C}^{n \times n}$ finden, so dass $T^{-1}AT$ in sogenannte "Jordan-Blöcke" J_k zerfällt, die gegeben sind durch

$$J_k = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & 1 & \\ & & \ddots & \ddots \\ & & & \lambda_k \end{bmatrix},$$

wobei wir sehen werden, dass λ_k gerade eine der Nullstellen des charakteristischen Polynoms $p_A(\lambda)$ ist, und es deshalb bereits so benennen. Ist die Matrix A die Systemmatrix eines Systems $\dot{x} = Ax + Bu$, das die technische Bedingung der Steuerbarkeit erfüllt, dann gibt es zu jedem der Werte λ_k (mit $k = 1, \dots, r$) genau einen Jordanblock, der die Dimension der sogenannten algebraischen Vielfachheit n_k hat. Ausgeschrieben ergibt sich also

$$T^{-1}AT = J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{bmatrix}$$

mit $J_k \in \mathbb{C}^{n_k \times n_k}$. Somit gilt tatsächlich wegen

$$p_A(\lambda) = p_J(\lambda) = \det(\lambda I - J) = \prod_{k=1}^r \det(\lambda I - J_k)$$

und

$$\det(\lambda I - J_k) = \det \left(\begin{bmatrix} (\lambda - \lambda_k) & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & (\lambda - \lambda_k) \end{bmatrix} \right) = (\lambda - \lambda_k)^{n_k},$$

dass

$$p_A(\lambda) = \prod_{k=1}^r (\lambda - \lambda_k)^{n_k}.$$

Wir haben hier wegen der Annahme der Steuerbarkeit annehmen dürfen, dass es zu jedem Eigenwert λ_k der Matrix A nur einen einzigen Jordanblock gibt.

Wegen eines Zusammenhangs mit den Polstellen der sogenannten "Übertragungsfunktion", den wir in Kapitel 6 erläutern werden, werden die Nullstellen des charakteristischen Polynoms, die zugleich die Eigenwerte der Systemmatrix sind, auch oft "Polstellen der Übertragungsfunktion" genannt, oder kurz "Polstellen". Wir werden in diesem Skript ab jetzt sogar aktiv vermeiden, den (mathematisch völlig korrekten) Ausdruck "Nullstellen des charakteristischen Polynoms" für die Eigenwerte/Polstellen eines Systems zu gebrauchen, um Verwirrung zu vermeiden, da in der Systemtheorie eine andere Form von "Nullstellen" - nämlich die Nullstellen des Zählerpolynoms der Übertragungsfunktion - auch eine große Rolle spielen werden.

4.4 Stabilität bei negativen Realteilen der Polstellen

Eine wichtige Frage zu jedem dynamischen System ist, ob es stabil ist, d.h., ob es bei verschwindenden oder sehr kleinen Inputs $u(t)$ einen verschwindenden oder sehr kleinen Output $y(t)$ hat. Wir werden sehen, dass ein System genau dann stabil ist, wenn die Realteile aller Polstellen in der linken Halbebene der komplexen Zahlenebene liegen. Hierzu betrachten wir ein System mit Input $u(t) = 0$, dass entweder durch die Differentialgleichung

$$\frac{d^n y}{dt^n} + a_{n-1} \frac{d^{(n-1)} y}{dt^{n-1}} + \dots + a_1 \frac{dy}{dt} + a_0 y = 0$$

beschrieben werden kann oder durch die Zustandsform

$$\dot{x} = Ax \quad \text{und} \quad y = Cx$$

mit passenden Matrizen A und C . Wenn wir alle Polstellen des Systems kennen, also die r verschiedenen Lösungen λ_k der Gleichung

$$\lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0 = 0,$$

oder, was äquivalent ist, die Eigenwerte der Matrix A , dann können wir zwei sehr starke Aussagen über die Stabilität des Systems machen:

1. Das System ist stabil, wenn alle Polstellen negativen Realteil haben.

2. Das System ist instabil, wenn eine Polstelle einen positiven Realteil hat.

Stabil oder instabil bezieht sich hier auf das Abklingen oder Anwachsen eines anfänglich von Null verschiedenen, beliebigen Anfangszustands x_0 .

Wir zeigen dies mit Hilfe der Zustandsraumdarstellung, für die der folgende explizite Ausdruck für den Output $y(t)$ gilt:

$$y(t) = Ce^{At}x_0. \quad (4.4)$$

Durch Transformation von A auf Jordan-Form $A = TJT^{-1}$ erhalten wir wegen

$$e^{TJT^{-1}} = \sum_{k=0}^{\infty} \frac{(TJT^{-1})^k}{k!} = T \sum_{k=0}^{\infty} \frac{(J)^k}{k!} T^{-1}$$

daraus den äquivalenten Ausdruck

$$y(t) = CT e^{Jt} T^{-1} x_0 = \tilde{C} e^{Jt} \tilde{x}_0 \quad (4.5)$$

mit den Definitionen $\tilde{C} := CT$ und $\tilde{x}_0 := x_0$. Da die Matrix J blockdiagonal ist und in Jordanblöcke J_k zerfällt, und die Matrix-Exponentialfunktion auf alle Diagonalblöcke separat wirkt, gilt

$$e^{Jt} = \begin{bmatrix} e^{J_1 t} & & & \\ & e^{J_2 t} & & \\ & & \ddots & \\ & & & e^{J_r t} \end{bmatrix}$$

wobei $e^{J_k t} \in \mathbb{C}^{n_k \times n_k}$. Wir können also das Verhalten jedes Jordanblock-Subsystems separat behandeln, denn es gilt

$$\tilde{C} e^{Jt} \tilde{x}_0 = [\tilde{C}_1 | \tilde{C}_1 | \dots | \tilde{C}_r] \begin{bmatrix} e^{J_1 t} \tilde{x}_{0,1} \\ \vdots \\ e^{J_r t} \tilde{x}_{0,r} \end{bmatrix} = \sum_{k=1}^r \tilde{C}_k e^{J_k t} \tilde{x}_{0,k}$$

für die entsprechenden Vektoraufteilungen

$$\tilde{C} = [\tilde{C}_1 | \tilde{C}_1 | \dots | \tilde{C}_r] \quad \text{und} \quad \tilde{x}_0 = \begin{bmatrix} \tilde{x}_{0,1} \\ \vdots \\ \tilde{x}_{0,r} \end{bmatrix}.$$

Man beachte, dass \tilde{x}_0 ein stehender Vektor und \tilde{C} für SISO Systeme eine $(1 \times n)$ -Matrix ist, also ein liegender Vektor. Analysieren wir jetzt also das Verhalten so eines Subsystems $\tilde{C}_k e^{J_k t} \tilde{x}_{0,k}$ mit Jordanblock J_k für ein beliebiges $k \in \{1, \dots, r\}$. Hierfür ist es hilfreich, die Matrix J_k in zwei Teile zu zerlegen:

$$J_k = \lambda_k I + N_k \quad \text{mit} \quad N_k = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & \ddots & \ddots \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix},$$

wobei $N_k \in \mathbb{R}^{n_k \times n_k}$. Man berechnet leicht, dass

$$N_k^2 = \begin{bmatrix} 0 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \end{bmatrix}, \quad N_k^3 = \begin{bmatrix} 0 & 0 & 0 & 1 & \\ & \ddots & \ddots & \ddots & \ddots \\ & & & & & 1 \end{bmatrix}, \quad \text{etc.}$$

gilt, und dass $N_k^{n_k} = 0$. Damit ergibt sich der folgende Ausdruck für $e^{N_k t}$:

$$e^{N_k t} = \begin{bmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{(n_k-1)}}{(n_k-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & & & \frac{t^2}{2} \\ & & & & t \\ & & & & 1 \end{bmatrix}.$$

Da die Matrizen $\lambda_k I$ und J kommutieren, also ihre Reihenfolge in Matrixprodukten keine Rolle spielt, gilt zudem

$$e^{\lambda_k I t + N_k t} = e^{\lambda_k I t} e^{N_k t} = e^{\lambda_k t} \begin{bmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{(n_k-1)}}{(n_k-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & & & \frac{t^2}{2} \\ & & & & t \\ & & & & 1 \end{bmatrix}.$$

Man beachte, dass die Lösung des Anfangswertproblems damit nur aus Termen der Form $e^{\lambda_k t}$, $e^{\lambda_k t} t$, \dots , $e^{\lambda_k t} t^{(n_k-1)}$ besteht, also Produkten der Exponentialfunktion $e^{\lambda_k t}$ mit Polynomen vom Grad kleiner als n_k . Es gilt für den Ausdruck $\tilde{C}_k e^{J_k t} \tilde{x}_{0,k}$ also, dass

$$\tilde{C}_k e^{J_k t} \tilde{x}_{0,k} = (c_0 + c_1 t + c_2 t^2 + \dots + c_{n_k-1} t^{n_k-1}) e^{\lambda_k t},$$

wobei $c_0, \dots, c_{n_k-1} \in \mathbb{C}$ komplexe Koeffizienten sind. Ist nun der Realteil von λ_k negativ, dann konvergiert die Funktion $e^{\lambda_k t}$ schneller gegen Null als jedes Polynom, und es gilt

$$\lim_{t \rightarrow \infty} \tilde{C}_k e^{J_k t} \tilde{x}_{0,k} = 0.$$

Sind also die Realteile aller Polstellen $\lambda_1, \dots, \lambda_r$ negativ, dann gilt wegen $y(t) = \sum_{k=1}^r \tilde{C}_k e^{J_k t} \tilde{x}_{0,k}$ auch $\lim_{t \rightarrow \infty} y(t) = 0$. Hat umgekehrt nur einer der Eigenwerte einen positiven Realteil, dann wächst der zugehörige Beitrag zu $y(t)$ über alle Grenzen: das System ist instabil. Damit sind unsere beiden Aussagen oben bewiesen.

Eine ähnliche Argumentation kann man auch für die Eingangs-Ausgangsdifferentialgleichung führen, wobei man den Lösungsansatz $y(t) = y_0 e^{\lambda t}$ machen kann, und sieht, dass dieser Ansatz für jedes λ , das eine Nullstelle des

charakteristischen Polynoms ist, tatsächlich eine Lösung liefert, denn:

$$\begin{aligned} & \frac{d^n y}{dt^n}(t) + a_{n-1} \frac{d^{(n-1)}y}{dt^{n-1}}(t) + \dots a_1 \frac{dy}{dt}(t) + a_0 y(t) \\ = & y_0 e^{\lambda t} [\lambda^n + a_{n-1} \lambda^{n-1} + \dots a_1 \lambda + a_0] \\ = & 0. \end{aligned}$$

Man sieht daran zum Beispiel, dass eine einzige Polstelle λ_1 mit positivem Realteil eine Lösung $y_0 e^{\lambda_1 t}$ zulässt, die exponentiell wächst.¹

¹Um alle möglichen Lösungen zu analysieren, was zum Beweis der Stabilität nötig wäre, könnte man die Tatsache verwenden, dass jede mögliche Lösung die Form $y(t) = c_0 t^m e^{\lambda t}$ hat, mit geeignetem $m \in \mathbb{N}$ und $\lambda \in \mathbb{C}$ und $c_0 \in \mathbb{C}$, und diesen Ausdruck dann in die Differentialgleichung einsetzen.

Kapitel 5

Dynamisches Verhalten linearer Systeme

In diesem Kapitel werden wir grundlegende Eigenschaften linearer zeitinvarianter Systeme mit einem Input und einem Output (LTI-SISO Systeme). Wir interessieren uns für die Frage, welches Ausgangssignal entsteht, wenn man ein gegebenes Eingangssignal in das System schickt. Die entscheidende Rolle spielt dabei das sogenannte “Superpositionsprinzip”, das eigentlich nichts anderes ist als die praktische Konsequenz aus Linearität und Zeitinvarianz. Das Superpositionsprinzip besagt, dass man, wenn man die Outputsignale zu gegebenen Inputsignalen kennt, man auch für jede Überlagerung dieser Eingangssignale das Ausgangssignal berechnen kann, indem man einfach die entsprechenden Outputsignale überlagert. Das Superpositionsprinzip erlaubt es zum Beispiel, bei der Schallfortpflanzung in einem Raum von einem Lautsprecher ins Ohr eines Zuhörers die verschiedenen Frequenzen separat zu behandeln: weiß man aus unabhängigen Experimenten, wie zwei verschiedene Frequenzen durch den Raum geleitet werden, dann weiß man auch, wie ihre Überlagerung aussieht. Eine ganz wichtige und auch sehr anschauliche Rolle spielt die sogenannte “Sprungantwort”, mit der wir das Kapitel beginnen wollen.

5.1 Die Sprungantwort

Was passiert, wenn man für ein ruhendes LTI-SISO System eine Sprungfunktion, die zum Zeitpunkt Null bei Null startet und dann auf den Wert eins springt, als Input in das System gibt? Wir nennen diese Eingangsfunktion den “Einheitssprung” und nutzen dafür das Symbol $\sigma(t)$

$$\sigma(t) := \begin{cases} 0 & \text{wenn } t < 0 \\ 1 & \text{wenn } t \geq 0 \end{cases}$$

Wenn wir also den Input auf $u(t) = \sigma(t)$ setzen, dann kann man das daraus resultierende Outputsignal beobachten und aufzeichnen. Wir nennen dieses Output-Signal die “Sprungantwort” (oder auch “Übergangsfunktion”), und nennen diese $h(t)$.

Wegen der Kausalität ist diese Funktion für negative t Null, und wir wollen diese Eigenschaft für alle Funktionen in diesem Kapitel annehmen. Für stabile Systeme konvergiert die Sprungantwort nach einer gewissen Zeit gegen einen konstanten Wert, den wir etwas salopp $h(\infty)$ nennen können. Diesen Wert nennen wir den “Verstärkungsfaktor” oder auf Englisch auch *DC-Gain*. Jedes System ist eindeutig durch seine Sprungantwort charakterisiert. Noch mehr: man kann jedes Eingangssignal $u(t)$ als eine Überlagerung verschieden großer und in der Zeit verschobener Einheits sprünge auffassen (auch negative sind zugelassen), und dann aufgrund der Linearität und Zeitinvarianz den Output als eine Überlagerung der daraus resultierenden Sprungantworten auffassen.

Wie können wir die Sprungantwort berechnen? Im Falle der Zustandsraumdarstellung hilft uns dabei die analytische Formel (2.43), die für ein LTI-System mit Inputs und Outputs gegeben ist durch

$$y(t) = Ce^{At}x_0 + \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau + Du(t). \quad (5.1)$$

Nimmt man an, dass das System in Ruhe startet, der Anfangswert x_0 also exakt Null ist, dann gilt die folgende Formel:

$$y(t) = \int_0^t \underbrace{Ce^{A(t-\tau)}B}_{=g(t-\tau)}u(\tau)d\tau + Du(t). \quad (5.2)$$

Wir nennen die Funktion $g(t)$ die “Gewichtsfunktion”. Ihren Wert an der Stelle $t = 0$ lassen wir noch einen Moment offen.

Wählt man für den Eingang $u(t)$ den Einheitsprung $\sigma(t)$, dann wird der Output $y(t)$ zur Sprungantwort $h(t)$, also für $t \geq 0$

$$h(t) = \int_0^{t-} g(t-\tau)d\tau + D\sigma(t) = \int_{0+}^t g(\tau)d\tau + D \quad (5.3)$$

Die Sprungantwort $h(t)$ ist also nichts anderes als das Integral der Gewichtsfunktion $g(t)$, plus den Wert D . Da wir den Wert der Gewichtsfunktion an der Stelle Null noch offengelassen haben, enden bzw. beginnen wir das Integral infinitesimal kurz vor bzw. nach diesem Wert, was man durch die obere Integrationsgrenze $t-$ bzw. untere Integrationsgrenze $0+$ in den Integralen ausdrücken kann. Die Sprungantwort ist ein so anschauliches Objekt, dass man sie in Blockschaltbildern oft zur Kennzeichnung des entsprechenden Blocks verwendet. Die Sprungantworten für einige wichtige Übertragungsglieder sind in Abbildung 3.3 gezeigt.

5.2 Der Dirac-Impuls und die Impulsantwort

Gibt es einen Input, der als Output die Gewichtsfunktion $g(t)$ erzeugt? Da die Sprungantwort $h(t)$ als Integral der Gewichtsfunktion $g(t)$ gewonnen wird, ist die Gewichtsfunktion nichts anderes als die Ableitung der Sprungantwort, also

$$g(t) = \frac{dh}{dt}(t) \quad \text{für } t \in (0, \infty).$$

Da die Sprungantwort durch den Einheitssprung erzeugt wird, müßte man nur den Einheitssprung ableiten und diese Ableitung als Input in das System geben; aufgrund der Linearität des Systems wäre der Output dann die Gewichtsfunktion. Wenn wir also $u(t) = \frac{d\sigma}{dt}(t)$ setzen, dann wird als Systemantwort $y(t) = \frac{dh}{dt}(t)$, also $y(t) = g(t)$. Die Ableitung des Einheitssprungs ist allerdings ein seltsames Objekt, das man eigentlich nicht mehr Funktion nennen darf, sondern Distribution, aber das wir manchmal trotzdem die "Dirac'sche Deltafunktion" oder "Dirac-Impuls" nennen, und mit dem Symbol $\delta(t)$ bezeichnen. Es ist definiert durch

$$\delta(t) = \frac{d\sigma}{dt}(t)$$

und hat die Eigenschaft, dass für jede beliebige stetige Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ gilt, dass

$$\int_{-\infty}^{\infty} f(t)\delta(t) = f(0).$$

Der Zusammenhang zwischen Einheitssprung und Dirac-Impuls ist in Abb. 5.1 illustriert. Mit Hilfe der Dirac'schen Deltafunktion können wir auch den Wert der Gewichtsfunktion g an der Stelle Null bestimmen: im Falle $D \neq 0$ springt die Sprungantwort $h(t)$ bei $t = 0$ vom Wert 0 auf den Wert D . Man kann also definieren:

$$g(t) := \begin{cases} 0 & \text{wenn } t < 0 \\ D\delta(t) & \text{wenn } t = 0 \\ Ce^{At}B & \text{wenn } t > 0 \end{cases}$$

und bekommt dann die einfache Formel

$$y(t) = \int_0^t g(t-\tau)u(\tau)d\tau, \quad (5.4)$$

wobei wir bei der Integration den Dirac-Impuls an der Stelle Null explizit einschließen, was wir in Zukunft immer annehmen möchten (aber was man auch durch die untere Integrationsgrenze $0-$ ausdrücken kann). Sobald wir die Sprungantwort $h(t)$ eines Systems wissen, wissen wir auch die Impulsantwort $g(t) = \dot{h}(t)$, und diese Impulsantwort $g(t)$ erlaubt uns, eine explizite Formel für den Output $y(t)$ bei beliebigen Inputsignalen $u(t)$ aufzustellen. Die Impulsantwort erlaubt uns also, das System aus der Input-Outputperspektive vollständig zu beschreiben. Etwas überspitzt, aber mathematisch völlig korrekt könnte man zu einem unbekanntem System, dass man zufällig auf einer Party trifft, sagen: "Sag mir, welche Impulsantwort Du hast, und ich sage Dir, was für ein System Du bist". Wir werden im nächsten Kapitel sehen, dass die Laplace-Transformierte der Impulsantwort $g(t)$ eine sehr wichtige Rolle beim Verstehen der Systemeigenschaften spielt; sie wird die "Übertragungsfunktion" (engl. *transfer function*) genannt.

5.2.1 Totzeitglieder

Die Formel (5.4) kann nicht nur mit Hilfe der expliziten Lösungsformel für Differentialgleichungen gewonnen werden, sondern kann noch allgemeiner für alle Systeme

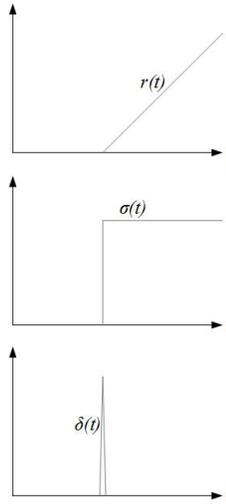


Abbildung 5.1: Zusammenhang zwischen Rampenfunktion $r(t)$, Einheitssprung $\sigma(t)$, und Dirac-Impuls $\delta(t)$. Wenn wir die Rampenfunktion $r(t)$ (mit $r(t) = t$ für $t \geq 0$ und $r(t) = 0$ für $t < 0$) einmal “ableiten” erhalten wir den Einheitssprung $\sigma(t) = \frac{d}{dt}r(t)$, wenn wir diesen noch einmal im verallgemeinerten Sinne “ableiten”, erhalten wir den Dirac-Impuls $\delta(t) = \frac{d}{dt}\sigma(t)$.

gezeigt werden, für die Zeitinvarianz und Linearität gelten. Ein Beispiel für ein System, das nicht als gewöhnliche Differentialgleichung geschrieben werden kann, das aber auch ein LTI-SISO System ist, ist das sogenannte “Totzeitglied”, das den Input einfach nur eine gewisse Zeit t_{tot} später wieder ausgibt (siehe Abb. 5.2). Es gilt also für ein Totzeitglied, dass

$$y(t) = u(t - t_{\text{tot}}), \quad \text{für } t \in [t_{\text{tot}}, \infty).$$

Die Impulsantwort eines Totzeitgliedes ist gegeben durch

$$g(t) = \delta(t - t_{\text{tot}}).$$

Man rechnet leicht nach, dass diese Funktion g tatsächlich das gewünschte Ergebnis gibt:

$$\begin{aligned} y(t) &= \int_0^t g(t - \tau)u(\tau)d\tau \\ &= \int_0^t \delta(t - \tau - t_{\text{tot}})u(\tau)d\tau \\ &= \begin{cases} u(t - t_{\text{tot}}) & \text{wenn } t \geq t_{\text{tot}} \\ 0 & \text{wenn } t < t_{\text{tot}} \end{cases} \end{aligned}$$

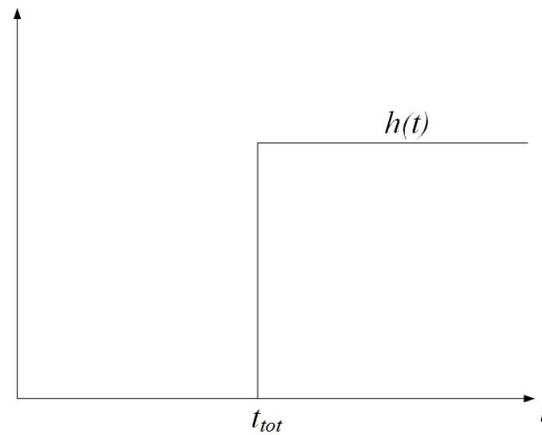


Abbildung 5.2: Darstellung eines Totzeitglieds.

5.2.2 Approximation des Totzeitglieds durch ein PT_n -Glied

Da das Totzeitglied keine Darstellung durch eine Differentialgleichung oder ein Zustandsraummodell hat, muss man es für die meisten in diesem Skript behandelten Methoden durch ein solches approximieren. Eine mögliche und sehr einfache Approximation des Totzeitgliedes mit Totzeit t_{tot} ist gegeben durch ein sogenanntes PT_n -Glied, das nichts anderes ist als eine Hintereinanderschaltung von n PT_1 -Gliedern, von denen jedes die Verzögerung $T = \frac{t_{tot}}{n}$ hat. In Zustandsform ist dies gegeben durch

$$\dot{x}_1(t) = \frac{1}{T}(u(t) - x_1(t)), \quad (5.5)$$

$$\dot{x}_2(t) = \frac{1}{T}(x_1(t) - x_2(t)), \quad (5.6)$$

$$\vdots$$

$$\dot{x}_n(t) = \frac{1}{T}(x_{n-1}(t) - x_n(t)), \quad (5.7)$$

$$y(t) = x_n(t). \quad (5.8)$$

Eine Simulation der Sprungantwort zeigt, dass der dazugehörige Output keine exakte Sprungfunktion ist, sondern abgerundete Ecken hat, der Sprung wird sozusagen “verschmiert”, was man im Kontext der Simulation von Partiellen Differentialgleichungen (engl. partial differential equation, PDE) auch “numerische Diffusion” nennt. Je höher der Grad n des Systems ist, desto besser ist die Approximation des Totzeitgliedes. Im Limit $n \rightarrow \infty$ entspricht das obige Differentialgleichungssystem einem sogenannten “System mit verteilten Parametern”, es wird eine instationäre PDE mit einem auf dem Interval $[0, L]$ verteiltem Zustand $x : [0, L] \times \mathbb{R} \rightarrow \mathbb{R}, (z, t) \mapsto x(z, t)$. Die PDE ist

gegeben durch

$$\dot{x}(z, t) = -\frac{L}{t_{\text{tot}}} \frac{\partial x}{\partial z}(z, t), \quad \text{für } z \in [0, L] \quad (5.9)$$

mit der linksseitigen Randbedingung $x(0, t) = u(t)$ und mit der Outputgleichung $y(t) = x(L, t)$. Das durch diese PDE beschriebene System ist äquivalent zum originalen Totzeitglied (ohne Approximationsfehler), und Diskretisierung der partiellen Ableitung nach dem Index z mit finiten Differenzen führt auf das ODE System (5.5)-(5.8), also das PT_n -Glied.

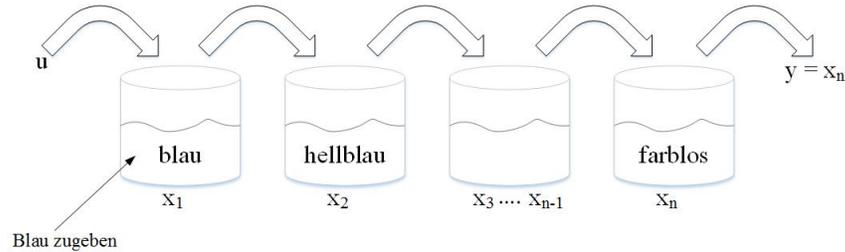


Abbildung 5.3: Das PT_n -Glied als Hintereinanderschaltung von Wassertanks die mit blaugefärbtem Wasser gespeist werden.

Eine Realisierung des PT_n -Gliedes ist gegeben durch eine Hintereinanderschaltung von Wassertanks, durch die ein kontinuierlicher Wasserstrom fließt, und bei der man blaue Tinte in den Zustrom des ersten Tanks zufügt, siehe Abb. 5.3. Das Wasser fließt immer vom linksseitigen Tank in den rechts davon stehenden. Der Input $u(t)$ ist gegeben durch die Konzentration der blauen Tinte im Wasser des Zustroms in den ersten Tank, und die Zustände x_1 bis x_n durch die Tintenkonzentration in jedem der Tanks (alle in Mol/Liter). Wenn V (in Liter) das Volumen jedes Tanks ist und Q (in Liter/s) die jeweils ein- und ausfließende Wassermenge pro Sekunde ist, dann ist die Tintenmenge in jedem Tank gegeben durch Vx_i (in Mol/Liter). Die Veränderung der Tintenmenge $V\dot{x}_i$ ergibt sich durch die Differenz der im Zustrom enthaltenen Tinte und der ausfließenden Tintenmenge. Wir erhalten die Differentialgleichung jedes Zustands also aus Massenerhaltung der Tintenmenge, und für den ersten Tank ergibt sich

$$V\dot{x}_1 = Qu - Qx_1,$$

was man noch durch V dividieren kann. Ebenso ist die Differentialgleichung für die anderen Zustände gegeben durch

$$\dot{x}_i = \frac{Q}{V}(x_{i-1} - x_i), \quad i = 2, \dots, n.$$

Man erhält erneut das System (5.5)-(5.8), wobei die Zeitkonstante jedes Tanks $T = \frac{V}{Q}$ ist.

Wenn man mit farblosem Wasserzustrom und klaren Wassertanks startet, und zur Zeit $t = 0$ beginnt, dem Zustrom mit konstanter Konzentration blau gefärbtes Wasser

zuzugeben ($u(t) = u_{\text{const}}$ für $t \geq 0$), dann braucht die blaue Farbe eine Weile, bis sie rechts angekommen ist - nämlich in etwa die Zeit nT . Diese Zeit ist gleich dem Gesamtwasservolumen $V_{\text{tot}} = nV$ dividiert durch die Wasserflussmenge Q . Wenn die Sprungantwort gänzlich vollzogen ist, sind alle Tanks gleich blau gefärbt und haben die Tintenkonzentration $x_i = u_{\text{const}}$. Dies gilt insbesondere auch für den letzten Zustand, der als Ausgang des Systems aufgefasst wird.

Ließe man den Wasserstrom Q gleich, verteilte aber das Wasservolumen V_{tot} aller Tanks auf noch mehr miteinander verbundene Tanks, dann erhielten wir im Grenzfalle von unendlich vielen Tanks das Totzeitglied mit Totzeit $t_{\text{tot}} = V_{\text{tot}}/Q$, das ebenso durch die partielle Differentialgleichung 5.9 beschrieben werden kann. Wenn das gesamte Tanksystem die Länge L hat und die Querschnittsfläche A , dann ist das Gesamtvolumen gegeben durch $V_{\text{tot}} = LA$. Das Wasser fließt mit Geschwindigkeit $v = Q/A$ über die Strecke L . Die Gesamtdauer, die es dafür braucht, ist natürlich gegeben durch $\frac{L}{v} = \frac{LA}{Q} = t_{\text{tot}}$.

5.3 Eingangs-Ausgangs-Stabilität (BIBO-Stability)

Wir haben bereits eine Bedingung kennengelernt, die garantiert, dass ein System bei auf Null gesetzten Inputs $u(t) = 0$ mit seinem Output $y(t)$ auch gegen Null konvergiert: dies war dann der Fall, wenn alle Polstellen des Systems in der linken Halbebene der komplexen Zahlenebene liegen. Was passiert aber, wenn der Input $u(t)$ klein, aber von Null verschieden ist? Um diesen Fall zu behandeln, definieren wir zunächst einen neuen Stabilitätsbegriff: "Eingangs-Ausgangsstabilität", oder, genauer und auf Englisch: *bounded input bounded output (BIBO) stability*. Ein System heisst genau dann BIBO-stabil, wenn man für jede beliebig klein gewählte Output-Obergrenze y_{max} eine Input-Obergrenze u_{max} finden kann, so dass für jedes durch u_{max} begrenzte Input Signal $u : \mathbb{R} \rightarrow \mathbb{R}$ das dadurch erzeugte Ausgangssignal $y : \mathbb{R} \rightarrow \mathbb{R}$ durch y_{max} begrenzt wird. Es soll also für das System gelten:

$$|u(t)| \leq u_{\text{max}} \quad \text{für alle } t \in \mathbb{R} \quad \text{impliziert dass} \quad |y(t)| \leq y_{\text{max}} \quad \text{für alle } t \in \mathbb{R}.$$

Für LTI-SISO Systeme kann man BIBO-Stabilität wegen der Linearität auch durch folgende Bedingung charakterisieren: für ein BIBO-stabiles System gibt eine positive Konstante $K_{\text{max}} \in \mathbb{R}$ so dass

$$|u(t)| \leq 1 \quad \text{für alle } t \in \mathbb{R} \quad \text{impliziert dass} \quad |y(t)| \leq K_{\text{max}} \quad \text{für alle } t \in \mathbb{R}.$$

Man kann leicht zeigen (Übungsaufgabe), dass diese Konstante K_{max} explizit berechnet werden kann als das Integral über den Betrag der Impulsantwort, also

$$K_{\text{max}} = \int_0^{\infty} |g(t)| dt. \quad (5.10)$$

Ein LTI-SISO System ist also genau dann BIBO-stabil, wenn das obige Integral existiert und endlich ist.

Falls die Impulsantwort eine Summe aus endlich vielen Termen der Form $t^m e^{\lambda t}$ ist, mit $m \in \mathbb{N}$ und $\lambda \in \mathbb{C}$ mit negativem Realteil, dann ist das Integral endlich. Alle

bisher behandelten LTI-SISO Systeme in Differentialgleichungsform sind also BIBO-stabil, wenn alle Polstellen negative Realteile haben. Das gilt auch, falls g zusätzlich noch eine endliche Anzahl von Dirac-Impulsen enthält (Totzeitglieder mit direktem Durchgriff).

Als ein Gegenbeispiel, also ein System, das nicht BIBO-stabil ist, können wir ein PT_2 -Glied ohne Dämpfung betrachten, das durch die Differentialgleichung

$$T^2 \ddot{y}(t) + y(t) = u(t)$$

beschrieben ist. Dieses System hat die Sprungantwort

$$h(t) = 1 - \cos \frac{t}{T},$$

wie man leicht durch Einsetzen von $\ddot{h}(t) = \frac{1}{T^2} \cos \frac{t}{T}$ in die Systemgleichung und Verifikation der Anfangsbedingungen $h(0) = 0$ und $\dot{h}(0) = 0$ sieht. Seine Impulsantwort ist gegeben durch die Ableitung der Sprungantwort, $g(t) = \frac{dh}{dt}(t)$, also

$$g(t) = \frac{1}{T} \sin \frac{t}{T}.$$

Das zum Testen der BIBO-Stabilität entscheidende Integral in Gleichung (5.10) ist nicht endlich, denn es gilt für eine beliebig wachsende Integrationsobergrenze $k\pi T$ mit $k \in \mathbb{N}$ und $k \rightarrow \infty$ dass:

$$\int_0^{k\pi T} |g(t)| dt = k \underbrace{\int_0^{\pi T} \frac{1}{T} \sin \frac{t}{T} dt}_{=2} = 2k \rightarrow \infty.$$

Selbst einige Systeme mit abklingender Impulsantwort sind nicht BIBO-stabil, z.B. ist das LTI-SISO System, das die Impulsantwort

$$g(t) = \frac{1}{1+t}$$

hat, nicht BIBO-stabil, da für wachsende Integralobergrenze $\tau \rightarrow \infty$ gilt:

$$\int_0^\tau |g(t)| dt = \int_0^\tau \frac{1}{1+t} dt = \log(1+\tau) \rightarrow \infty.$$

Wir kennen auch ein anderes wichtiges Glied, das nicht BIBO-stabil ist: das Integrierglied. Es hat den Einheitssprung als Impulsantwort, und das Integral über den Betrag des Einheitssprungs ist nicht endlich.

5.4 Charakteristische Größen von BIBO-Systemen

Betrachtet man die Sprungantwort eines BIBO-Systems, dann konvergiert diese am Ende gegen einen Gleichgewichtszustand, den wir auch den *Verstärkungsfaktor* oder das

DC-Gain nennen (von Engl. *direct current*) und in dieser Vorlesung mit dem Symbol $h(\infty)$ benennen. Wir definieren also:

$$h(\infty) = \int_0^{\infty} g(t) dt.$$

Meist sind die zwei wesentlichen Ziele des Reglerentwurfs, dass das geregelte System BIBO-stabil ist und ein DC-Gain mit dem Wert eins hat, das heisst, dass es auf einen Einheitssprung nach etwas Wartezeit auch den Ausgangswert eins erzeugt. Zur Charakterisierung des Übergangsverhaltens nutzt man oft noch drei wichtige Größen: die Überschwinghöhe h_{\max} , die Anstiegszeit t_{rise} (engl. *rise time*) und die Einschwingzeit t_{settling} (engl. *settling time*) (siehe Abbildung 5.4).

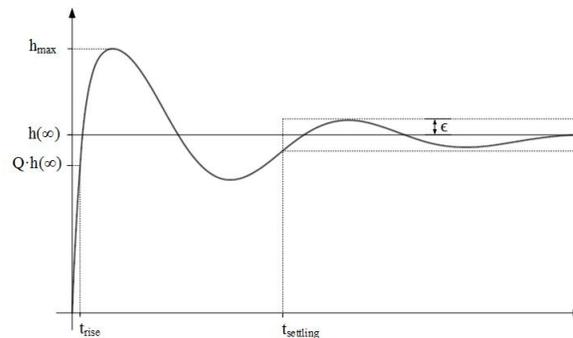


Abbildung 5.4: Definition der Anstiegszeit t_{rise} , der Einschwingzeit t_{settling} und der Überschwinghöhe h_{\max} .

Die Überschwinghöhe h_{\max} ist der höchste Wert, den die Sprungantwort jemals annimmt, oder äquivalent die kleinste Zahl, so dass für alle $t \in [0, \infty)$ gilt, dass

$$h(t) \leq h_{\max}.$$

Meist möchte man durch guten Reglerentwurf erreichen, dass die relative Überschwinghöhe, $\frac{h_{\max} - h(\infty)}{h(\infty)}$, nicht zu gross ist.

Die Anstiegszeit t_{rise} ist der erste Zeitpunkt, bei dem der Output einen gewissen Anteil Q des DC-Gains erreicht, z.B. 90%. Es gilt also

$$h(t_{\text{rise}}) = Qh(\infty)$$

meist mit $Q = 9/10$, und für alle $t \in [0, t_{\text{rise}})$ gilt $\frac{h(t)}{h(\infty)} < Q$.

Die Einschwingzeit t_{settling} ist der erste Zeitpunkt, von dem ab der Output in einem engen Intervall um den gewünschten Endwert $h(\infty)$ bleibt. Als Intervallgröße wählt man eine für die jeweilige Anwendung ausreichend kleine Zahl ϵ , z.B. 1%. Die Einschwingzeit t_{settling} ist also der früheste Zeitpunkt, für den gilt, dass:

$$\text{für alle } t > t_{\text{settling}} : \frac{|h(t) - h(\infty)|}{h(\infty)} \leq \epsilon.$$

Kapitel 6

Die Übertragungsfunktion

In diesem Kapitel werden wir eine mächtige Technik erlernen, die es erlaubt, auf einfache Weise anzugeben, welches Verhalten komplexe LTI-SISO Systeme haben die z.B. aus einer Hintereinanderschaltung oder Rückkopplung verschiedener Systeme entstehen. Dafür abstrahieren wir weiter und verlassen den Zeitbereich, um uns fast vollständig in der Welt der komplexen Zahlen und der Laplace-Transformation weiterzubewegen.

Wir haben im vorigen Kapitel gesehen, dass man alles über ein LTI-SISO System weiß, wenn man seine Impulsantwort (oder Gewichtsfunktion) $g : \mathbb{R} \rightarrow \mathbb{R}$ kennt. Diese Funktion ist für negative t Null, und es gilt die entscheidende Formel:

$$y(t) = \int_0^t g(t - \tau)u(\tau)d\tau.$$

Dabei hatten wir angenommen, dass das System zur Zeit Null im Ruhezustand startet, und erst ab dann die Inputs $u(t)$ auf das System wirken. Wir werden annehmen, dass der Input davor Null war. Da die Gewichtsfunktion $g(t)$ (die ja nichts anderes als die Impulsantwort ist) für negative t auch den Wert Null annimmt, kann man das obige Integral auch wie folgt schreiben:

$$y(t) = \int_{-\infty}^{\infty} g(t - \tau)u(\tau)d\tau. \quad (6.1)$$

Diese Formel ist für BIBO-Systeme selbst dann wohldefiniert, wenn der Input $u(t)$ in der vor dem Zeitpunkt Null liegenden Vergangenheit ungleich Null war. Wir wollen in diesem Kapitel solche Inputtrajektorien zulassen, und interessieren uns besonders für eine spezielle Klasse.

Bevor wir beginnen, noch eine kleine Bemerkung: das Integral in Gleichung (6.1) bezeichnet man auch als "Faltung" der Funktionen u und g (siehe Abb. 6.1). Man kann das Integral auch rückwärts durchlaufen, und bekommt dann die äquivalente Formel

$$y(t) = \int_{-\infty}^{\infty} g(\tau)u(t - \tau)d\tau. \quad (6.2)$$

Da (6.1) und (6.2) das gleiche Ergebnis liefern, können wir je nach Bedarf die eine oder andere Formel verwenden.

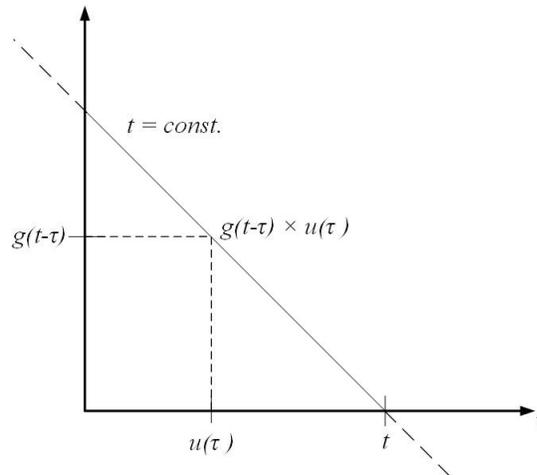


Abbildung 6.1: Faltung der Funktionen u und g .

6.1 Antwort auf exponentielle Inputs

Was passiert, wenn wir als Input in das System ein exponentialfunktionsförmiges Signal, als ein Signal der Form

$$u(t) = U e^{st}$$

mit gegebenen konstanten Zahlen $U, s \in \mathbb{C}$ geben? Es wird wegen der Formel (6.1) zu folgendem Output führen:

$$y(t) = U \int_{-\infty}^{\infty} g(\tau) e^{s(t-\tau)} d\tau.$$

Aus diesem Integral können wir den Faktor e^{st} herausziehen, und erhalten die Formel:

$$y(t) = U e^{st} \underbrace{\int_{-\infty}^{\infty} g(\tau) e^{-s\tau} d\tau}_{=G(s)}. \quad (6.3)$$

Zwei Dinge sind daran bemerkenswert: zum ersten, dass der Output wieder die Form einer Exponentialfunktion hat, und zum zweiten, dass der exponentielle Output aus dem exponentiellen Input durch eine reine Multiplikation mit der Zahl $G(s)$ hervorgeht! Dies ist eine Konsequenz aus Linearität und Zeitinvarianz des Systems. Da die

Funktion $g(t)$ für negative t Null war können wir die komplexe Zahl $G(s)$ auch wie folgt definieren:

$$G(s) = \int_0^{\infty} g(\tau)e^{-s\tau} d\tau.$$

Man kann nun $G(s)$ selbst wieder als Funktion der komplexen Variablen s auffassen, und die Funktion G ist nichts anderes als die Laplace-Transformierte der Impulsantwort. Man nennt sie auch die Übertragungsfunktion.

Fassen wir zusammen: auf einen exponentialförmigen Input der Form $u(t) = Ue^{st}$ reagiert das System mit einem Output $y(t) = Ye^{st}$ wobei $Y = G(s)U$ eine konstante komplexe Zahl ist.

6.2 Laplace-Transformation

Wir betrachten jetzt allgemeine Signale $f(t)$, für die wir meist annehmen, dass sie für negative Zeiten verschwinden. Signale sind Zeitfunktionen von den reellen Zahlen in die reellen Zahlen, also $f : \mathbb{R} \rightarrow \mathbb{R}$.

Die Laplace Transformierte zu einem Signal $f(t)$ benennen wir mit $F(s)$, also dem gleichen Buchstaben wie das Zeitsignal, aber als Grossbuchstaben. Als Argument benutzen wir immer die Variable s , die eine komplexe Zahl ist. Wir definieren die Funktion $F(s)$ durch

$$F(s) = \int_0^{\infty} f(t)e^{-st} dt.$$

Die Laplace-Transformierte ist eine Funktion von den komplexen Zahlen in die komplexen Zahlen, also $F : \mathbb{C} \rightarrow \mathbb{C}$. Man kann zeigen, dass dieser Ausdruck für alle uns interessierenden Zeitsignale wohldefiniert ist, wenn der Realteil von s ausreichend groß gewählt wurde. Aber auch für die meisten Werte von s mit kleinerem Realteil kann man die Funktion analytisch fortsetzen. Es können in der Funktion $F(s)$ jedoch Singularitäten (Polstellen) auftreten, also Werte von s , für die $F(s)$ den Wert unendlich annimmt. Für eine gegebene Funktion $F(s)$ gibt es jedoch in allen uns interessierenden Fällen ein $a \in \mathbb{R}$ als untere Grenze, so dass für alle s mit $\text{Re}(s) > a$ die Funktion $F(s)$ wohldefiniert und endlich ist. Man nennt diesen Bereich der komplexen Ebene die "Konvergenzhalbebene" für F . Ähnlich wie bei der Fourier-Transformation gibt es für die Laplace-Transformation eine Rücktransformation, und diese ist für ein gegebenes $F(s)$ mit einer unteren Konvergenzgrenze a gegeben durch das folgende Integral:

$$f(t) = \frac{1}{2\pi j} \int_{s=a-j\infty}^{a+j\infty} F(s)e^{st} ds.$$

Man beachte, dass der Integrationspfad so gewählt ist, dass er ganz in der Konvergenzhalbebene verläuft. Das Zeitsignal $f(t)$ und seine Laplace-Transformierte $F(s)$ enthalten also genau die gleiche Information und können ineinander umgewandelt werden. Sie sind eng miteinander verbunden und können als verschiedene Sichtweisen auf das gleiche Objekt angesehen werden. Wir benutzen für die Paarung zwischen einem Zeitsignal und seiner Laplace-Transformierten oft das Symbol $\circ \text{---} \bullet$ mit dem vollen Kreis

auf der Seite der Laplace-Transformierten, wir schreiben also

$$f(t) \quad \circ \text{---} \bullet \quad F(s)$$

um auszudrücken, dass f und F ein Laplace-Paar sind. Wir berechnen zur Übung drei Beispiele für Laplace-Transformierte wichtiger Zeitfunktionen:

$$f(t) = \sigma(t) \quad \circ \text{---} \bullet \quad F(s) = \int_0^{\infty} e^{-st} dt = \frac{1}{s}, \quad (6.4)$$

$$f(t) = \delta(t) \quad \circ \text{---} \bullet \quad F(s) = \int_0^{\infty} e^{-st} \delta(t) dt = 1, \quad (6.5)$$

$$f(t) = \delta(t - t_0) \quad \circ \text{---} \bullet \quad F(s) = \int_0^{\infty} e^{-st} \delta(t - t_0) dt = e^{-st_0}. \quad (6.6)$$

6.3 Eigenschaften der Laplace-Transformation

Die Laplace-Transformation erfüllt viele interessante Eigenschaften, von denen wir jedoch in diesem Kapitel nur die vier wichtigsten besprechen wollen, mit deren Hilfe wir bereits viele Berechnungen durchführen können.

6.3.1 Überlagerung

Für zwei Zeitfunktionen f_1, f_2 und Zahlen $a_1, a_2 \in \mathbb{C}$ gilt:

$$a_1 f_1(t) + a_2 f_2(t) \quad \circ \text{---} \bullet \quad a_1 F_1(s) + a_2 F_2(s),$$

da die Laplace-Transformation eine lineare Operation ist.

6.3.2 Zeitableitung

Für die Zeitableitung einer Zeitfunktion f gilt:

$$\frac{df}{dt}(t) \quad \circ \text{---} \bullet \quad sF(s) - f(0).$$

Zum Beweis definieren wir zunächst der eindeutigen Notation wegen $\bar{f}(t) = \frac{df}{dt}(t)$ und berechnen die Laplace-Transformierte:

$$\begin{aligned} \bar{F}(s) &= \int_0^{\infty} e^{-st} \bar{f}(t) dt \\ &= \int_0^{\infty} e^{-st} \frac{df}{dt}(t) dt \\ &= \int_0^{\infty} \frac{d}{dt} (e^{-st} f(t)) dt - \int_0^{\infty} (-se^{-st} f(t)) dt \\ &= [e^{-st} f(t)]_0^{\infty} + s \int_0^{\infty} e^{-st} f(t) dt \\ &= -f(0) + sF(s). \end{aligned}$$

Oft nehmen wir für den Anfangswert $y(0)$ und alle seine Ableitungen an, dass sie verschwinden, die Zeitfunktion also zu allen Zeitpunkten $t < 0$ identisch Null war.

6.3.3 Höhere Ableitungen

Man kann die Differentiationsregel mehrmals auf sich selbst anwenden. Betrachtet man die Zeitableitung der Funktion $\bar{f} = \frac{df}{dt}$ aus dem vorigen Unterabschnitt, bekommt man für die zweite Ableitung die Relation

$$\begin{aligned} \frac{d^2 f}{dt^2}(t) & \circ \longrightarrow \bullet \quad s\bar{F}(s) - \bar{f}(0) \\ & = s^2 F(s) - sf(0) - \frac{df}{dt}(0). \end{aligned}$$

Nimmt man an, dass kurz vor dem Zeitpunkt $t = 0$ alle Zeitableitungen von y verschwinden, dann erhält man auch für höhere Ableitungen die einfache Formel

$$\frac{d^n f}{dt^n}(t) \circ \longrightarrow \bullet \quad s^n F(s)$$

6.3.4 Faltung

Besonders wichtig ist die Tatsache, dass die Transformierte der Faltung zweier Zeitfunktionen nichts anderes ist als das Produkt der beiden Laplace-Transformierten. Wenn wir das Resultat der Faltung zweier Funktionen f_1, f_2 definieren als

$$\bar{\bar{f}}(t) = \int_0^t f_1(t - \tau) f_2(\tau) d\tau$$

dann gilt

$$\bar{\bar{f}}(t) \circ \longrightarrow \bullet \quad \bar{\bar{F}}(s) = F_1(s) F_2(s).$$

Zum Beweis starten wir mit der Definition und formen den Ausdruck dann Schritt für Schritt um.

$$\begin{aligned}
 \bar{\bar{F}}(s) &= \int_0^{\infty} e^{-st} \bar{\bar{f}}(t) dt \\
 &= \int_0^{\infty} e^{-st} \int_0^t f_1(t-\tau) f_2(\tau) d\tau dt \\
 &= \int_0^{\infty} \int_0^t e^{-s(t-\tau)} e^{-s\tau} f_1(t-\tau) f_2(\tau) d\tau dt \\
 &= \int_0^{\infty} \int_0^t e^{-s(t-\tau)} f_1(t-\tau) e^{-s\tau} f_2(\tau) d\tau dt \\
 &= \int_0^{\infty} \int_{\tau}^{\infty} e^{-s(t-\tau)} f_1(t-\tau) e^{-s\tau} f_2(\tau) dt d\tau \\
 &= \int_0^{\infty} e^{-s\tau} f_2(\tau) \left(\int_{\tau}^{\infty} e^{-s(t-\tau)} f_1(t-\tau) dt \right) d\tau \\
 &= \int_0^{\infty} e^{-s\tau} f_2(\tau) \left(\int_0^{\infty} e^{-st'} f_1(t') dt' \right) d\tau \\
 &= \left(\int_0^{\infty} e^{-s\tau} f_2(\tau) d\tau \right) \left(\int_0^{\infty} e^{-st'} f_1(t') dt' \right) \\
 &= F_2(s) \cdot F_1(s)
 \end{aligned}$$

Die erstaunliche Tatsache, dass die Faltung im Laplace-Bereich ein ganz normales Produkt wird, ist der Hauptgrund für unser Interesse an der Laplace-Transformation und macht viele Rechnungen viel einfacher im Laplace-Bereich als im Zeitbereich. Es folgt daraus insbesondere, dass der Systemausgang eines LTI-SISO Systems im Laplace-Bereich durch einfache Multiplikation erhalten werden kann, denn es gilt folgende Korrespondenz:

$$\begin{aligned}
 y(t) &= \int_0^t g(t-\tau)u(\tau) d\tau \\
 \circ \\
 \bullet \\
 Y(s) &= G(s)U(s)
 \end{aligned}$$

Im Laplacebereich können wir mit der Übertragungsfunktion rechnen, unter Verwendung der normalen Rechenregeln für komplexe Zahlen, wie z.B. der Multiplikation und Division. Dies wird es sehr einfach machen, hintereinandergeschaltete oder sogar zurückgekoppelte Systeme zu beschreiben. Ganz allgemein kann man die Übertragungsfunktion definieren als den Quotient aus zueinander passenden Ausgangs- und Eingangssignalen im Laplace-Bereich:

$$G(s) = \frac{Y(s)}{U(s)}. \quad (6.7)$$

6.4 Übertragungsfunktion aus der Eingangs-Ausgangs-Differentialgleichung

Wir können die Übertragungsfunktion mit Hilfe der Definition (6.7) und den Umformungsregeln aus dem vorigen Abschnitt berechnen, wenn wir wissen, dass das System durch eine Eingangs-Ausgangs-Differentialgleichung beschrieben wird. Wir betrachten dafür erneut die Differentialgleichung und übertragen sie dann in den Laplace-Bereich:

$$\begin{aligned} \frac{d^n y}{dt^n}(t) + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}}(t) + \dots + a_0 y(t) &= b_q \frac{d^q u}{dt^q}(t) + \dots + b_0 u(t) \\ &\Downarrow \\ s^n Y(s) + a_{n-1} s^{n-1} Y(s) + \dots + a_0 Y(s) &= b_q s^q U(s) + \dots + b_0 U(s) \\ (s^n + a_{n-1} s^{n-1} + \dots + a_0) Y(s) &= (b_q s^q + \dots + b_0) U(s). \end{aligned}$$

Hieraus folgert man direkt, dass

$$\frac{Y(s)}{U(s)} = \frac{b_q s^q + \dots + b_0}{s^n + a_{n-1} s^{n-1} + \dots + a_0}.$$

Dieser Quotient ist nach Definition nichts anderes als die Übertragungsfunktion $G(s)$, die wir damit bereits erhalten haben, und die also eine gebrochene rationale Funktion ist:

$$G(s) = \frac{b_q s^q + \dots + b_0}{s^n + a_{n-1} s^{n-1} + \dots + a_0} = \frac{Z(s)}{N(s)} = b_q \frac{\prod_{i=1}^{r_0} (s - \lambda_{0,i})^{n_{0,i}}}{\prod_{i=1}^r (s - \lambda_i)^{n_i}} \quad (6.8)$$

mit einem Nennerpolynom $N(s)$ der Ordnung n und eine Zählerpolynom $Z(s)$ der Ordnung q . Das Nennerpolynom ist identisch mit dem charakteristischen Polynom $p_A(s)$ des Systems, und dadurch wird nachträglich der Name ‘‘Polstellen’’ für die Eigenwerte λ_i der Systemmatrix gerechtfertigt: wird der Nenner Null, hat man eine Polstelle der Übertragungsfunktion. Wird hingegen der Zähler Null, hat man eine Nullstelle der Übertragungsfunktion. Die Übertragungsfunktion hat genau n Polstellen und q Nullstellen, wenn wir mehrfache Null- und Polstellen mehrfach zählen, d.h. es gilt für die Nullstellenzerlegung der Nenner- und Zähler-Polynome, dass $n_1 + \dots + n_r = n$ und $n_{0,1} + \dots + n_{0,r_0} = q$. Wir erinnern noch einmal an die Tatsache, dass für physikalisch realisierbare Systeme immer gilt, dass $q \leq n$, und dass man den Nennergrad n als die *Ordnung* oder den *Grad* des Systems bezeichnet, und die Differenz $n - q$ des Nenner- und Zählergrades als den *relativen Grad* des Systems. Der Nennergrad n entspricht zudem der Zustandsdimension, wenn wir das selbe System in Zustandsform repräsentieren, was wir im folgenden Abschnitt tun möchten.

6.5 Übertragungsfunktion aus Zustandsdifferentialgleichung

Man kann mit Hilfe der Rechenregeln für Laplace-Transformierte auch aus der Zustandsraumdarstellung eines Systems die Übertragungsfunktion berechnen. Dafür be-

trachten wir ein System in Zustandsform und berechnen die Laplace-Transformierte davon

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x_0 = 0, \quad (6.9)$$

$$y(t) = Cx(t) + Du(t), \quad (6.10)$$

⋮

$$sX(s) = AX(s) + BU(s), \quad (6.11)$$

$$Y(s) = CX(s) + DU(s). \quad (6.12)$$

Man beachte, dass $X(s)$ eine Vektorfunktion mit n Komponenten ist, also $X : \mathbb{C} \rightarrow \mathbb{C}^n$ mit

$$X(s) = \begin{bmatrix} X_1(s) \\ \vdots \\ X_n(s) \end{bmatrix}.$$

Durch Umordnen von Gleichung (6.11) erhält man $(sI - A)X(s) = BU(s)$, was man invertieren kann zu

$$X(s) = (sI - A)^{-1}BU(s),$$

wobei die Matrixinverse genau dann existiert, wenn s kein Eigenwert λ_i der Systemmatrix A ist, also keine Polstelle. Durch Einsetzen in Gleichung (6.12) erhalten wir

$$Y(s) = [C(sI - A)^{-1}B + D]U(s),$$

also einen expliziten Ausdruck für die Übertragungsfunktion:

$$G(s) = C(sI - A)^{-1}B + D. \quad (6.13)$$

Diese Funktion ist für SISO Systeme trotz der auftretenden Matrixinversion im Inneren wieder eine ganz normale skalare Funktion, $G : \mathbb{C} \rightarrow \mathbb{C}$, da die inverse Matrix mit dem liegenden Vektor C von links und dem stehenden Vektor B von rechts multipliziert wird, und da D bereits eine reelle Zahl war.

Vergleicht man dies mit dem Ausdruck (6.8), den wir aus der Eingangs-Ausgangs-Differentialgleichung gewonnen hatten, so ist die Tatsache, dass beide für ein gegebenes System identisch sind, nicht auf den ersten Blick deutlich. Eine genauere Inspektion der Matrixinversen $(sI - A)^{-1}$ zeigt jedoch, dass sie in jedem Feld einen gebrochen rationalen Ausdruck stehen hat, in dessen Nenner immer das charakteristische Polynom $p_A(s)$ steht. Diese Ausdrücke werden durch Links- und Rechtsmultiplikation mit C und B einfach nur linearkombiniert und ergeben einen gebrochen rationalen Ausdruck

$$C(sI - A)^{-1}B = \frac{p'(s)}{p_A(s)}.$$

Man kann zeigen, dass der Grad des Zählerpolynoms $p'(s)$ geringer ist als n . Die Addition der Zahl D kann man dann durch Erweiterung mit $p_A(s)$ durchführen, es gilt also

$$C(sI - A)^{-1}B + D = \frac{p'(s) + Dp_A(s)}{p_A(s)} = \frac{Z(s)}{N(s)}.$$

wobei die Nenner und Zählerpolynome identisch sind mit denen in (6.8). Eine interessante Beobachtung ist, dass das Zählerpolynom genau dann den Grad $q = n$ hat, wenn $D \neq 0$ ist. Der relative Grad des Systems ist also genau dann Null, wenn man “direkten Durchgriff” des Inputs u auf den Output y hat.

6.6 Gekoppelte und geregelte Systeme

Die Tatsache, dass wir den Ausgang eines Systems im Laplace-Bereich durch eine Multiplikation des Inputs mit der Übertragungsfunktion berechnen können, erlaubt es uns nun, einige Rechenregeln für das Zusammenkoppeln von LTI-SISO Systemen abzuleiten. Wir werden hier die drei wichtigsten nennen. Da ein LTI-SISO System komplett durch seine Übertragungsfunktion $G(s)$ beschrieben wird, schreiben wir der Einfachheit halber oft “das System $G(s)$ ” anstatt des längeren, aber genaueren Ausdrucks “das System mit der Übertragungsfunktion $G(s)$ ”.

6.6.1 Reihenschaltung

Wenn man zwei LTI-SISO Systeme $G_1(s)$ und $G_2(s)$ hintereinanderschaltet (siehe Abbildung 6.2), also den Output des ersten Systems als Input des zweiten Systems nehmen, dann ist die Übertragungsfunktion des resultierenden Gesamtsystems $G(s)$ gegeben durch:

$$G(s) = G_2(s)G_1(s).$$

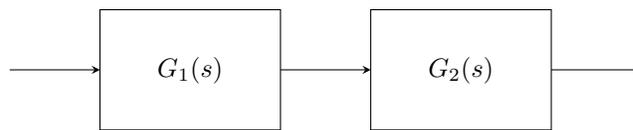


Abbildung 6.2: Blockschaltbild einer Reihenschaltung.

6.6.2 Parallelschaltung

Gibt man in zwei Systeme $G_1(s)$ und $G_2(s)$ denselben Eingang und addiert dann die resultierenden Ausgänge, um den Ausgang des Gesamtsystems zu erzeugen (siehe Abbildung 6.3), dann ist die Übertragungsfunktion des Gesamtsystems gegeben durch:

$$G(s) = G_1(s) + G_2(s).$$

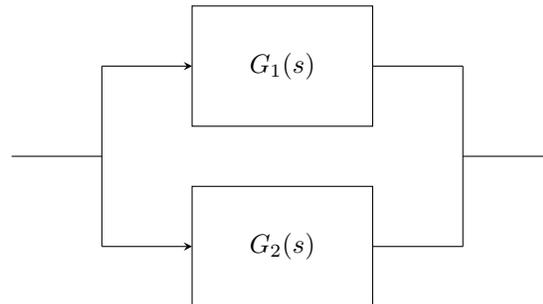


Abbildung 6.3: Blockschaltbild einer Parallelschaltung.

6.6.3 Rückkopplungsschaltung

In der Regelungstechnik wird oft der Output $Y(s)$ eines Systems $G_1(s)$ mit einem gewünschten Output verglichen, den wir als Input des Gesamtsystems auffassen und deshalb $U(s)$ nennen. Die Differenz $U(s) - Y(s)$ wird dann als Input in das System gegeben (siehe Abbildung 6.4). Daraus ergibt sich die Gleichung:

$$Y(s) = G_1(s)(U(s) - Y(s))$$

Wir können dies mit Hilfe der normalen Rechenregeln für Subtraktion und Division so umformen, dass $Y(s)$ eine explizite Funktion von $U(s)$ wird:

$$Y(s)(1 + G_1(s)) = G_1(s)U(s) \quad \Leftrightarrow \quad Y(s) = \frac{G_1(s)}{1 + G_1(s)}U(s).$$

Die Übertragungsfunktion des zurückgekoppelten Gesamtsystems ist also gegeben durch

$$G(s) = \frac{G_1(s)}{1 + G_1(s)}.$$

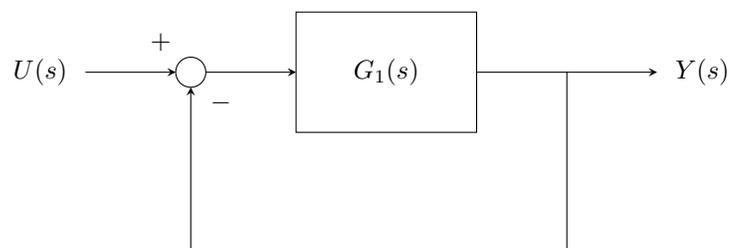


Abbildung 6.4: Blockschaltbild einer Rückkopplungsschaltung.

6.6.4 Beispiel: der einfach geregelte Traktor

Das kinematische Traktormodell aus Abschnitt 2.2 kann bei Fahrt in X -Richtung als SISO-System beschrieben werden, wenn man die Y -Koordinate als Output $y(t)$ und den Lenkwinkel $\alpha(t)$ als Input $u(t)$ auffasst. Durch Linearisierung kann man dieses Modell als Eingangs-Ausgangs-Differentialgleichung $\ddot{y}(t) = b_0 u(t)$ beschreiben, wobei der Koeffizient b_0 sich aus der Traktorgeschwindigkeit V und dem Achsabstand L ergibt als $b_0 = \frac{V^2}{L}$. Im Laplace-Bereich ergibt sich die Gleichung

$$s^2 Y(s) = b_0 U(s) \quad \Leftrightarrow \quad Y(s) = \frac{b_0}{s^2} U(s).$$

womit bereits die Übertragungsfunktion des Traktors gefunden ist:

$$G_{\text{Tr}}(s) = \frac{b_0}{s^2}.$$

Ein sehr einfacher Proportional-Regler könnte jetzt wie folgt aussehen: man gibt eine Referenz $r(t)$ vor und vergleicht den Output $y(t)$ damit. Den eingestellten Lenkwinkel $u(t)$ erhält man aus der Differenz durch Multiplikation mit einem Verstärkungsfaktor (engl. *gain*) K_P , man setzt also

$$u(t) = K_P(r(t) - y(t)),$$

oder, im Laplace-Bereich,

$$U(s) = K_P(R(s) - Y(s)).$$

Die Gesamt-Übertragungsfunktion $G(s) = Y(s)/R(s)$ vom Referenzwert $R(s)$ zum Ausgang $Y(s)$ kann man berechnen, indem man z.B. die Relation $U(s) = Y(s)/G_{\text{Tr}}(s)$ nutzt und in die obige Gleichung einsetzt, woraus man die Gleichung

$$\frac{Y(s)}{G_{\text{Tr}}(s)} = K_P(R(s) - Y(s)) \quad \Leftrightarrow \quad Y(s)(1 + K_P G_{\text{Tr}}(s)) = K_P G_{\text{Tr}}(s) R(s)$$

erhält, also

$$G(s) = \frac{Y(s)}{R(s)} = \frac{K_P G_{\text{Tr}}(s)}{1 + K_P G_{\text{Tr}}(s)}.$$

Einsetzen von $G_{\text{Tr}}(s) = \frac{b_0}{s^2}$ ergibt:

$$G(s) = \frac{K_P b_0}{s^2 + K_P b_0}.$$

Diese Übertragungsfunktion hat den Grad $n = 2$, und seine zwei Polstellen sind

$$\lambda_1 = j\sqrt{K_P b_0}, \quad \lambda_2 = -j\sqrt{K_P b_0}.$$

Das System ist also leider nicht stabil, ganz unabhängig davon, wie man den Parameter K_P des Proportional-Reglers wählt. Ein einfacher Proportionalregler ist nicht geeignet, den Traktor zu stabilisieren.

6.6.5 Der PID-Regler

Könnten wir statt des konstanten Parameters K_P für den Regler eine Übertragungsfunktion $K(s) = K_P + K_D s$ wählen, dann bekämen wir für den geschlossenen Kreis die Gesamtübertragungsfunktion

$$G(s) = \frac{K(s)b_0}{s^2 + K(s)b_0} = \frac{K_D b_0 s + K_P b_0}{s^2 + K_D b_0 s + K_P b_0}$$

mit den Polstellen

$$\lambda_{1,2} = -\frac{K_D b_0}{2} \pm j \sqrt{K_P b_0 - \frac{K_D^2 b_0^2}{4}}.$$

Für positive Werte von K_D und K_P wäre der geschlossene Kreis stabil. Ein Regler der Form $K(s) = K_P + K_D s$ wird oft gebraucht, er besteht aus einem Proportionalglied und einem Differenzierglied, und heißt PD-Regler. Eine noch allgemeinere einsetzbare Reglerstruktur erhält man, wenn man noch ein Integrierglied hinzufügt, also eine Reglertransferfunktion der Form $K(s) = K_P + K_I \frac{1}{s} + K_D s$ zulässt. Einen solchen Regler nennt man PID-Regler, und durch geschickte Wahl seiner drei Tuning-Parameter K_P, K_I, K_D kann man die allermeisten Regelungsprobleme für LTI-SISO-Systeme gut lösen.

6.7 Zeitkonstantenform

Wir haben bisher zwei Varianten kennengelernt, die Zähler- und Nennerpolynome einer Übertragungsfunktion zu schreiben. Wir diskutieren hier der Einfachheit halber erst das Nennerpolynom, das Zählerpolynom können wir dann analog behandeln. Zum ersten hatten wir das Nennerpolynom $N(s)$ (das natürlich identisch zum charakteristischen Polynom $p_A(s)$ ist), mit Hilfe seiner reellen Koeffizienten a_0, \dots, a_{n-1} beschrieben, also

$$N(s) = s^n + a_{n-1}s^{n-1} + \dots + a_1 s + a_0.$$

Zum zweiten hatten wir die Polstellenzerlegung genutzt, um das selbe Polynom mit Hilfe der Polstellen λ_i in der Form

$$N(s) = \prod_{i=1}^r (s - \lambda_i)^{n_i}$$

zu schreiben, wobei die Vielfachheit jeder Polstelle λ_i mit n_i bezeichnet wurde, und $\sum_{i=1}^r n_i = n$. Hierbei ist zu beachten, dass die Polstellen auch komplex sein können. Allerdings treten komplexe Polstellen immer nur in komplex-konjugierten Paaren auf. Falls für zwei Indizes i und j die Pole λ_i und λ_j solch ein Paar formen, dann sind n_i und n_j gleich, und es tritt genau n_i mal der Faktor

$$(s - \lambda_i)(s - \bar{\lambda}_i) = s^2 - 2\operatorname{Re}(\lambda_i) s + |\lambda_i|^2$$

auf. Man beachte, dass in diesem Term nur noch reelle Zahlen vorkommen. Um dies zu erreichen, muss man nur die Linearfaktoren jedes komplex-konjugierten Polstellenpaares zu einem Polynom zweiter Ordnung ausmultiplizieren.

Eine dritte Form, die das Nennerpolynom nur mit reellen Zahlen darstellt, und die manchmal anschaulicher als die Polstellenzerlegung ist, nutzt die sogenannten "Zeitkonstanten" $T_i = -\frac{1}{\lambda_i}$ für alle reellen und von Null verschiedenen Eigenwerte $\lambda_i \in \mathbb{R} \setminus \{0\}$. Statt $(s - \lambda_i)$ schreibt man dann einfach den identischen Term

$$s - \lambda_i = \frac{1}{T_i}(T_i s + 1).$$

Man beachte, dass T_i die Dimension der Zeit hat, also z.B. in Sekunden gemessen werden kann. Für komplex-konjugierte Eigenwerte nutzt man eine ähnliche Form, und man schreibt

$$s^2 - 2\operatorname{Re}(\lambda_i) + |\lambda_i|^2 = \frac{1}{T_i^2}(T_i^2 s^2 + 2T_i d_i + 1),$$

mit $T_i = |\lambda_i|^{-1}$ und $d_i = \frac{-\operatorname{Re}(\lambda_i)}{|\lambda_i|}$, wobei T_i wieder die Dimension der Zeit hat, und die sogenannte Dämpfung d_i ohne Einheit ist. Ausserdem kann natürlich auch Null als Polstelle auftreten, wodurch ein Term der Form s^m im Nennerpolynom verbleiben kann. Zusammenfassend kann man das Nennerpolynom also auch wie folgt schreiben:

$$N(s) = s^m \prod_{\lambda_i \text{ reell}} \left[\frac{1}{T_i}(T_i s + 1) \right]^{n_i} \prod_{\lambda_i \text{ komplex}} \left[\frac{1}{T_i^2}(T_i^2 s^2 + 2T_i d_i + 1) \right]^{n_i}.$$

Die gleiche Form kann man für das Zählerpolynom der Übertragungsfunktion erstellen. Man kann dann alle auftretenden Konstanten sammeln, und bekommt dann z.B. im Falle, dass es nur einfache reelle Nullstellen und Polstellen ungleich Null gibt, für die Übertragungsfunktion den Ausdruck

$$G(s) = \frac{Z(s)}{N(s)} = k_s \frac{\prod_{i=1}^q (T_{0,i} s + 1)}{\prod_{i=1}^n (T_i s + 1)}.$$

Der Faktor k_s heißt die statische Verstärkung, und ist nichts anderes als das DC-Gain, das wir zuvor mit der Größe $h(\infty)$ bezeichnet hatten. Setzt man $s = 0$ in die Übertragungsfunktion ein, hat sie genau diesen Wert:

$$G(0) = k_s.$$

Kapitel 7

Frequenzgang und Bode-Diagramm

Wir wissen bereits, dass stabile LTI-SISO Systeme die Eigenschaft haben, dass sie, wenn sie mit einem sinusförmigen Eingangssignal angeregt werden, nach einer kurzen Einschwingphase ein ebenso sinusförmiges Ausgangssignal ausgeben, das in Amplitude und Phase verschoben ist. Mit Hilfe der Übertragungsfunktion $G(s)$ kann man diesen Zusammenhang sehr einfach ausdrücken, denn für ein unendlich lang wirkendes Eingangssignal

$$u(t) = U_0 e^{j\omega t} \quad (7.1)$$

mit Amplitude $U_0 \in \mathbb{R}_+$ und Frequenz $\omega \in \mathbb{R}_+$ ergibt sich das folgende Ausgangssignal:

$$y(t) = G(j\omega) U_0 e^{j\omega t}. \quad (7.2)$$

Die Verwendung der komplexen Zahlen ist hier sehr hilfreich und wir werden in diesem Kapitel sehr intensiven Gebrauch von ihnen machen, aber für eine physikalische Realisierung müsste man sich statt (7.1) auf reelle Eingangssignale z.B. der Form

$$u(t) = U_0 \frac{1}{2} (e^{j\omega t} + e^{-j\omega t})$$

beschränken, also nur den Realteil des vorherigen Signals, und würde dann statt (7.2) folgendes Ausgangssignal erhalten:

$$y(t) = \frac{1}{2} (G(j\omega) U_0 e^{j\omega t} + G(-j\omega) U_0 e^{-j\omega t}).$$

Glücklicherweise stellt dieses Signal wieder einfach nur den Realteil des Ausgangssignals (7.2) dar, da $G(-j\omega) = \overline{G(j\omega)}$ gilt. Diese Eigenschaft der Übertragungsfunktion folgt aus ihrer Definition mit Hilfe der Laplace-Transformation und der Tatsache, dass die Impulsantwort $g(t)$ eine reelle Funktion ist:

$$\overline{G(j\omega)} = \int_0^\infty \overline{e^{-j\omega t} g(t)} dt = \int_0^\infty e^{j\omega t} g(t) dt = G(-j\omega).$$

Wir können uns deshalb in der weiteren Analyse auf den einfacheren Ausdruck (7.2) beschränken, und behalten im Hinterkopf, dass man sich für die physikalische Realisierung nur auf den Realteil (oder alternativ den Imaginärteil) beschränken kann. Der entscheidende Grund für die Popularität der Frequenzanalyse ist, dass man diese sehr einfach auch experimentell erstellen kann, z.B. mit Hilfe eines Oszilloskops: man regt das System nacheinander mit sinusförmigen Eingangssignalen verschiedener Frequenzen ω an, und misst jeweils die Amplitudenverstärkung

$$M(\omega) := |G(j\omega)|$$

und die Phasenverschiebung

$$\phi(\omega) := \arg G(j\omega),$$

die positiv ist, wenn das Ausgangssignal dem Eingangssignal vorausschneidet, und negativ, wenn es dem Eingangssignal nachschneidet.

7.1 Bode-Diagramm

Eine der vielen entscheidenden Ideen, die auf Hendrik Bode zurückgehen, der in der ersten Hälfte des 20sten Jahrhunderts gemeinsam mit anderen Pionieren der Regelungs- und Nachrichtentechnik wie Harry Nyquist oder Harold Black an den US-amerikanischen Bell-Labs gearbeitet hat, ist die folgende: wenn man sowohl die Frequenzen als auch die Amplitudenverstärkung logarithmisch aufträgt, dann wird die Kurve der Amplitudenverstärkungsfunktion $M(\omega)$ zu großen Teilen fast gerade und damit besonders einfach interpretierbar. Dies liegt an zwei Tatsachen. Zum ersten geht durch die Logarithmierung von $|G(s)|$ das Produkt der aus den Nullstellen und Polstellen resultierenden Terme in eine Summe über, denn

$$\begin{aligned} \log |G(s)| &= \log \left(b_q \frac{\prod_{i=1}^{r_0} (s - \lambda_{0,i})^{n_{0,i}}}{\prod_{i=1}^r (s - \lambda_i)^{n_i}} \right) \\ &= \log |b_q| + \sum_{i=1}^{r_0} n_{0,i} \log |s - \lambda_{0,i}| - \sum_{i=1}^r n_i \log |s - \lambda_i|. \end{aligned}$$

Man nimmt als Basis des Logarithmus meist die Basis 10, aber verwendet zur Messung des Logarithmus oft die historisch entstandene Einheit Dezibel [dB], die jede Dekade noch einmal in 20 Untereinheiten teilt, also

$$|G(s)|_{\text{dB}} = 20 \log_{10} |G(s)|.$$

Mathematisch gesehen könnte man ein Dezibel alternativ auch einfach als den Logarithmus zur Basis $10^{\frac{1}{20}} = 1.12202$ definieren. Welche Basis man für die Berechnung des Logarithmus verwendet, ist für die grundlegenden Zusammenhänge nicht entscheidend.

Zum zweiten wird jeder der Terme $|s - \lambda_i|$ durch die doppelte Logarithmierung besonders einfach, denn es gilt mit der Variablen $\omega_{\log} := \log \omega$ für jede Null- oder Polstelle:

$$\log |j\omega - \lambda_i| = \log |j10^{\omega_{\log}} - \lambda_i| = \log \sqrt{(10^{\omega_{\log}} - \text{Im}\lambda_i)^2 + (\text{Re}\lambda_i)^2}.$$

Für $\omega \gg |\lambda_i|$ dominiert der erste Term unter der Wurzel und es gilt die extrem einfache Approximationsformel

$$\log |j\omega - \lambda_i| \approx \omega_{\log},$$

also ein linearer Zusammenhang zwischen der logarithmischen Amplitudenverstärkung und der logarithmischen Frequenz. Für sehr kleine $\omega \ll |\lambda_i|$ dominieren hingegen die anderen Terme, und es gilt die noch einfachere, weil konstante Approximation

$$\log |j\omega - \lambda_i| \approx \log |\lambda_i|.$$

Im Zwischenbereich, also für $\omega \approx |\lambda_i|$, tritt ein Knick in der Kurve auf. Dieser kann im Falle einer reellen Null- oder Polstelle sehr einfach beschrieben werden und stellt eine rundgezogene Ecke zwischen den beiden Approximationstangenten dar. Im Falle einer komplexen Null- oder Polstelle ist es am sinnvollsten, direkt das komplex-konjugierte Paar zu betrachten, also

$$\log (|j\omega - \lambda_i| |j\omega - \bar{\lambda}_i|) = \log \sqrt{(10^{\omega_{\log}} - \operatorname{Im}\lambda_i)^2 + (\operatorname{Re}\lambda_i)^2} + \log \sqrt{(10^{\omega_{\log}} + \operatorname{Im}\lambda_i)^2 + (\operatorname{Re}\lambda_i)^2}. \quad (7.3)$$

Bei sehr großen ω tendiert dieser Term gegen $2\omega_{\log}$, und für kleine gegen den konstanten Wert $2 \log |\lambda_i|$, so wie man es für zwei Null- oder Polstellen mit gleichem Betrag erwarten würde. Im Zwischenbereich kann jedoch eine Amplitudenüberhöhung auftreten, die umso größer ist, je größer der Imaginärteil relativ zum Realteil ist, oder äquivalent, je kleiner die Dämpfung $d_i = \frac{-\operatorname{Re}\lambda_i}{|\lambda_i|}$ ist.

Es gibt noch einen anderen interessanten Spezialfall: für die Null- oder Polstelle $\lambda_i = 0$ ergibt sich die Identität

$$\log |j\omega - 0| = \omega_{\log},$$

die dem Magnituden-Diagramm eine einfache Gerade der Steigung eins bzw. minus eins hinzufügt, die sich im Gegensatz zu allen anderen Beiträgen auch in den Bereich der unendlich kleinen Frequenzen fortsetzt.

Zu einem Bode-Diagramm gehören immer zwei Kurven: zum ersten die doppelt-logarithmisch aufgetragene Amplitudenverstärkungsfunktion

$$M_{\log-\log}(\omega_{\log}) := \log |G(j10^{\omega_{\log}})|,$$

die wir auch den *Amplitudengang* nennen, und zum anderen die Phasenverschiebungsfunktion mit logarithmisch aufgetragenen Frequenzen:

$$\phi_{\log}(\omega_{\log}) := \arg G(j10^{\omega_{\log}}),$$

die wir den *Phasengang* nennen. Beim Phasengang addieren sich die Einzelphasen der Null- und Polstellen gemäß der Formel

$$\arg G(s) = \sum_{i=1}^{r_0} n_{0,i} \arg(s - \lambda_{0,i}) - \sum_{i=1}^r n_i \arg(s - \lambda_i),$$

mit $s = j\omega$. Die Phasenverschiebung nimmt damit für $\omega \rightarrow 0$ einen konstanten Wert an, und für sehr große ω den Wert $-\frac{\pi}{2}(n - q)$ wobei $(n - q)$ der relative Grad des

Systems ist, mit $n = \sum_{i=1}^r n_i$ und $n_0 = \sum_{i=1}^{r_0} n_{0,i}$. Bei langsam ansteigenden Werten für ω_{\log} springt die Phasenverschiebung im Übergangsbereich jeder Null- oder Polstelle auf eine abgerundete Weise jeweils um den Wert $\frac{\pi}{2}$ (=90 Grad) nach oben (Nullstelle in linker Halbebene oder Polstelle in rechter Halbebene) oder nach unten (Polstelle in linker Halbebene oder Nullstelle in rechter Halbebene). Bei komplex-konjugierten Null- oder Polstellenpaaren tritt jeweils ein doppelt so großer Phasensprung auf, um plus oder minus π (= 180 Grad).

Die meisten technisch interessanten Systeme sind stabil, so dass alle Polstellen in der linken Halbebene oder in der Null liegen. Liegen nun auch alle Nullstellen in der linken Halbebene, dann spricht man von einem *minimalphasigen System* und man kann aus der Steigung im Amplitudenplot des Bode-Diagramms direkt auch die Phasenverschiebung ablesen: ist die Steigung in einem der geraden Stücke im doppelt-logarithmischen Amplitudenplot gegeben durch $m \in \mathbb{Z}$, dann ist die Phasenverschiebung in diesem Bereich gegeben durch $m \frac{\pi}{2}$ (= m -Mal 90 Grad). Man beachte, dass eine Steigung von eins im doppelt-logarithmischen Plot bei Verwendung der Einheit Dezibel im Magnitudenplot einer Steigung von 20 dB pro Frequenzdekade entspricht.

Da es aber auch stabile und technisch interessante Systeme gibt, die Nullstellen in der rechten Halbebene haben, gilt der beschriebene Zusammenhang zwischen Amplitudengang und Phasengang in einigen wenigen wichtigen Fällen nicht. Eine Nullstelle in der rechten Halbebene hat im Vergleich zu einer entsprechenden Nullstelle in der linken Halbebene genau den gleichen Einfluss auf den Amplitudengang, aber ihr Phasenbeitrag startet bei 180 Grad für sehr kleine ω und endet für große ω bei 90 Grad. Für kleine Frequenzen hat die Phasenverschiebung damit einen im Vergleich zum entsprechenden minimalphasigen System einen um 180 Grad höheren Wert, und man spricht dann von einem *nicht-minimalphasigen System*. Die Regelung nicht-minimalphasiger Systeme ist etwas schwieriger als die Regelung minimalphasiger Systeme.

7.1.1 Bode-Diagramm des PT₁-Glieds

Wir hatten als wichtiges und immer wieder auftretendes System bereits das PT₁-Glied kennengelernt, dass durch die Übertragungsfunktion

$$G(s) = \frac{k_s}{Ts + 1} = \frac{k_s \omega_0}{s + \omega_0}$$

beschrieben ist, mit positiven reellen k_s und T (siehe Abb. 7.1). Seine Zeitkonstante T tritt in der Pol-Nullstellenform als Knickfrequenz $\omega_0 = \frac{1}{T}$ in Erscheinung. Diese Frequenz steht zur einzigen reellen Polstelle λ_1 des PT₁-Gliedes im Zusammenhang $\lambda_1 = -\omega_0$. Es ergibt sich der Amplitudengang

$$|G(j\omega)| = \frac{k_s \omega_0}{|j\omega + \omega_0|} = \frac{k_s \omega_0}{\sqrt{\omega^2 + \omega_0^2}}$$

und wegen

$$\frac{k_s \omega_0}{j\omega + \omega_0} = \frac{k_s \omega_0}{\omega^2 + \omega_0^2} (-j\omega + \omega_0)$$

der Phasengang

$$\phi(\omega) = \arctan \frac{-\omega}{\omega_0},$$

der für kleine ω gleich Null ist, und für große ω gleich $-\frac{\pi}{2}$. Das Bode-Diagramm eines PT_1 -Gliedes ist in Abbildung 7.1 links gezeigt.

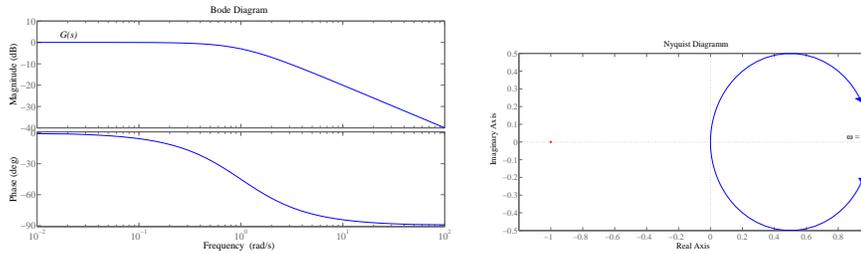


Abbildung 7.1: Bode- und Nyquist-Diagramm eines PT_1 -Gliedes mit $k_s = 1$ und $T = 1$ s, also $G(s) = \frac{1}{1+s}$.

7.2 Nyquist-Diagramm, oder Ortskurve

Eine interessante Variante, die Information des Bode-Diagramms darzustellen, liegt darin, die komplexen Werte der Funktion $G(j\omega)$ als Kurve in der komplexen Ebene zu zeichnen, für $\omega \in [0, \infty)$. Die Kurve dieses sogenannten *Nyquist-Diagramms*, das auch *Ortskurve* genannt wird, beginnt für den Wert $\omega = 0$ bei der Zahl $G(0)$, die immer auf der reellen Achse liegt, und deren Wert gleich der statischen Verstärkung des Systems ist. Die Kurve geht dann durch die typischerweise komplexen Werte $G(j\omega)$, und nähert sich für $\omega \rightarrow \infty$ wieder einer reellen Zahl, die gleich ist mit dem Koeffizienten b_n in der Eingangs-Ausgangsdifferentialgleichung, und die in der Zustandsraumdarstellung durch die 1×1 -Matrix D gegeben war, und die wir als *Durchgriff* bezeichnen, und die nur dann von Null verschieden ist, wenn der relative Grad des Systems Null ist. Viele physikalische Systeme haben keinen Durchgriff, und die Ortskurve endet in der Null.

Man beachte, dass das Bode-Diagramm zwar alle Information des Nyquist-Diagramms enthält, aber umgekehrt man das Bode-Diagramm nicht zeichnen kann, wenn man nur das Nyquist-Diagramm kennt, da die Zuordnung zu den Frequenzen ω verlorengegangen ist. Das Nyquist-Diagramm eines PT_1 -Gliedes ist in Abb. 7.1 rechts gezeigt.

7.3 Zusammenhang mit Übertragungsfunktion in der komplexen Ebene

Die Übertragungsfunktion $G(s)$ ist als Funktion $G: \mathbb{C} \rightarrow \mathbb{C}$ nicht leicht visualisierbar, da man einen vier-dimensionalen Plot dafür bräuchte, aber wir Menschen uns nur Ob-

jekte in drei Dimensionen vorstellen können. Eine interessante Darstellung der Übertragungsfunktion beschränkt sich auf ihre Amplitude $|G(s)|$, die eine reelle Zahl ist, und damit eine Abbildung von der komplexen Ebene in die reellen Zahlen darstellt, die man sich als Gebirge vorstellen kann. Die Gipfel dieses Gebirges werden durch die unendlich hohen Spitzen an den Polstellen der Übertragungsfunktion gebildet. Die Nullstellen sind die tiefsten Punkte des Gebirges, und liegen auf der Höhe "Normalnull". Wenn man allerdings den Logarithmus der Amplitude betrachtet, $\log |G(s)|$, dann treten die Nullstellen als unendlich tiefe Erdlöcher auf, die exakt die gleiche Form wie die Spitzen der Polstellen haben, nur nach unten statt nach oben gerichtet. Eine Visualisierung dieses "Pol-Nullstellen-Gebirges" für ein komplexes System mit vier Polstellen und drei Nullstellen ist in Abbildung 7.2 gezeigt.

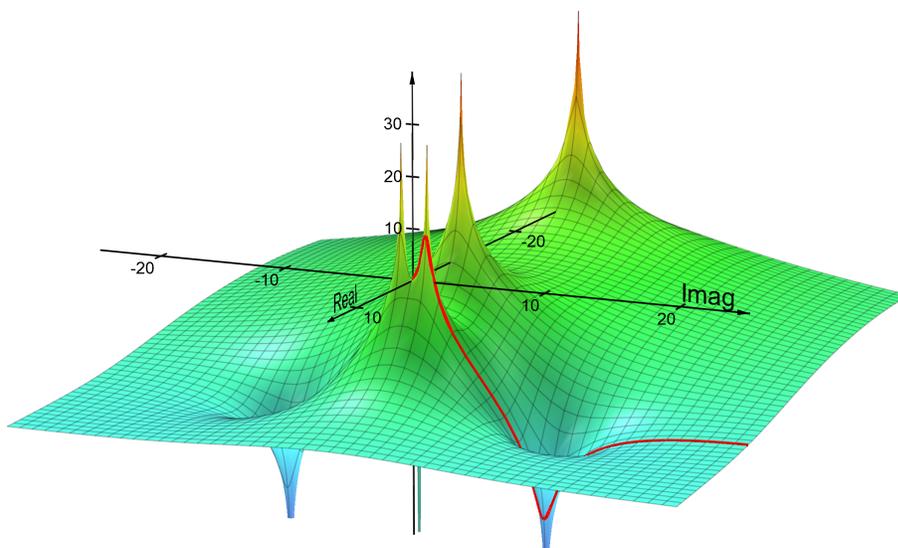


Abbildung 7.2: Pol-Nullstellen-Gebirge für die komplexe Übertragungsfunktion $G(s) = \frac{(s+1)(1+0.01s+0.01s^2)}{(1+0.5s+s^2)(1+0.11s)(1+0.03s)}$. Man beachte die rote Linie, die im Bode-Diagramm in Abbildung 7.3 wieder auftaucht.

In dieser Darstellung findet man die Werte des Amplitudenganges $\log |G(j\omega)|$ auf der positiven imaginären Achse, wie einen Querschnitt durch das Pol-Nullstellen-Gebirge (rote Kurve in Abbildung 7.2). Aus dieser Darstellung wird deutlich, dass komplexe Polstellen, die sehr nahe an der imaginären Achse liegen, zu einer deutlichen Erhöhung des Amplitudenganges führen, der sogenannten Resonanzüberhöhung, die ebenso in Formel (7.3) sichtbar war: der Term

$$-\log \sqrt{(\omega - \text{Im}\lambda_i)^2 + (\text{Re}\lambda_i)^2}$$

nimmt bei $\omega = \text{Im}\lambda_i$ ein Maximum an, das umso ausgeprägter in Erscheinung tritt, je kleiner der Realteil $\text{Re}\lambda_i$ ist. Umgekehrt treten Nullstellen, die dicht bei der imaginären

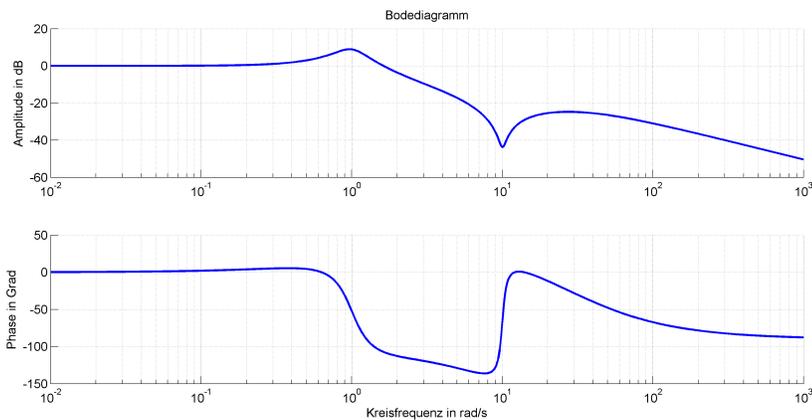


Abbildung 7.3: Bode-Diagramm des Systems aus Abbildung 7.2.

Achse liegen, als starke Vertiefungen des Amplitudengangs bei bestimmten Frequenzen auf. Das komplette Bode-Diagramm des betrachteten Systems ist in Abbildung 7.3 dargestellt.

7.4 Bode-Diagramm und Verhalten des PT₂-Gliedes

Ein PT₂-Glied wird in Zeitkonstantenform durch die Übertragungsfunktion

$$G(s) = \frac{k_s}{T^2 s^2 + 2Tds + 1}$$

beschrieben, mit positiver reeller Zeitkonstante T und reeller Dämpfung d . Die beiden Polstellen sind gegeben durch

$$\lambda_{1,2} = \frac{1}{T} \left(-d \pm \sqrt{d^2 - 1} \right).$$

Für den Fall, dass die Dämpfung größer ist als 1, sind beide Polstellen reell, und die Übertragungsfunktion zerfällt in zwei miteinander multiplizierte Terme mit jeweils nur einer reellen Polstelle; man kann dieses Glied auch als Hintereinanderschaltung zweier PT₂-Glieder auffassen. Das gleiche gilt für den Fall einer stark negativen Dämpfung $d < -1$, der zu zwei instabilen reellen Polstellen führt. Die einzigen Werte für d , bei denen die Polstellen nicht reell sind, und bei denen das PT₂-Glied seine ganz eigenen speziellen Eigenschaften zeigt, sind die Werte $d \in [-1, 1]$. Die Polstellen haben dann die Werte

$$\lambda_{1,2} = \frac{1}{T} \left(-d \pm j\sqrt{1 - d^2} \right).$$

Das Winkelargument jeder Polstelle hängt dabei nur von d ab, und es gilt

$$\arg \lambda_1 = \arccos(-d).$$

Auch die relative Resonanzüberhöhung und Form des Bode-Diagramms hängt nur von d ab, denn es gilt mit $\omega_0 = \frac{1}{T}$

$$G(j\omega) = \frac{k_s}{1 - \left(\frac{\omega}{\omega_0}\right)^2 + j2d\left(\frac{\omega}{\omega_0}\right)}.$$

An der Stelle $\omega = 0$ gilt einfach nur $G(0) = k_s$, und an der Stelle $\omega = \omega_0$ gilt

$$G(j\omega_0) = -j\frac{k_s}{2d},$$

was für kleine d beinahe dem Maximum des Amplitudengangs entspricht, das für $|d| < \frac{1}{\sqrt{2}}$ bei $\omega^* = \omega_0\sqrt{1 - 2d^2}$ angenommen wird. Für höhere Absolutwerte der Dämpfung gibt es keine Resonanzüberhöhung mehr. Die Bode-Diagramme von PT_2 -Gliedern mit Zeitkonstante $T = 1$ und unterschiedlich großen Dämpfungswerten sind in Abbildung 7.4 gezeigt.

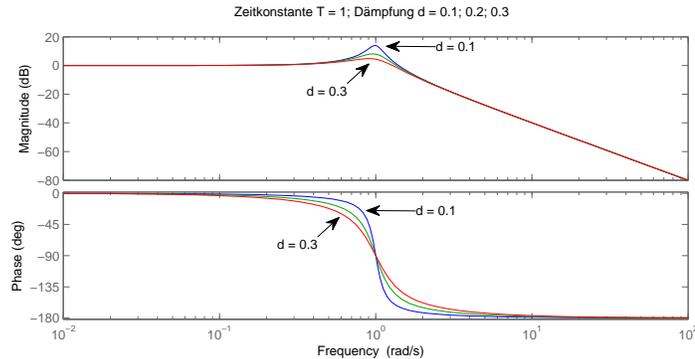


Abbildung 7.4: Bode-Plot eines PT_2 -Gliedes mit $k_s = 1$, $T = 1$ s, und schwacher Dämpfung $d = 0.1, 0.2, 0.3$ ($G_{0.1}(s) = \frac{1}{1+0.2s+s^2}$, $G_{0.2}(s) = \frac{1}{1+0.4s+s^2}$, $G_{0.3}(s) = \frac{1}{1+0.6s+s^2}$).

7.4.1 Polstellen und Zeitverhalten des PT_2 -Gliedes

Das Verhalten des PT_2 -Gliedes kann man auf viele verschiedene Weisen visualisieren: als Pol-Nullstellen-Diagramm, als Bode-Diagramm, als Nyquist-Plot, oder als Zeitverlauf der Sprungantwort. Um ein besseres Gefühl für den Zusammenhang der Pol-Orte mit dem Zeitverlauf der Sprungantwort zu bekommen, illustrieren wir in den Abbildungen 7.5, 7.6 und 7.7 die Veränderung der Sprungantwort bei Radial-, Horizontal- und Vertikalverschiebungen der Pole.

Bei einer Radialverschiebung der Pole, die man durch Veränderung von T bzw. ω_0 aber konstantes d erhält, ändert man eigentlich nur die Geschwindigkeit der Bewegung bei einem ansonsten identischen Zeitverlauf, wie Abb. 7.5 illustriert. Insbesondere bleibt der Overshoot identisch.

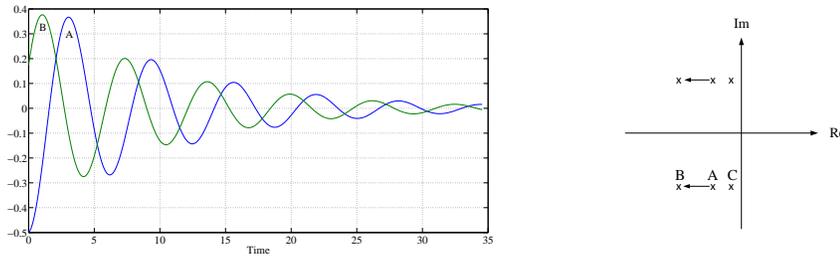


Abbildung 7.5: Zeitverhalten für verschiedene Zeitkonstante T bei gleicher Dämpfung d , und die zugehörigen Polstellen (Achtung: das Zeitverhalten ist hier nicht 100% korrekt dargestellt).

Bei einer Linksverschiebung der Pole, die man durch Erhöhung der Dämpfung d und gleichzeitige Erhöhung der Frequenz ω_0 erhält, so dass $\omega_0 \sqrt{1-d^2}$ konstant bleibt, bleibt die Frequenz der Oszillationen identisch, während die Höhe des Overshoots abnimmt. Dies ist in Abb. 7.6 illustriert.

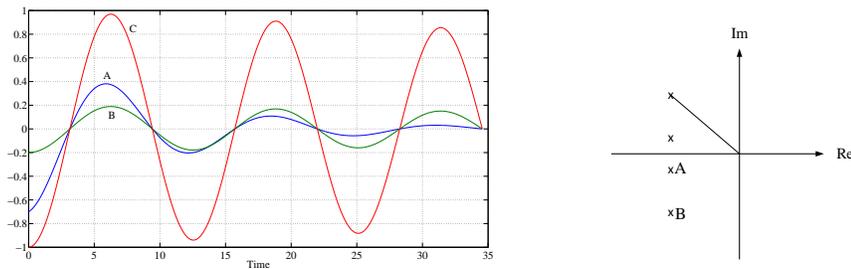


Abbildung 7.6: Zeitverhalten für verschieden starke Dämpfung bei gleicher Grundfrequenz, und die zugehörigen Polstellen.

Bei einer Verschiebung der Pole nach oben und unten, weg von der reellen Achse – was man dadurch erreicht, dass man ω_0 erhöht und d vermindert, so dass das Produkt $d\omega_0$ konstant bleibt – behält man die gleiche Einhüllende der Sprungantwort, während sich die Frequenz der Oszillationen erhöht (siehe Abb. 7.7).

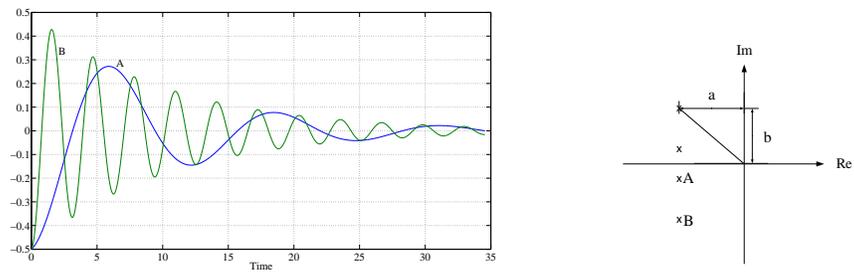


Abbildung 7.7: Zeitverhalten für verschieden große Grundfrequenz bei gleichem Realteil, der die Einhüllende bestimmt, und die zugehörigen Polstellen.

Kapitel 8

Regelungssysteme

In dem folgenden zweiten Teil der Vorlesung betrachten wir den Entwurf von Regelungssystemen. Als gegeben nehmen wir dabei ein Systemmodell $G(s)$ an, und gesucht ist ein Regler $K(s)$, so dass der resultierende geschlossene Regelkreis gewünschte Eigenschaften hat. Wichtige gewünschte Eigenschaften des geschlossenen Kreises sind:

- **Stabilität:** der geschlossene Kreis soll auf jeden Fall stabil sein. Dies ist ein nicht-trivialer Wunsch, denn selbst für ein stabiles System $G(s)$ kann der geschlossene Kreis instabil werden.
- **Sollwertfolge:** der Regelkreis soll den gewünschten Referenzwerten folgen. Man unterscheidet
 - **Festwertregelung:** der Referenzwert bleibt konstant oder ändert sich sehr selten, wie z.B. bei der Zimmertemperaturregelung.
 - **Folgeregelung:** der Output soll einer zeitlich variierenden Referenztrajektorie folgen, wie z.B. bei der Flugregelung eines von einem Piloten gesteuerten Flugzeuges, oder beim Übertragen von Audiosignalen durch einen Verstärker.
- **Störkompensation:** das geregelte System soll Störungen unterdrücken können, z.B. bei einem Temperaturregler den Einfluss der langsam variierenden Umgebungstemperatur (Störgröße) auf die Raumtemperatur (Regelgröße) kompensieren.
- **Robustheit:** der Regelkreis soll die gewünschten Eigenschaften, insbesondere die Stabilität, auch behalten, wenn das wirkliche System anders ist als das Systemmodell, das zum Entwurf verwendet wurde.

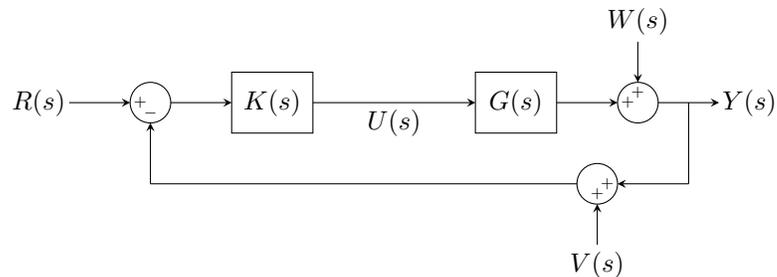


Abbildung 8.1: Blockdiagramm des Standardregelkreises mit Störungen $W(s)$ und Messrauschen $V(s)$.

8.1 Der Standardregelkreis und Gütekriterien

Wir betrachten ein System $G(s)$ mit Steuerungseingang $U(s)$ und Ausgang $Y(s)$. Der Ausgang wird durch eine Störung $W(s)$ modifiziert, die wir nicht kennen, es gilt also

$$Y(s) = G(s)U(s) + W(s).$$

Das Ausgangssignal selbst können wir nicht ohne Fehler messen, sondern für die Regelung steht nur das um das Messrauschen $V(s)$ modifizierte Signal

$$Y'(s) = Y(s) + V(s)$$

zur Verfügung. Im Standardregelkreis, dessen Blockdiagramm in Abbildung 8.1 gezeigt ist, betrachten wir die Differenz zwischen dem Referenzsignal $R(s)$ und dem gemessenen Output $Y'(s)$, die wir den Regelfehler $E(s)$ nennen:

$$E(s) = R(s) - Y'(s).$$

Dieses Signal formt den Eingang des Reglers $K(s)$, der daraus den Steuerungsinpud $U(s)$ generiert, also

$$U(s) = K(s)E(s).$$

Uns interessieren vor allem drei Übertragungsfunktionen:

- Führungsverhalten: wie reagiert $Y(s)$ auf die Referenztrajektorie $R(s)$?
- Störverhalten: wie reagiert $Y(s)$ auf Störungen $W(s)$?
- Messrauschempfindlichkeit: wie reagiert $Y(s)$ auf Messrauschen $V(s)$?

Um diese drei Übertragungsfunktionen zu erhalten, stellen wir unter Elimination von $U(s)$, $E(s)$ und $Y'(s)$ durch $U(s) = K(s)(R(s) - Y(s) - V(s))$ die Gleichung des geschlossenen Kreises auf, die wie folgt lautet:

$$Y(s) = W(s) + G(s)K(s)(R(s) - Y(s) - V(s))$$

und die wir algebraisch so umformen können, dass $Y(s)$ allein auf der linken Seite erscheint:

$$Y(s) = \underbrace{\frac{G(s)K(s)}{1+G(s)K(s)}}_{\text{Führungsverhalten}} R(s) + \underbrace{\frac{1}{1+G(s)K(s)}}_{\text{Störverhalten}} W(s) + \underbrace{\frac{-G(s)K(s)}{1+G(s)K(s)}}_{\text{Messrauschempfindlichkeit}} V(s).$$

Wir definieren uns zur Abkürzung drei Übertragungsfunktionen, die wir noch häufig verwenden werden: die *offene Kette* (engl: *open-loop system*)

$$G_0(s) = G(s)K(s),$$

die *Sensitivitätsfunktion*

$$S(s) = \frac{1}{1+G_0(s)},$$

die unter anderem das Störverhalten beschreibt, und die *komplementäre Sensitivitätsfunktion*

$$T(s) = \frac{G_0(s)}{1+G_0(s)},$$

die unter anderem das Führungsverhalten als auch die (negative) Messrauschempfindlichkeit beschreibt. Man beachte, dass für alle Werte $s \in \mathbb{C}$ gilt, dass

$$T(s) + S(s) = 1.$$

Dies gilt insbesondere auch für sinusförmige Anregungen des Systems, die mit $s = j\omega$ durch $T(j\omega)$ und $S(j\omega)$ beschrieben werden.

Da man idealerweise das Störverhalten klein haben möchte, also $|S(j\omega)| \ll 1$, aber andererseits auch die Messrauschempfindlichkeit $|T(j\omega)| \ll 1$, wird die Unmöglichkeit, beide Terme gleichzeitig sehr klein zu bekommen, manchmal auch als "Fundamentaldilemma der Regelungstechnik" bezeichnet. Glücklicherweise sind die Störungen, die der Regler kompensieren soll, meist niederfrequent, während das Messrauschen meist hochfrequent ist. Man strebt in der Praxis deshalb danach, dass für hohe Frequenzen ω gilt, dass $|T(j\omega)| \ll 1$, und für niedrige Frequenzen ω , dass $|S(j\omega)| \ll 1$, wie in Abb. 8.2 illustriert.

Die Wahl der Übergangsfrequenz, bei der dieser Umschlag von Störunterdrückung zu Messrauschunterdrückung stattfindet, spielt eine entscheidende Rolle beim Reglerentwurf, und beeinflusst auch das Führungsverhalten, das ebenso wie die Messrauschempfindlichkeit durch die Funktion $T(s)$ beschrieben wird. Man wird einem Referenzsignal von sehr hoher Frequenz nicht sehr gut folgen können, wenn man nicht gleichzeitig eine erhöhte Messrauschempfindlichkeit in diesem Frequenzbereich zulässt. Man kann nicht schneller regeln, als man messen kann. Die Übergangsfrequenz ω_B , unterhalb derer die Führungsübertragungsfunktion $T(j\omega)$ nahe bei eins ist, nennt man auch *Bandbreite* (engl. *bandwidth*). Es soll also gelten, dass $T(j\omega) \approx 1$ für alle $\omega \in [0, \omega_B]$. Die Bandbreite eines geregelten Systems ist oft exakt definiert als der Wert, bei dem die Amplitude der Führungsübertragungsfunktion $T(j\omega)$ auf -3 dB unter die statische Verstärkung abgefallen ist, also

$$|T(j\omega_B)|_{\text{dB}} = |T(0)|_{\text{dB}} - 3\text{dB}.$$

Die Definition der Bandbreite ist im Bode-Diagramm in Abb. 8.3 illustriert.

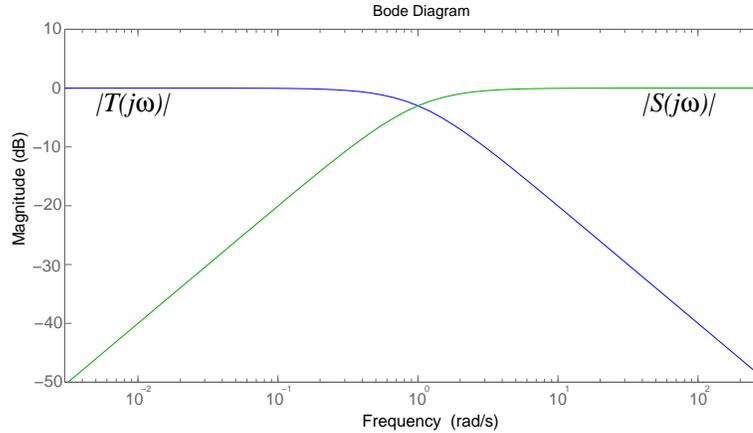


Abbildung 8.2: Illustration und Lösung des Fundamentaldilemmas $T(s) + S(s) = 1$ (am Beispiel $T(s) = \frac{1}{1+s}$ und $S(s) = \frac{s}{1+s}$).

8.2 Gütekriterien im Zeitbereich

Gütekriterien für ein geregeltes System (also die geschlossene Regelschleife) kann man auch im Zeitbereich definieren, wie wir das zuvor bereits in Abschnitt 5.4 getan haben. Man betrachtet dafür die Antwort $h_{\text{Führung}}(t)$ des Outputs $y(t)$ des geschlossenen Kreises auf einen Einheitssprung im Referenzwert $r(t)$, also die Laplacetransformation des Produktes der Übertragungsfunktion $\frac{Y(s)}{R(s)} = T(s)$ mit dem Einheitssprung $\frac{1}{s}$:

$$h_{\text{Führung}}(t) \circ \bullet \frac{T(s)}{s}.$$

Charakteristische Größen sind dabei die Anstiegszeit T_{rise} (rise time), die man gerne kurz haben möchte, und die in engem Zusammenhang mit dem Kehrwert der Bandbreite steht, die Überschwinghöhe Δh und Überschwingzeit, die Abklingzeit (settling time) T_{settling} , sowie eine möglicherweise nicht verschwindende bleibende Regelabweichung, gegeben durch $h(\infty) - 1$.

Neben dem Führungssprung, also der Antwort des Systems auf einen Sprung des Referenzeingangs, kann man auch einen ‘‘Störungssprung’’ $h_{\text{Störung}}(t)$, also die Antwort des Systems auf einen Sprung in der Störgröße betrachten. Aufgrund der Relation $S(s) = 1 - T(s)$ gilt

$$h_{\text{Störung}}(t) \circ \bullet \frac{S(s)}{s} = \frac{1}{s} - \frac{T(s)}{s}.$$

Zusammen mit der Linearität der Laplacetransformation, und der Transformierten des Einheitssprungs $\sigma(t) \circ \bullet \frac{1}{s}$ gilt also

$$h_{\text{Störung}}(t) = \sigma(t) - h_{\text{Führung}}(t).$$

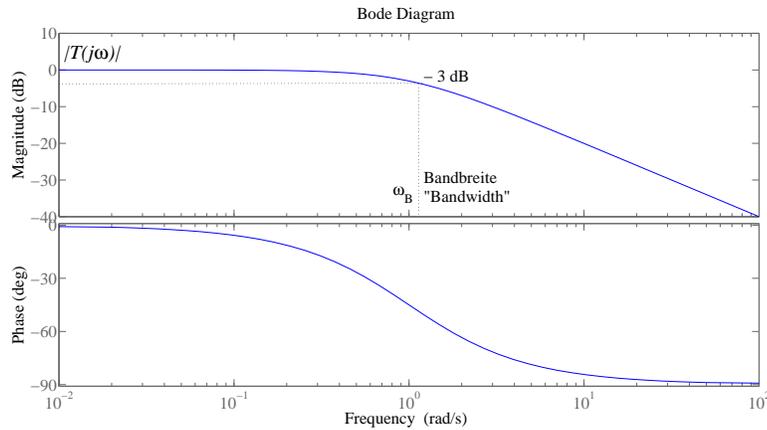


Abbildung 8.3: Bandbreite der Übertragungsfunktion $T(s)$ des geschlossenen Kreises (illustriert für $T(s) = \frac{1}{1+s}$).

Damit erbt das Störverhalten die gleichen Charakteristiken wie das Führungsverhalten. Eine plötzliche auftretende sprungförmige Störung wird z.B. nach der selben Abklingzeit T_{settling} weggeregelt sein, in der das System einem Führungssprung folgt.

Kapitel 9

Stabilität von Regelungssystemen

In diesem Kapitel untersuchen wir die Stabilität des geschlossenen Kreises. Dafür nehmen wir an, dass die offene Kette durch

$$G_0(s) = \frac{Z_0(s)}{N_0(s)}$$

gegeben ist, mit einem Zählerpolynom

$$Z_0(s) = b_q s^q + \dots + b_0 = b_q \prod_{i=1}^q (s - s_{0i})$$

und Nennerpolynom

$$N_0(s) = a_n s^n + \dots + a_0 = a_n \prod_{i=1}^n (s - s_i),$$

Die q Nullstellen von $G_0(s)$ sind je nach Mehrfachheit mehrfach gezählt und gegeben durch s_{01}, \dots, s_{0q} und die n Polstellen durch s_1, \dots, s_n , wobei $n \geq q$.

Der geschlossene Kreis wird durch die Sensitivitätsfunktion $S(s)$ und komplementäre Sensitivitätsfunktion $T(s)$ beschrieben, die folgendermassen dargestellt werden können:

$$S(s) = \frac{1}{1 + G_0(s)} = \frac{N_0(s)}{N_0(s) + Z_0(s)}$$

und

$$T(s) = \frac{G_0(s)}{1 + G_0(s)} = \frac{Z_0(s)}{N_0(s) + Z_0(s)}.$$

Das Nennerpolynom beider Übertragungsfunktionen ist gleich, und dieses bestimmt über die Polstellen des geschlossenen Kreises, also über Stabilität oder Instabilität des geregelten Systems. Wir definieren das Nennerpolynom der Kürze halber als

$$N_{\text{CL}}(s) = N_0(s) + Z_0(s),$$

und in Koeffizientenform ist es gegeben durch

$$N_{\text{CL}}(s) = \underbrace{a_n}_{=:c_n} s^n + \dots + \underbrace{a_{q+1}}_{=:c_{q+1}} s^{q+1} + \underbrace{(a_q + b_q)}_{=:c_q} s^q + \dots + \underbrace{(a_0 + b_0)}_{=:c_0}.$$

Die Polstellenzerlegung dieses Polynoms ist zunächst unbekannt, aber wir wissen, dass es prinzipiell auch in der Form

$$N_{\text{CL}}(s) = c_n \prod_{i=1}^n (s - \bar{s}_i)$$

geschrieben werden kann, wobei $\bar{s}_1, \dots, \bar{s}_n$ die n Polstellen des geschlossenen Kreises sind. Der Regelkreis ist nun genau dann stabil, wenn die Realteile aller \bar{s}_i negativ sind. Wie können wir dies testen?

9.1 Stabilitätstest mit Hilfe von Polynomkoeffizienten

Wenn die Koeffizienten des Nennerpolynoms bekannt sind, also die Zahlen c_n, \dots, c_0 aus

$$N_{\text{CL}}(s) = c_n s^n + \dots + c_0$$

bekannt sind, kann man auf verschiedene Weisen bestimmen, ob alle Nullstellen negativen Realteil haben. Die erste und heutzutage gebräuchlichste Variante ist, ein numerisches Tool wie z.B. MATLAB's Funktion `roots(c)` zu Hilfe zu nehmen, und die Realteile der ausgegebenen Wurzeln des Polynoms zu checken.

Eine andere, und historisch sehr bedeutsame andere Familie von Methoden geht auf die Mathematiker Routh und Hurwitz zurück, und umgeht die Berechnung der Nullstellen des Polynoms, um stattdessen durch genaue Betrachtung der Koeffizienten direkt zu einer hinreichenden und notwendigen Stabilitätsaussage zu kommen. Wir werden das Routh-Hurwitz Kriterium, das in der Praxis heute nicht mehr so bedeutsam ist wie früher, aus Platzgründen hier nicht beschreiben, aber geben ein verwandtes notwendiges Stabilitätskriterium, um einen kleinen Geschmack von dieser Familie von Stabilitätskriterien zu geben. Die einfache Aussage, die wir selbst beweisen können, ist: wenn alle Nullstellen $\bar{s}_1, \dots, \bar{s}_n$ des reellen Polynoms $N_{\text{CL}}(s) = c_n s^n + \dots + c_0$ negativen Realteil haben, dann sind notwendigerweise alle reellen Koeffizienten c_i von Null verschieden, und haben das gleiche Vorzeichen. Der Beweis geht von der Nullstellenzerlegung $c_n \prod_{i=1}^n (s - \bar{s}_i)$ aus und multipliziert den Ausdruck $\prod_{i=1}^n (s - \bar{s}_i)$ aus, um alle Koeffizienten $\frac{c_0}{c_n}, \dots, \frac{c_{n-1}}{c_n}$ zu finden und zu sehen, dass sie alle positiv sind. Hierbei werden komplex-konjugierte Paare $\bar{s}_{i,i+1} = -\alpha_i \pm j\beta_i$ erst miteinander multipliziert, um einen reellen Ausdruck $(s - \bar{s}_i)(s - \bar{s}_{i+1}) = (s^2 + 2\alpha_i s + \alpha_i^2 + \beta_i^2)$ zu erhalten, der genau dann positive Koeffizienten hat, wenn $\alpha_i > 0$. Ebenso gilt natürlich für reelle Nullstellen $\bar{s}_i = -\alpha_i$, dass $(s - \bar{s}_i) = s + \alpha_i$, was wiederum ein Polynom mit positiven Koeffizienten ist. Multipliziert man nun mehrere Polynome, die allesamt positive Koeffizienten haben, erhält man wieder nur positive Koeffizienten, und der einfache Satz ist bewiesen. Die Kriterien von Routh und Hurwitz gehen nun einige Schritte weiter, um zu einem notwendigen und hinreichenden Stabilitätskriterium zu kommen, das nur auf den Koeffizienten c_i beruht.

9.2 Das Nyquist Kriterium

Wir hatten im vorhergehenden Abschnitt gesehen, dass der geschlossene Kreis eines offenen Kreises $G_0(s) = \frac{Z_0(s)}{N_0(s)}$ genau dann stabil ist, wenn alle Nullstellen des Nennerpolynoms $N_{CL}(s) = N_0(s) + Z_0(s)$ in der linken komplexen Halbebene liegen, also keine Nullstelle in der rechten Halbebene liegt. Diese Nullstellen sind interessanterweise auch genau die Nullstellen der Funktion

$$\frac{1}{S(s)} = 1 + G_0(s) = \frac{N_0(s) + Z_0(s)}{N_0(s)}.$$

Wir nennen diese Funktion der Einfachheit halber $F(s) = 1 + G_0(s)$. Wir können diese Funktion schreiben als

$$F(s) = k \frac{\prod_{i=1}^n (s - \bar{s}_i)}{\prod_{i=1}^n (s - s_i)},$$

und es gilt, wie für jede komplexe Zahl, dass

$$F(s) = |F(s)| e^{j\Phi_F(s)}$$

wobei sich der Argumentwinkel $\Phi_F(s) = \arg F(s)$ als Summe der Argumentwinkel der Null- und Polstellen ergibt:

$$\Phi_F(s) = \arg k + \sum_{i=1}^n \arg(s - \bar{s}_i) - \sum_{i=1}^n \arg(s - s_i).$$

Für die Vorbereitung auf die weitere Argumentation betrachten wir zunächst nur eine einzige Nullstelle und nehmen für einen Moment eine vereinfachte Funktion $F(s) = k(s - s_0)$ an, mit $k, s_0 \in \mathbb{C}$. Betrachtet man nun eine geschlossene, kreisförmige Kurve $C \subset \mathbb{C}$ in der komplexen Ebene, dann kann man sich fragen, welches Bild diese Kurve unter der Abbildung $F(s)$ hat. Wir nennen das Bild $F(C)$, und dieses Bild formt wieder eine neue Kurve in der komplexen Zahlenebene, denn $F(C) \subset \mathbb{C}$. Man kann für diese neue Kurve fragen, wie oft sie die Null umrundet. Die Antwort auf diese Frage hängt davon ab, ob die ursprüngliche Kurve C die Nullstelle s_0 umrundet hat oder nicht. Falls ja, dann umrundet die Kurve $F(C)$ die Null in genau derselben Richtung, und falls nicht, dann umrundet die Kurve $F(C)$ die Null nicht.

Wenn es nun mehrere Nullstellen gibt, dann ergibt sich aufgrund der Additivität der Argumentwinkel der Nullstellen, dass die Bildkurve $F(C)$ den Nullpunkt für jede umkreiste Nullstelle genau einmal umkreist. Falls es auch noch Polstellen gibt, dann nimmt jede umkreiste Polstelle wieder eine Umdrehung der Null zurück.

Die Idee des Nyquistkriteriums nutzt nun eine Kurve C_{Nyquist} , die die gesamte rechte Halbebene der komplexen Ebene umschliesst, und die man auch die Nyquist-Kurve nennt. Diese Kurve ergibt sich, wenn man bei $s = 0$ startet, dann auf der imaginären Achse senkrecht nach oben bis $s = j\infty$ geht, und dann in einem unendlich grossen Halbkreis im Uhrzeigersinn die rechte Halbebene umschliesst, bis man unten beim Wert $s = -j\infty$ ankommt, und von dort wieder senkrecht nach oben geht, bis man zurück beim Punkt $s = 0$ angekommen ist.

Interessanterweise ist die Bildkurve $F(C_{\text{Nyquist}})$ genau das Nyquistdiagramm der Funktion $F(s)$, und dieses kann wegen $F(s) = 1 + G_0(s)$ ganz einfach durch Rechtsverschiebung des Nyquist-Diagramms von $G_0(s)$ um die reelle Zahl $+1$ erhalten werden. Dieser Zusammenhang ist in Abbildung 9.1 illustriert.

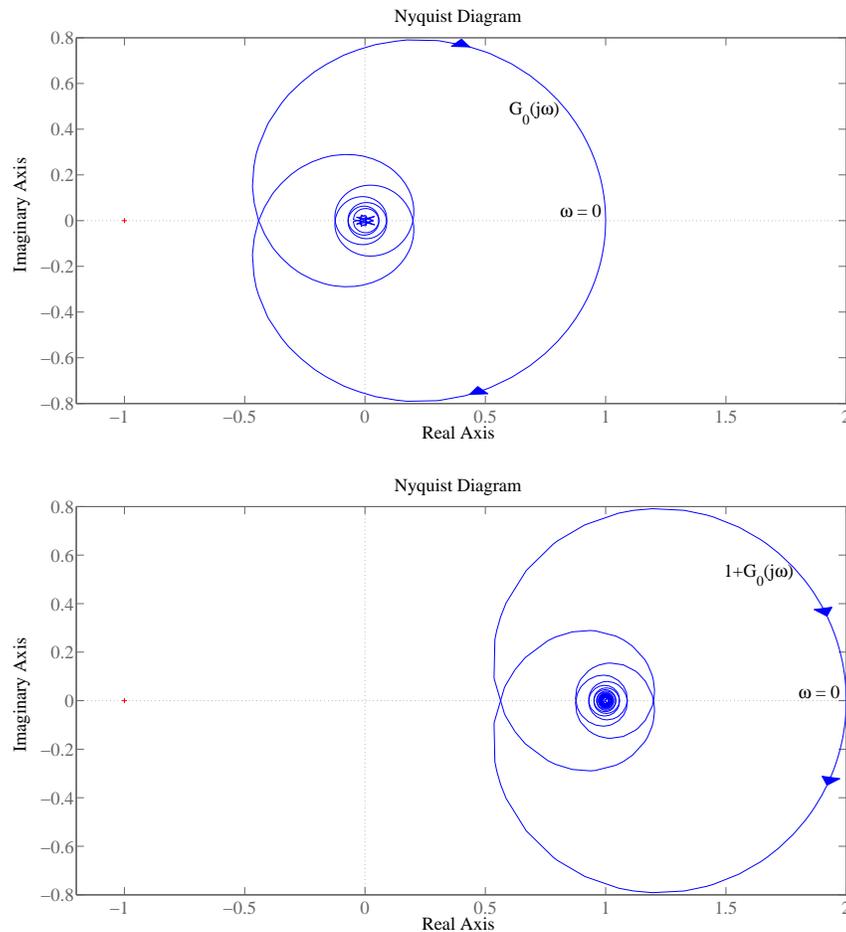


Abbildung 9.1: Nyquist-Diagramm des offenen Kreises $G_0(s)$ (oben) und der Funktion $F(s) = 1 + G_0(s)$ (unten), visualisiert für $G_0(s) = \frac{1}{1+s} \cdot e^{-s}$.

Wenn wir zunächst annehmen, dass der offene Kreis $G_0(s)$ stabil ist, also dass keine Nullstelle von $N_0(s)$ in der rechten Halbebene liegt, dann ist der geschlossene Kreis genau dann stabil, wenn die Funktion $F(s)$ weder Pole noch Nullstellen in der rechten Halbebene hat, wenn also die Null niemals von $F(C_{\text{Nyquist}})$ umkreist wird. Anstatt zu fragen, wie oft der Punkt 0 von der Kurve $F(C_{\text{Nyquist}})$ umrundet wird, kann man

wegen $F(s) = 1 + G_0(s)$ natürlich auch einfach fragen, wie oft der Punkt -1 von der Bildkurve $G_0(C_{\text{Nyquist}})$ umrundet wird, also vom Nyquist-Diagramm des offenen Kreises. Im oberen Diagramm der Abbildung 9.1 ist der Punkt -1 rot markiert, und man sieht, dass das Nyquist-Diagramm des offenen Kreises diesen Punkt links liegen lässt, der geschlossene Kreis also stabil ist. Wir fassen das bisherige Resultat in dem folgenden Satz zusammen.

Theorem 1 (Nyquist-Kriterium für stabile offene Kette) *Ist die offene Kette $G_0(s)$ stabil, dann ist der geschlossene Kreis $\frac{G_0(s)}{1+G_0(s)}$ genau dann stabil, wenn das Nyquist-Diagramm von $G_0(s)$ die komplexe Zahl -1 links liegen lässt, und diese nicht umrundet.*

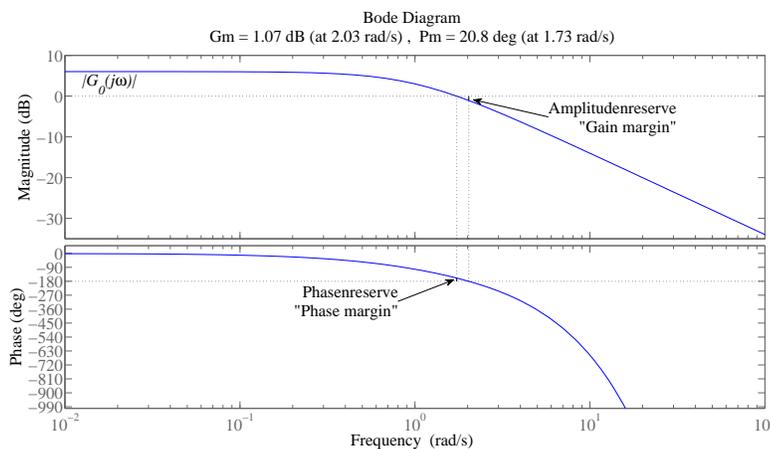


Abbildung 9.2: Visualisierung der Amplituden- und Phasenreserve ($G_0(s) = \frac{2}{1+s}e^{-s}$).

Von besonderem Interesse ist der Abstand, den das Nyquist-Diagramm einer stabilen offenen Kette vom Punkt -1 einhält, denn je größer dieser Abstand, umso eher bleibt der geschlossene Regelkreis auch bei Abweichungen des reellen Systems vom nominellen System stabil. Man spricht deshalb bei den folgenden zwei Größen auch von “Robustheitsmaßen”.

Definition 1 (Phasenreserve (phase margin, PM)) *Die Frequenz, bei der die Amplitudenverstärkung $|G_0(s)|$ des offenen Kreises den Wert eins schneidet, heißt “Schnittfrequenz” ω_s , es gilt also*

$$|G_0(j\omega_s)| = 1 \quad \text{bzw.} \quad |G_0(j\omega_s)|_{\text{dB}} = 0.$$

Der Winkelabstand, den der Wert $G_0(j\omega_s)$ vom Punkt -1 hat, nennen wir die “Phasenreserve” (engl. phase margin) Φ_R . Die Phasenreserve ist definiert als

$$\Phi_R = \arg(G_0(j\omega_s)) - (-\pi).$$

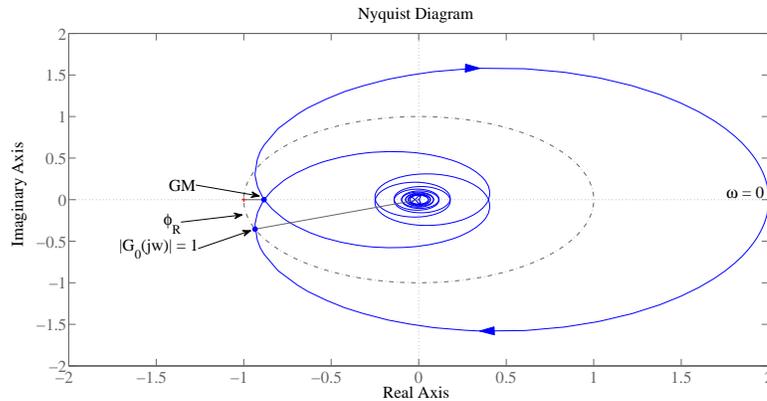


Abbildung 9.3: Visualisierung der Amplituden- und Phasenreserve ($G_0(s) = \frac{2}{1+s}e^{-s}$).

Die Phasenreserve, die hier im Bogenmaß definiert ist, aber auch oft in Grad gemessen wird, wird in MATLAB mit dem Kürzel PM für “Phase Margin” bezeichnet. Man kann die Schnittfrequenz und Phasenreserve sehr gut sowohl aus dem Bode-Diagramm als auch dem Nyquist-Diagramm ablesen.

Definition 2 (Amplitudenreserve (gain margin, GM)) Nennen wir die Frequenz, bei der das Nyquist-Diagramm des offenen Kreises die negative reelle Achse schneidet ω_{-180° , also $\arg G_0(j\omega_{-180^\circ}) = -\pi$, dann definiert man die “Amplitudenreserve” (engl. gain margin) als den Faktor GM, mit dem man das Nyquist-Diagramm des offenen Kreises noch multiplizieren könnte, bevor der Punkt -1 berührt wird

$$\text{GM} = \frac{1}{|G_0(j\omega_{-180^\circ})|}.$$

Die Definitionen der Phasen- und Amplitudenreserve sind in Abbildung 9.2 und Abbildung 9.3 verdeutlicht.

Beispiel: P-geregeltes PT₂-Glied

In Abbildung 9.4 sind zwei Systeme gezeigt, die beide das Nyquist-Stabilitätskriterium erfüllen, aber verschieden große Phasenreserven haben. Sie sind dadurch entstanden, dass man ein PT₂-Glied der Form

$$G(s) = \frac{1}{1 + 0.8s + 2s^2}$$

mit einem P-Regler der Form $K(s) = k_s$ verbindet, so dass der offene Kreis

$$G_0(s) = \frac{k_s}{1 + 0.8s + 2s^2}$$

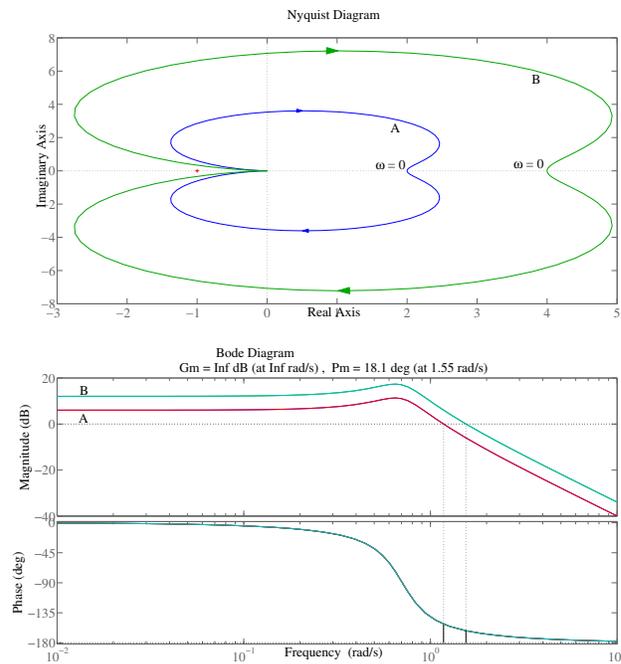


Abbildung 9.4: Visualisierung zweier offener Kreise $G_0(s)$ mit verschieden großer Phasenreserve (A: $G_0(s) = \frac{2}{1+0.8s+2s^2}$ und B: $G_0(s) = \frac{4}{1+0.8s+2s^2}$).

entsteht. Wählt man den Regelfaktor k_s (das *gain*) des P-Reglers höher, so verkleinert sich die Phasenreserve. In Abbildung 9.4 sind die Plots für $k_s = 2$ und $k_s = 4$ gezeigt. Die Amplitudenreserve ist unendlich, egal wie hoch man das Gain k_s wählt.

Grenzstabiler offener Kreis

Sobald man einen Integrationsanteil in der offenen Kette $G_0(s)$ hat, scheint das Nyquist-Kriterium in der bisherigen Formulierung nicht mehr anwendbar zu sein, da der offene Kreis nicht stabil ist. Er ist nur grenzstabil, da ein Eigenwert auf der imaginären Achse liegt. Man kann das Nyquist-Kriterium jedoch auch auf diesen Fall grenzstabiler offener Ketten anwenden, wenn man die Nyquist-Kurve C_{Nyquist} so wählt, dass sie infinitesimal rechts von der imaginären Achse verläuft. Man kann sich vorstellen, dass statt des I-Anteils ein ganz langsam abfallendes PT_1 -Glied in der Kette vorhanden ist, man also die folgende Ähnlichkeit nutzt

$$\frac{1}{s} \approx \frac{1}{\epsilon + s} = \frac{\frac{1}{\epsilon}}{1 + \frac{s}{\epsilon}}$$

Wenn man $\epsilon > 0$ gegen Null gehen läßt, dann werden die Kreise des PT_1 -Gliedes immer größer, aber bleiben in der rechten Halbebene. Das Nyquist Diagramm eines

I-Gliedes, das von $-j\infty$ nach $+j\infty$ geht, schliesst sich also auf der rechten Halbebene. Wenn das Nyquist-Diagramm also den Punkt -1 links liegen läßt, dann ist also der aus einer grenzstabilen offenen Kette resultierende geschlossene Kreis nach dem Nyquist-Kriterium stabil. Ein Beispiel für ein $G_0(s)$, das zu einem stabilen geschlossenen Kreis $\frac{G_0(s)}{1+G_0(s)}$ führt, ist in Abbildung 9.5 gezeigt. Es handelt sich um das System

$$G_0(s) = \frac{1}{s \cdot (s+1) \cdot (s+3)},$$

das zwei stabile Polstellen und eine grenzstabile Polstelle hat.

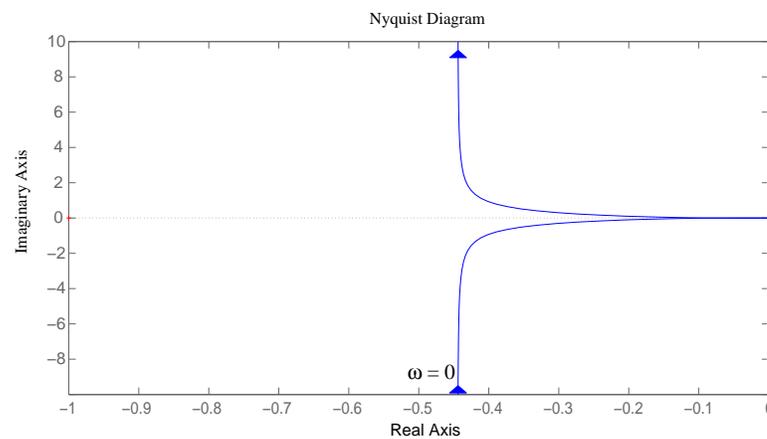


Abbildung 9.5: Das Nyquist-Diagramm eines grenzstabilen offenen Kreises, der zu einem stabilen geschlossenen Kreis führt, da er den Punkt -1 links liegen läßt, und sich in der rechten Halbebene schließt. ($G_0(s) = \frac{1}{s \cdot (s+1) \cdot (s+3)}$).

Allgemeine Form des Nyquist-Kriteriums

Das Zählen der Pol- und Nullstellen von $F(s)$, die von der Nyquist Kurve C_{Nyquist} umrundet werden, kann man auch anwenden, wenn der offene Kreis instabil ist, also $F(s) = \frac{N_0(s)+Z_0(s)}{N_0(s)}$ (bereits bekannte) Polstellen in der rechten Halbebene hat. Da die Polstellen negativ zählen und wir für Stabilität des geschlossenen Kreises keine Nullstellen von $F(s)$ haben möchten, muss eine negative Zahl von Umrundungen des Punktes -1 vorkommen, um Stabilität zu garantieren. Es gilt folgender Satz.

Theorem 2 (Nyquist-Kriterium für instabile offene Kette) *Ist die offene Kette $G_0(s)$ instabil und hat $n^+ > 0$ instabile Pole, dann ist der geschlossene Kreis $\frac{G_0(s)}{1+G_0(s)}$ genau dann stabil, wenn das Nyquist-Diagramm von $G_0(s)$ die komplexe Zahl -1 genau n^+ -mal gegen den Uhrzeigersinn umrundet.*

Zur Illustration betrachten wir das folgende Beispiel eines P-geregelten instabilen Systems der Form

$$G(s) = \frac{1}{(s-1)(s+3)} \quad \text{mit} \quad K(s) = K_P.$$

mit $K_P > 0$, das zu einem offenen Kreis

$$G_0(s) = \frac{K_P}{(s-1)(s+3)}$$

führt. Der Start des Nyquist-Diagramms liegt jetzt in der linken Halbebene bei $G(0) = \frac{-K_P}{3}$, und es ist in Abbildung 9.6 zu sehen, dass eine gewisse Mindestgröße des Proportionalglied-Faktors K_P notwendig ist, nämlich $K_P > 3$, um das instabile System zu stabilisieren. Eine Nullstellen-Analyse des charakteristischen Polynoms des geschlossenen Kreises, $N_0(s) + Z_0(s) = s^2 + 2s - 3 + K_P$, zeigt dasselbe.

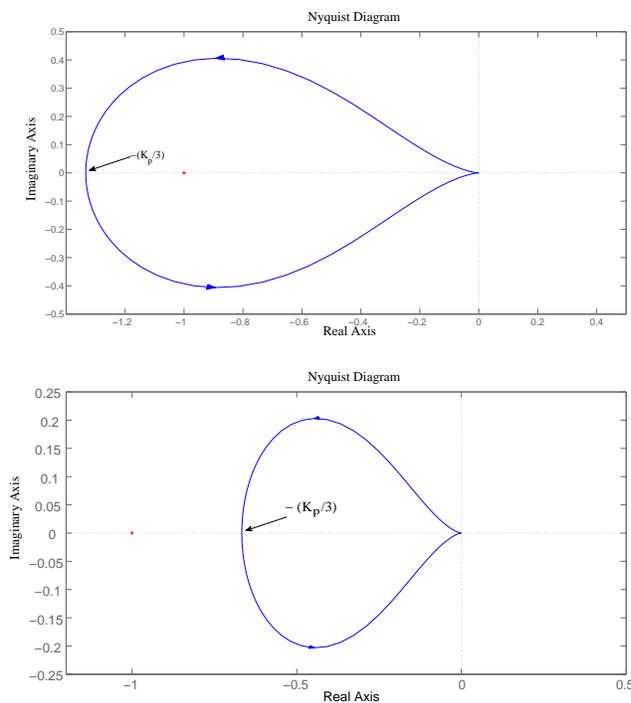


Abbildung 9.6: Der instabile offene Kreis $G_0(s) = \frac{K_P}{(s-1)(s+3)}$ mit einer instabilen Polstelle führt nach dem Nyquist-Kriterium genau dann auf einen stabilen geschlossenen Regelkreis, wenn der Punkt -1 einmal gegen den Uhrzeigersinn umrundet wird. Dies ist nur oben für $K_P = 4$ der Fall, unten für $K_P = 2$ nicht. Ein instabiles System kann nur stabilisiert werden, wenn der Regler aktiv eingreift.

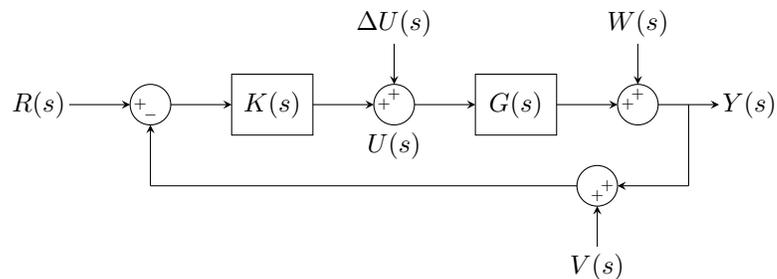


Abbildung 9.7: Blockdiagramm

Der vorliegende Abschnitt des Skriptes ist nicht vollständig. Bitte lesen Sie zusätzlich im Buch von Lunze den Abschnitt 8.5 “Stabilitätsprüfung von Regelkreisen anhand des Frequenzganges der offenen Kette”.

9.3 Innere Stabilität

Die Eingangs/Ausgangs-Stabilität vom Referenzsignal $R(s)$ zum Ausgangssignal $Y(s)$, also der Übertragungsfunktion $T(s)$, ist nicht die einzige Stabilitätsbedingung, die ein Regelkreis erfüllen muss, um in der Praxis brauchbar zu sein. Zusätzlich sollte ein Regelkreis auch im Innern stabil sein, er sollte also nicht instabil werden, wenn auf seine inneren Signale kleine Störungen wirken. Um die innere Stabilität zu definieren, betrachtet man alle möglichen Paare von möglichen Eingängen und Ausgängen, auch die auf die Steuerung wirkenden Störungen, siehe Abbildung 9.7. Dann berechnet man alle zu diesen Paaren gehörigen Übertragungsfunktionen, und nur wenn jede von ihnen stabil ist, sagt man, dass das System innerlich stabil ist. Wir betrachten neben $Y(s)$ auch das vom Regler erzeugte Signal $U(s)$ als Systemausgang, und betrachten alle vier in Abbildung 9.7 gezeigten Eingänge. Die Übertragungsfunktionen von $R(s)$, $W(s)$, $V(s)$ und $\Delta U(s)$ zu den zwei Ausgangsfunktionen $Y(s)$ und $U(s)$ führen zu acht möglichen Eingangs/Ausgangs-Paaren. Alle diese Eingangs-Ausgangs-Paare müssen stabil sein, damit das System die Eigenschaft der inneren Stabilität hat.

Achtung: Dieser Abschnitt des Skriptes ist nicht fertig. Bitte lesen Sie im Buch von Lunze den Abschnitt 8.4.2 “Innere Stabilität von Regelkreisen”.

Kapitel 10

Der PID Regler

Dieses Kapitel behandelt einen in der Praxis sehr häufig gebrauchten Reglertyp, den PID Regler. Er besteht aus der Parallelschaltung eines P, I, und D Gliedes, und seine Übertragungsfunktion hat die Form:

$$K_{\text{PID}}(s) = K_{\text{P}} + \frac{K_{\text{I}}}{s} + K_{\text{D}}s, \quad (10.1)$$

wobei die reellen Zahlen K_{P} , K_{I} , K_{D} frei wählbar sind und die Tuning-Parameter des PID Reglers sind. Ziel ist es, ein gegebenes System $G(s)$ zu regeln, und durch geschicktes Einstellen der drei PID Reglerparameter die resultierende Transferfunktion des offenen Kreises $G_0(s) = K_{\text{PID}}(s)G(s)$ günstig zu beeinflussen, um letztendlich dem geschlossenen Regelkreis $\frac{G_0(s)}{1+G_0(s)}$ gewünschte Eigenschaften zu geben, wie z.B. Stabilität, gute Führungsfolge, oder gute Störungskompensation. Im Zeitbereich kann man den PID Regler als ein Glied darstellen, das auf folgende Weise aus dem Fehler-signal $e(t) = r(t) - y(t)$ das Steuerungssignal $u(t)$ erzeugt:

$$u(t) = K_{\text{P}}e(t) + K_{\text{I}} \int_0^t e(\tau) d\tau + K_{\text{D}} \frac{de}{dt}(t).$$

Die Motivation, die jeweils hinter jedem der drei Glieder steht, kann man etwas überspitzt wie folgt formulieren:

- Proportional: Je mehr Regelabweichung, umso mehr Stelleingriff.
- Integral: Solange Abweichung besteht, erhöhe Stelleingriff.
- Derivative: Sobald Abweichung wächst, reagiere direkt mit Stelleingriff.

Die Wahl der drei Parameter ist entscheidend und eine der wichtigsten Aufgaben der regelungstechnischen Praxis. Der I-Anteil hilft dabei, bleibende Regelabweichungen zu eliminieren, und der D-Anteil, schnell genug auf Störungen zu reagieren. In der Praxis muss das D-Glied als DT_1 -Glied realisiert werden, da alle physikalisch realisierbaren

Systeme mindestens ebenso viele Pole wie Nullstellen haben müssen. Das DT_1 -Glied hat die Form

$$\frac{K_D s}{(1 + Ts)},$$

wobei die Zeitkonstante T kleiner gewählt wird als alle anderen vorkommenden Zeitkonstanten, d.h. die Frequenz $\frac{1}{T}$ ist höher als alle anderen für den Reglerentwurf wichtigen Frequenzen.

10.1 Analyse des PID Gliedes

Zur Analyse des PID Gliedes ist es hilfreich, statt der Form in Gleichung (10.1) die Zeitkonstantenform

$$K_{\text{PID}}(s) = K_P \left(1 + \frac{1}{T_I s} + T_D s \right), \quad (10.2)$$

zu verwenden, wobei nur der Parameter K_P identisch bleibt. Die ursprünglichen Parameter des I- und D-Gliedes können daraus wie folgt erhalten werden:

$$K_I = \frac{K_P}{T_I} \quad \text{und} \quad K_D = K_P T_D.$$

Die Sprungantwort des PID Gliedes besteht aus einem Dirac-Impuls aufgrund des D-Gliedes, einem endlichen Sprung aufgrund des P-Gliedes, und einem darauffolgenden rampenförmigen Anstieg aufgrund des I-Gliedes.

Die Übertragungsfunktion des PID Reglers kann in die Form

$$K(s) = \frac{K_P}{T_I} \frac{T_D T_I s^2 + T_I s + 1}{s}$$

gebracht werden, und die beiden Nullstellen im Zähler ergeben sich durch die Formel:

$$s_{01,2} = -\frac{1}{2T_D} \pm \sqrt{\frac{1}{4T_D^2} - \frac{1}{T_D T_I}}$$

Die Nullstellen sind für $T_I > 4T_D$ reell, was bei den meisten PID Reglern der Fall ist, und für $T_I < 4T_D$ werden sie imaginär.

Das Bode-Diagramm des PID-Reglers, das in Abb. 10.1 oben visualisiert ist, besteht bei zwei reellen Nullstellen, die als Knickfrequenzen auftreten, zunächst aus einem mit 20 dB pro Dekade abfallenden Teil mit Phase -90 Grad. Dieser geht bei der ersten Knickfrequenz in einen konstanten Teil mit Phase Null über, und nach der zweiten Knickfrequenz in einen Anstieg mit 20 dB pro Dekade und Phase +90 Grad. Das Nyquist-Diagramm des PID Reglers, das nur aus einer senkrecht von unten nach oben verlaufenden Linie besteht, ist in Abb. 10.1 unten gezeigt.

Durch die Realisierung des D-Gliedes als DT_1 -Glied ergibt sich am rechten Ende des Bodediagramms in der Realität noch ein weiterer Knick bei der Frequenz $\frac{1}{T}$, ab dem der Amplitudenplot konstant bleibt und die Phase auf Null Grad zurückfällt.

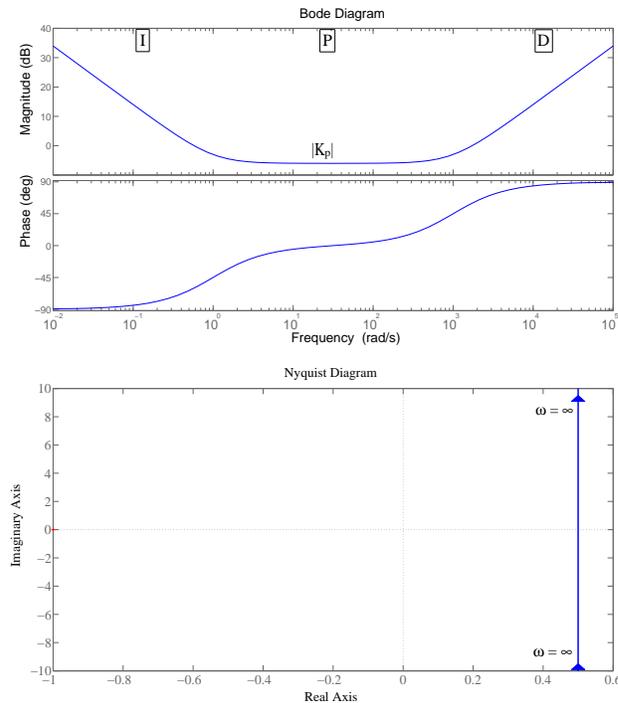


Abbildung 10.1: Bode- und Nyquist-Diagramm der Übertragungsfunktion $K(s)$ eines PID-Reglers. Das Bode-Diagramm beginnt typischerweise mit dem I-Anteil links bei den tiefen Frequenzen, dann dem P-Anteil in der Mitte, und dem D-Anteil für die hohen Frequenzen. Im Beispiel wurde $K(s) = \frac{1}{2} \cdot \left(1 + \frac{1}{s} + 0.001 \cdot s\right)$ gewählt.

10.2 PID Einstellregeln nach Ziegler-Nichols

Viele praktische Regelprobleme haben eine stabile Regelstrecke, die approximativ durch ein PT_1 Glied mit Totzeitanteil beschrieben werden kann, also durch

$$G(s) \approx \frac{k_s}{1 + T_s s} e^{-T_t s}.$$

Die Ortskurve beginnt dann beim reellen Wert k_s , geht erst durch die untere Hälfte der komplexen Ebene, um sich dann spiralförmig, sich im Uhrzeigersinn drehend, dem Ursprung zu nähern, siehe Abb. 10.2.

Die Idee der Regelungstechniker Ziegler und Nichols war es, unter der einfachen Annahme, dass das System stabil ist und approximativ durch obige Funktion beschrieben werden kann, einfache Regeln anzugeben, wie man die drei Parameter des PID Reglers einstellen kann, so dass für die allermeisten Fälle ein befriedigendes Regelverhalten erreicht wird. Entscheidend dabei ist, dass sie nicht davon ausgehen, dass ein

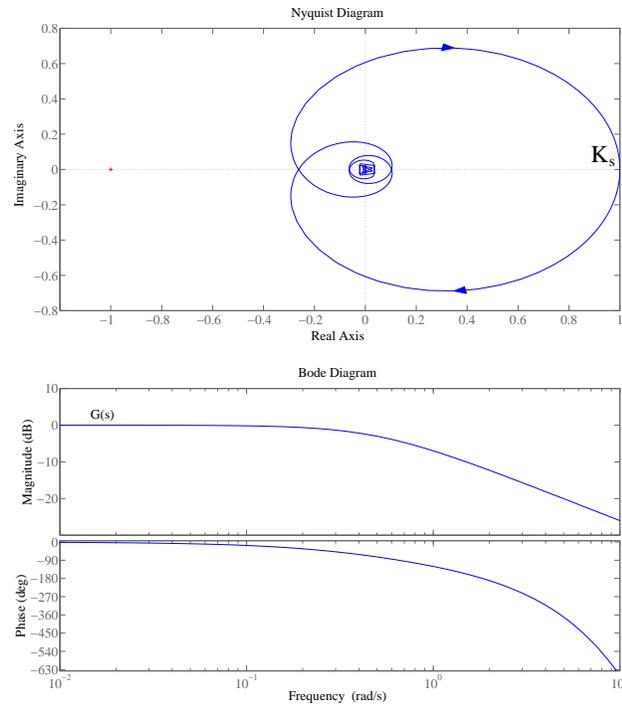


Abbildung 10.2: Nyquist und Bode-Plot eines typischen, für die PID-Regelung geeigneten Systems $G(s) = \frac{1}{1+2 \cdot s} \cdot e^{-s}$.

explizites Modell $G(s)$ der Strecke bekannt ist. Stattdessen findet man erst durch wenige reale Experimente einige wichtige Systemeigenschaften heraus, und entwirft danach direkt den PID Regler, indem man in einer einfachen Formeltabelle nachschaut, welches Tuning Ziegler und Nichols empfehlen. Wir diskutieren zwei Methoden, die erste extrem knapp, und die zweite noch immer kurz, aber etwas detaillierter.

Methode 1: Ermitteln der Sprungantwort

Die erste Methode verlangt nur, in einem einzigen Experiment einen Eingangssprung auf das unbekannte System zu geben, und die Sprungantwort aufzuzeichnen. Dies ist die Sprungantwort des offenen Kreises, und man kann die drei Streckenparameter k_s, T, T_t durch einen Kurvenfit der Sprungantwort an die Sprungantwort des Modells $G(s) = \frac{k_s}{1+T \cdot s} e^{-T_t \cdot s}$ ermitteln. Sobald man die drei Parameter k_s, T, T_t kennt, kann man mit Hilfe einer einfachen Formel daraus die drei Reglerparameter K_P, T_D, T_I ermitteln, und der Regler ist fertig. Man beachte, dass diese Methode zunächst ein sehr einfaches Streckenmodell ermittelt, und erst danach den Regler entwirft.

Methode 2: Ermitteln der Stabilitätsgrenze

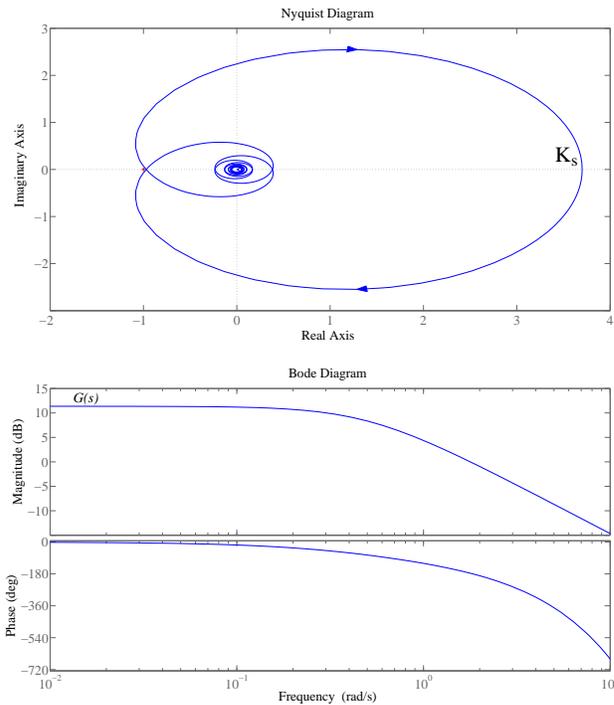


Abbildung 10.3: Die Stabilitätsgrenze ist für das System aus Abbildung 10.2 genau dann erreicht, wenn $K_P = 3.7$ ist, denn dann erfüllt der offene Kreis $G_0(s) = \frac{3.7}{1+2 \cdot s} \cdot e^{-s}$ das Nyquist-Kriterium beinahe nicht mehr. Der geschlossene Kreis beginnt zu oszillieren, siehe Abb. 10.4.

Die zweite Methode kommt ganz ohne ein Modell der Strecke aus, und geht folgendermaßen vor:

1. Schliesse den Regelkreis mit einem einfachen P-Glied mit sehr kleinem Verstärkungsfaktor (engl.: gain).
2. Erhöhe die Verstärkung, bis der geschlossene Kreis anfängt, zu oszillieren. Die Stabilitätsgrenze ist erreicht, und man merkt sich den kritischen Verstärkungsfaktor K_{kr} (siehe Abb. 10.3).
3. Man ermittelt zudem die kritische Periodendauer T_{kr} der Schwingung an der Stabilitätsgrenze (siehe Abb. 10.4).
4. Man verwendet jetzt die Formeln aus Tabelle 10.1, um die PID Reglerparameter zu finden.

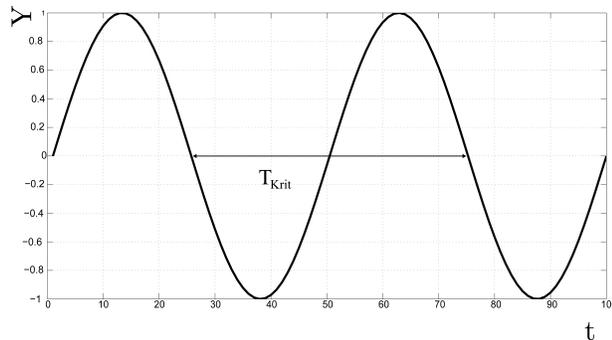


Abbildung 10.4: Periodisch schwingender Ausgang $y(t)$ sobald die Stabilitätsgrenze durch Erhöhen des P-Gains auf den Wert K_{kr} erreicht ist. Die Periodendauer T_{kr} der Oszillation am kritischen Punkt kann man aus dem Zeitverlauf direkt ermitteln.

Typ	Zeitkonstantenform in (10.2)	Konstanten K_I, K_D in (10.1)
P	$K_P = 0.5K_{kr}$	-
PI	$K_P = 0.45K_{kr}, T_I = 0.85T_{kr}$	$K_I = 0.53 \frac{K_{kr}}{T_{kr}}$
PID	$K_P = 0.6K_{kr}, T_I = 0.5T_{kr}, T_D = 0.12T_{kr}$	$K_I = 1.2 \frac{K_{kr}}{T_{kr}}, K_D = 0.072K_{kr}T_{kr}$

Tabelle 10.1: Bestimmung der P, PI, und PID Reglerparameter nach der Stabilitätsgrenzenmethode von Ziegler und Nichols (Methode 2).

Für den Fall des P-Reglers kann man den Ziegler-Nichols Verstärkungsparameter $K_P = 0.5K_{kr}$ sehr einfach interpretieren: man wählt den Faktor K_P so gross wie möglich, um bleibende Regelabweichungen so klein wie möglich zu machen, aber bleibt von der Stabilitätsgrenze soweit entfernt, dass die Amplitudenreserve (engl. gain margin) gerade 2 beträgt. Auf Englisch nennt man die Methode 2 auch *ultimate gain method*.

10.3 Eingangssaturation und Anti-Wind-Up

Bei vielen dynamischen Systemen ist die Eingang beschränkt. Wenn man mit einem PID-Regler mehr Steuerungsinpult $u_{PID}(t)$ verlangt, als möglich ist, wird dieser Input nicht an das wirkliche System weitergegeben. Man sagt dann, dass der Eingang ist *saturiert* oder *gesättigt*. Die Saturation eines von einem PID-Regler generierten Signals $u_{PID}(t)$ ist definiert als

$$u(t) = \max(\min(u_{PID}(t), u_{\max}), u_{\min}).$$

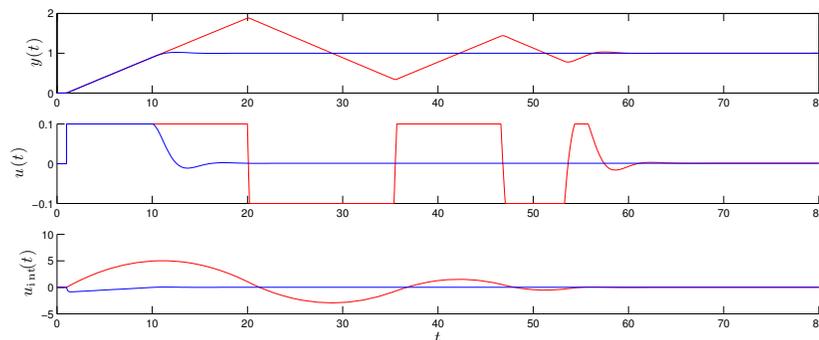


Abbildung 10.5: Anti-Wind-Up: PID (rot) und PID mit Anti-Wind-Up (blau)

Eine Eingangssaturation macht das System stark nichtlinear.¹

Diese Saturation ist bei einem P-Regler und einem D-Regler nur ein marginales Problem, da das zukünftige Eingangssignal unabhängig von vergangenen Ausgangssignalen ist. Eine Sättigung schneidet einfach nur das aktuell vom PD-Regler verlangte Signal ab.

Wenn man aber einen Regler benutzt, der ein Integrationsglied enthält, also einen PI- oder einen PID-Regler, dann muss man aufpassen. Denn falls das Eingangssignal für eine längere Zeit saturiert und der Ausgang dem Referenzsignal deshalb nicht folgen kann, dann wird der Integrator die Regeldifferenz immer weiter aufintegrieren und immer größere Steuerungen $u_{PID}(t)$ verlangen, ohne dass dies eine Auswirkung auf den Systemoutput hat. Der Integrationsanteil kann dabei beliebig groß werden und sogar ins Unendliche steigen. Und selbst falls der Ausgang durch die Dynamik des Systems oder günstige äußere Einflüsse irgendwann wieder in den “grünen Bereich” zurückkehrt, bleibt das vom I-Glied im Regler verlangte Eingangssignal noch sehr lange außerhalb der Sättigungsgrenze, so dass das wirkliche Steuersignal viel länger als nötig an dieser Grenze “haften” bleibt, was zu starken reglerinduzierten Oszillationen führen kann. Eine Illustration ist in Abbildung 10.5 gegeben, wobei die roten Signale durch eine Simulation des geschlossenen Regelkreises mit unmodifiziertem PID Regler und Eingangssaturation generiert sind. Man beachte, dass aufgrund der Sättigung die vom Integrationsglied verlangte Steuerung $u_{int}(t)$ sehr viel grösser ist als das wirklich realisierte Eingangssignal $u(t)$. Man nennt dieses Verhalten des I-Gliedes im Regelkreis “Wind-Up”, da der I-Anteil durch die Sättigung so wie die Feder einer Armbanduhr aufgezo-gen wird

Viele Anti-Wind-Up Techniken existieren, von denen wir hier nur die sogenannte *back calculation* Methode vorstellen wollen. Die Idee ist einfach: wenn das Eingangssignal saturiert, kann man die Differenz zwischen den gewünschten und den realen Eingangssignalen entweder messen oder modellieren. Nennen wir diese Differenz $\Delta u(t) = u_{PID}(t) - u(t)$. Wenn die Differenz $\Delta u(t)$ bekannt ist, kann man sie nutzen,

¹Große Teile dieses Abschnitts wurden von Mario Zanon erstellt, der auch angeregt hat, dieses wichtige Thema dem Kurs hinzuzufügen.

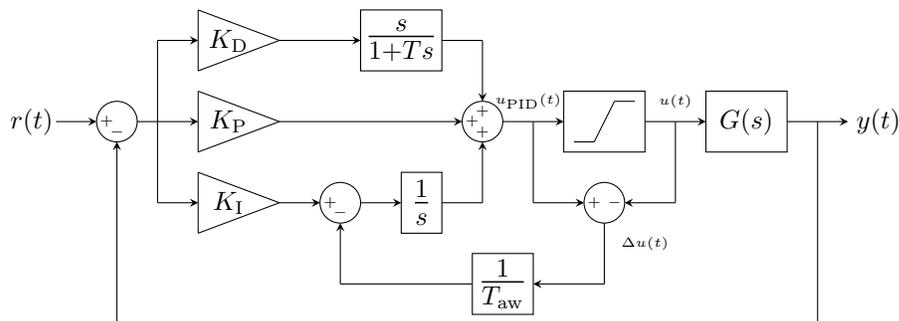


Abbildung 10.6: Blockdiagramm eines PID-Reglers mit Anti-Wind-Up Technik

um den Integrator immer dann zu stoppen, wenn das das Signal saturiert. Dies kann z.B. dadurch geschehen, dass man $\Delta u(t)$ erst durch die Anti-Wind-Up Zeitkonstante T_{aw} teilt, und diesen Wert dann vom Eingang des Integrationsglied des I-Reglers abzieht. Das Blockdiagramm eines PID-Reglers mit Anti-Wind-Up Modifikation ist in Abbildung 10.6 gezeigt, wobei die Sättigung in der Mitte ist, und der Anti-Wind-Up Anteil durch den zweituntersten Feedback-Pfeil gegeben ist.

Das Resultat der Anti-Wind-Up Technik ist, dass der I-Anteil umso stärker schrumpft, je mehr der Eingang gesättigt ist, so dass der I-Anteil nicht mehr ins unendliche wachsen kann. Empfohlene Werte für T_{aw} sind $T_{aw} = \sqrt{T_I T_D}$, falls man einen PID-Regler benutzt, und $T_{aw} = T_I$, falls man einen PI-Regler benutzt.

Im Beispiel in Abbildung 10.5 betrachten wir das System $G(s) = \frac{1}{s+0.001}$ und einen PID-Regler mit $K_P = 0.1$, $K_D = 1$ und $K_I = 1$. In der Abbildung sind die Sprungantworten eines normalen PID-Reglers und eines PID-Regler mit Anti-Wind-Up mit $T_{aw} = 0.1$ verglichen. Man kann in Rot sehen, dass, wenn kein Anti-Wind-Up benutzt wird, das Eingangssignal länger saturiert, der Integrator größere Eingangssignale generiert und der Ausgang länger braucht, bevor er zum Gleichgewicht kommt. In Blau kann man sehen, wie Anti-Wind-Up dieses Problem löst: das Eingangssignal saturiert nur solange, bis der Ausgang bei der Referenz ankommt, und das Integrationsglied generiert die gesamte Zeit nur kleine Signale.

Kapitel 11

Reglerentwurf im Frequenzraum

Das Ziel des Reglerentwurfes ist es, durch geschickte Wahl des Reglers $K(s)$ dem geschlossenen Regelkreis $G_{CL}(s)$ gewünschte Eigenschaften zu geben, die wir idealerweise im Zeitbereich angeben möchten, so wie eine kleine Überschwingweite, keine bleibende Regelabweichung, oder eine kurze Einschwingzeit.

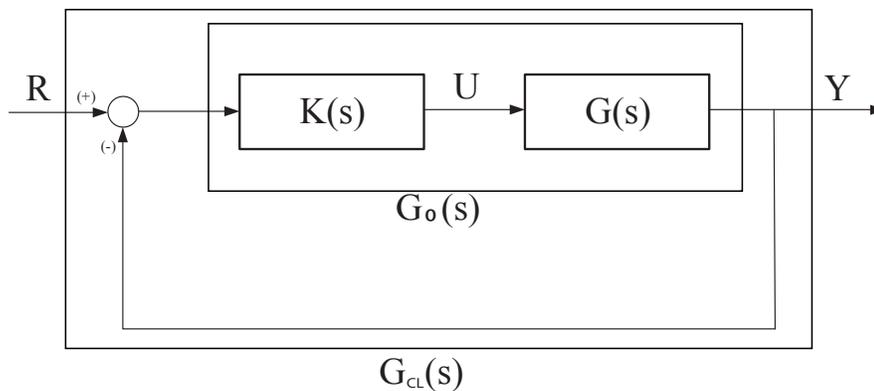


Abbildung 11.1: Geschlossener Regelkreis $G_{CL}(s)$.

Für ein gegebenes System $G(s)$ ist der Zusammenhang von $K(s)$ zu $G_{CL}(s)$ gegeben durch zwei Schritte, die in Abbildung 11.1 visualisiert sind: zunächst erhalten wir durch Zusammenschalten von $K(s)$ und $G(s)$ den offenen Kreis $G_0(s)$:

$$G_0(s) = G(s)K(s). \quad (11.1)$$

Zum zweiten erhält man aus dem offenen Kreis durch schliessen der Regelschleife den

geschlossenen Kreis (eng. closed-loop)

$$G_{\text{CL}}(s) = \frac{G_0(s)}{1 + G_0(s)}. \quad (11.2)$$

Der erste Zusammenhang in Gleichung (11.1) ist eine reine Multiplikation und kann in der logarithmischen Darstellung des Bode-Diagramms als Addition der zwei logarithmischen Übertragungsfunktionen aufgefasst werden. Der Zusammenhang zwischen $K(s)$, $G(s)$ und $G_0(s)$ im Frequenzraum ist unserer Intuition also leicht zugänglich. Der zweite Schritt in Gleichung (11.2) ist hingegen eine komplexere Operation, und der Zusammenhang zwischen $G_0(s)$ und $G_{\text{CL}}(s)$ ist zunächst weniger intuitiv. Sobald man jedoch ein Gespür für diesen Zusammenhang entwickelt hat, man also durch Betrachten des Bode-Diagramms von $G_0(s)$ die Eigenschaften des geschlossenen Kreises “sieht”, kann man für einen systematischen Reglerentwurf im Frequenzraum folgendermassen vorgehen: man betrachtet zunächst das Bode-Diagramm des unregulierten Systems $G(s)$ und überlegt dann, wie man durch “hinzufügen” oder “wegnehmen” von Amplitudenverstärkung und Phasenverschiebung in verschiedenen Frequenzbereichen mit Hilfe von $K(s)$ einen gewünschten offenen Kreis $G_0(s) = G(s)K(s)$ erzeugen kann. Das Ziel dieses Kapitels ist es, unsere Intuition für den Zusammenhang zwischen $G_0(s)$ und $G_{\text{CL}}(s)$, der bereits im Nyquist Stabilitätskriterium vorkam, weiter zu schärfen.

Einfachster Fall: integrierender offener Kreis

Im allereinfachsten Fall kann man den offenen Kreis einem Integrationsglied der Form

$$G_0(s) = \frac{1}{Ts} = \frac{\omega_0}{s}$$

ähnlich machen. Hierbei ist $\omega_0 = \frac{1}{T}$ die Frequenz, bei der der offene Kreis die Amplitudenverstärkung eins hat. Sie ist also die “Schnittfrequenz”, ab der im Bode-Amplitudendiagramm Null-dB-Wert unterschritten wird.

Der geschlossene Kreis ist dann beschrieben durch die folgende Übertragungsfunktion vom Referenzwert zum Ausgang:

$$G_{\text{CL}}(s) = \frac{G_0(s)}{1 + G_0(s)} = \frac{1}{\frac{1}{G_0(s)} + 1} = \frac{1}{Ts + 1}$$

Diese Übertragungsfunktion ist ein PT_1 -Glied mit statischer Verstärkung $G_{\text{CL}}(0) = 1$ und mit Bandbreite $\omega_B = \omega_0$, denn

$$|G_{\text{CL}}(j\omega_0)| = \frac{1}{|j + 1|} = \frac{1}{\sqrt{2}},$$

also $|G_{\text{CL}}(j\omega_0)|_{\text{dB}} = -3\text{dB}$. Man sieht einen ersten sehr einfachen Zusammenhang zwischen $G_0(s)$ und $G_{\text{CL}}(s)$, der auch für komplexere Regelkreise approximativ gilt: die Schnittfrequenz, bei der im offenen Kreis der Null-dB-Wert erreicht wird, ist im geschlossenen Kreis gleich der Bandbreite ω_B . Alle Frequenzen, die unterhalb von ω_B liegen, werden fast exakt und mit Verstärkung eins durch den geschlossenen Kreis übertragen, während die darüberliegenden Frequenzen gedämpft werden.

11.1 Approximation des geschlossenen Kreises durch ein PT₂-Glied

Der oben beschriebene einfachste Fall hatte nur einen einzigen Pol im geschlossenen Kreis, der genau auf der reellen Achse liegt. Eine realistischere Annahme für den geschlossenen Kreis ist, dass es mehrere stabile Pole gibt, die auch in komplex-konjugierten Paaren auftreten können. Das Verhalten des geschlossenen Kreises wird dann von demjenigen Pol-Paar dominiert, das am dichtesten an der imaginären Achse liegt; die Eigenfrequenzen der anderen Pole werden stärker gedämpft und treten im Bode-Diagramm des geschlossenen Kreises schwächer sichtbar in Erscheinung. Zudem erwarten wir von dem nach unseren Wünschen gestalteten geschlossenen Kreis, dass er die statische Verstärkung eins hat. Wir betrachten deshalb im folgenden einen geschlossenen Kreis, der die PT₂-Übertragungsfunktion

$$G_{CL}(s) = \frac{1}{1 + 2Tds + T^2s^2}$$

hat, wobei wir den schwach gedämpften Fall mit Dämpfungskonstante $d \in (0, 1]$ annehmen, und T die Zeitkonstante des Systems ist. Wir definieren wieder die natürliche Frequenz $\omega_0 = \frac{1}{T}$. Die beiden Pole des PT₂-Gliedes ergeben sich aus dem charakteristischen Polynom $p_A(\lambda) = \omega_0^2 + 2d\omega_0\lambda + \lambda^2$ zu

$$\lambda_{1,2} = \omega_0 \left(-d \pm j\sqrt{1 - d^2} \right).$$

Man sieht, dass die Grundfrequenz ω_0 den Abstand der Pole vom Nullpunkt bestimmt, während die Dämpfung d den Winkel α bestimmt, den der Polort mit der negativen reellen Achse formt: $d = \cos \alpha$.

11.2 Vom geschlossenen zurück zum offenen Kreis

Welche Übertragungsfunktion $G_0(s)$ der offenen Kette entspricht einem geschlossenen Kreis $G_{CL}(s)$, der durch ein PT₂-Glied beschrieben wird? Wir setzen einfach die Definition des geschlossenen Kreises gleich mit der Übertragungsfunktion des PT₂-Gliedes, und formen dies zu $G_0(s)$ um:

$$G_{CL}(s) = \frac{1}{\frac{1}{G_0(s)} + 1} = \frac{1}{T^2s^2 + 2Tds + 1}$$

woraus sich ergibt:

$$G_0(s) = \frac{1}{T^2s^2 + 2Tds}.$$

Dies ist ein integrierendes System, das noch einen Verzögerungsanteil hat, und das wir IT₁-Glied nennen. Wir können es als Hintereinanderschaltung eines I-Gliedes und eines PT₁-Gliedes auffassen:

$$G_0(s) = \frac{1}{2Tds} \frac{1}{\left(1 + \frac{T}{2d}s\right)}$$

und definieren uns unter Verwendung von $\omega_0 = \frac{1}{T}$ noch die beiden Frequenzen

$$\omega_s = \frac{1}{2d}\omega_0 \quad \text{und} \quad \omega_1 = 2d\omega_0,$$

so dass wir schreiben können:

$$G_0(s) = \frac{\omega_s}{s} \frac{1}{\left(1 + \frac{s}{\omega_1}\right)}.$$

Im Falle einer ausreichend großen Dämpfung $d > \frac{1}{2}$, die wir uns im geschlossenen Kreis fast immer wünschen, liegt die Knickfrequenz ω_1 höher als die Frequenz ω_s . Im Bode-Amplitudenplot des offenen Kreises sieht man für kleine Frequenzen eine negative Steigung von -20 dB pro Dekade, einen Schnittpunkt mit der 0 dB-Linie der dicht bei ω_s liegt, und ein Abknicken der Steigung auf -40 dB pro Dekade bei der Knickfrequenz ω_1 . Die Phase ist für kleine Frequenzen -90 Grad, und für hohe Frequenzen -180 Grad.

Es ist interessant, einen Zusammenhang mit der Phasenreserve zu machen, also dem Abstand der Phase zu -180 Grad bei der Schnittfrequenz.

Diese Phasenreserve ist fast 90 Grad, also sehr hoch, falls der Abstand $a = \frac{\omega_1}{\omega_s} = 4d^2$ zwischen den Frequenzen gross ist, wir also eine starke Dämpfung des geschlossenen Kreises erzielt haben. Wir sehen folgende Zusammenhänge zwischen der Form des Bode-Plots für $G_0(s)$ und den Eigenschaften des dominanten Pols des geschlossenen Kreises, der als PT₂-Glied approximiert werden kann:

- Der Bode-Plot des offenen Kreises $G_0(s)$ ist bei tiefen Frequenzen idealerweise einem Integrationsglied ähnlich, hat also die Steigung -20dB pro Dekade und Phase -90 Grad.
- Die Schnittfrequenz, bei der der Bode-Amplitudenplot von $G_0(s)$ den Wert 0 dB erreicht, ist in etwa gleich der Bandbreite ω_B .
- Je grösser der Abstand zwischen der Schnittfrequenz und der darauffolgenden Knickfrequenz, umso stärker die Dämpfung des geschlossenen Kreises.

11.3 Korrekturglieder

Achtung, dieser Abschnitt des Skripts ist nicht fertig. Bitte studieren Sie im Buch von Jan Lunze den Abschnitt 11.2 (Reglerentwurf unter Beachtung des Führungsverhaltens), insbesondere Abschnitt 11.2.2 (Entwurfsdurchführung) und Abbildung 11.11 (wichtige Korrekturglieder).

Kapitel 12

Regelung im Zustandsraum

In den folgenden zwei Kapiteln wollen wir einen anderen Ansatz zum Entwurf von Regelungssystemen betrachten, der nicht im Frequenzraum mit Übertragungsfunktionen arbeitet, sondern direkt mit Zustandsraummodellen des zu regelnden Systems. Wir betrachten also ab jetzt ein LTI-SISO System, das im Zeitbereich beschrieben wird, und dessen Eingangs-Ausgangsverhalten von $u(t)$ nach $y(t)$ durch ein Zustandsraummodell mit Matrizen $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times n}$ und $D \in \mathbb{R}^{m \times m}$ beschrieben wird, wie in Abbildung 12.1 gezeigt.

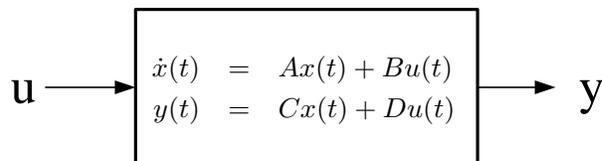


Abbildung 12.1: Zustandsraumdarstellung des zu regelnden Systems.

Zustandsbasierte Regelungsmethoden sind sehr leicht auf MIMO Systeme, auf mit relativ geringem Aufwand sogar auf Systeme mit Eingangs- oder Ausgangsbeschränkungen und nichtlineare Systeme verallgemeinerbar. Da der Fall von MIMO Systemen mit den selben Formeln beschrieben werden kann wie der Fall von SISO Systemen, erlauben wir in den Formeln in diesem Kapitel den Fall mehrerer Eingänge, also $u(t) \in \mathbb{R}^m$. Für den in dieser Vorlesung hauptsächlich behandelten Fall von LTI-SISO Systemen gilt dann natürlich einfach $m = 1$, und in den Beispielen und Abbildungen nehmen wir weiterhin an, dass $u(t) \in \mathbb{R}$ und $y(t) \in \mathbb{R}$. Da zustandsraum-basierte Regelungsmethoden einige Jahrzehnte später als die frequenzbasierten Methoden entwickelt und verwendet wurden, nennt man sie manchmal auch “moderne Methoden der Regelungstechnik”, aber wir wollen diesen selbst wieder etwas veralteten Term in diesem Skript nicht weiter verwenden. Zustandsbasierte Regelungsmethoden wur-

den insbesondere seit Ihrer Verwendung im Apollo-Mondlandungs-Programm in den 1960er Jahren sehr populär, und sind bis heute ein aktives Forschungsthema in der Industrie und an Universitäten, insbesondere im Bereich der optimierungsbasierten und der modell-prädiktiven Regelung.

Das Ziel des Reglerentwurfs ist zunächst die Stabilisierung des Systems, die wir zunächst betrachten wollen, und dann natürlich die Sollwertfolge. Die Idee, die allen zustandsbasierten Regelungsmethoden zugrunde liegt, ist, den aktuellen Zustandsvektor $x(t)$ des Systems entweder vollständig zu messen, oder ihn zu schätzen, und dann darauf basierend den Steuerungseingang $u(t)$ zu berechnen. Wir wollen in diesem Kapitel annehmen, dass der gesamte Zustandsvektor $x(t) \in \mathbb{R}^n$ gemessen werden kann – man nennt diesen Fall “Zustandsfeedback” – und erst im folgenden Kapitel das in der Praxis sehr wichtige Thema der Zustandsschätzung behandeln. Das Blockdiagramm eines geschlossenen Kreises mit Zustandsrückführung ist in Abbildung 12.2 illustriert, während wir in Abbildung 12.3 einen Ausblick auf den Zusammenhang zwischen Zustandsregler und die im folgenden Kapitel behandelte Zustandsschätzung geben.

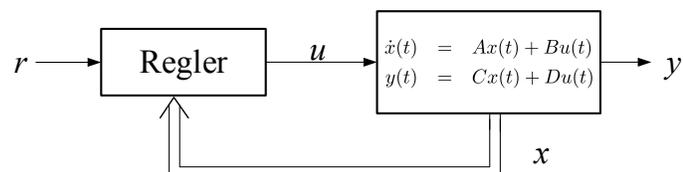


Abbildung 12.2: Regler mit Zustandsrückführung.

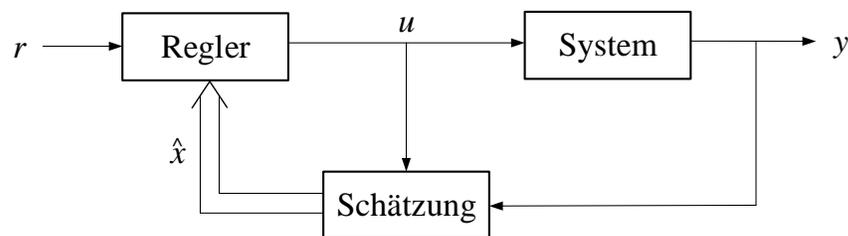


Abbildung 12.3: Zustandsregler mit Zustandschätzer (siehe Kapitel 13).

Um wieder ein lineares System zu bekommen, machen wir für den Zustandsregler, also für die Abbildung von $x(t)$ zu $u(t)$, den Ansatz

$$u(t) = -Kx(t)$$

mit einer Matrix $K \in \mathbb{R}^{m \times n}$. Für LTI-SISO Systeme ist die Matrix K ein liegender Vektor, und wir können n Koeffizienten k_0, \dots, k_{n-1} wählen:

$$K = [k_0, \dots, k_{n-1}].$$

Wenn man diesen Ausdruck in die Systemgleichungen $\dot{x} = Ax + Bu$ des offenen Kreises (*open-loop system*) einsetzt, dann bekommen wir die Systemgleichung des geschlossenen Kreises (*closed-loop system*)

$$\dot{x}(t) = Ax(t) + B(-Kx(t)) = \underbrace{(A - BK)}_{=: A_{CL}} x(t)$$

Die Frage, die sich jetzt stellt, ist, ob wir die Matrix K so wählen können, dass die closed-loop Systemmatrix $A_{CL} = A - BK$ stabil ist. Dies ist in der Tat meist der Fall, allerdings nur unter einer Bedingung, die wir im folgenden Abschnitt definieren und untersuchen wollen.

12.1 Steuerbarkeit

Um Zustandsregler entwerfen zu können, müssen wir eine wichtige Bedingung testen, die wir "Steuerbarkeit" nennen. Wir hatten diese als technische Bedingung schon einmal erwähnt im Zusammenhang mit der Regelungsnormalform, die nur unter dieser Bedingung existierte, aber wir wollen die Steuerbarkeit in diesem Abschnitt etwas genauer untersuchen. Dann werden wir eine sehr wichtige Eigenschaft steuerbarer Systeme kennenlernen, nämlich die Tatsache, dass wir die Pole des geschlossenen Kreises beliebig wählen können. Beginnen wir mit der exakten Definition der Steuerbarkeit.

Definition 3 (Steuerbarkeit, Controllability) Das Matrizenpaar (A, B) mit $A \in \mathbb{R}^{n \times n}$ und $B \in \mathbb{R}^{n \times m}$ heißt "steuerbar" (engl. *controllable*) genau dann, wenn die Steuerbarkeitsmatrix

$$C := [B, AB, A^2B, \dots, A^{n-1}B]$$

den vollen Rang n hat.

Diese Definition der Steuerbarkeit gilt ganz allgemein für LTI-MIMO Systeme mit m Eingängen. Man beachte, dass $C \in \mathbb{R}^{n \times (nm)}$. Für die in dieser Vorlesung betrachteten LTI-SISO Systeme mit $m = 1$ ist der volle Rang der dann quadratischen Steuerbarkeitsmatrix $C \in \mathbb{R}^{n \times n}$ gleichbedeutend damit, dass sie invertierbar ist, also $\det(C) \neq 0$.

Zur Motivation des Namens "Steuerbarkeit" ist es interessant zu bemerken, dass eine noch allgemeinere Definition von Steuerbarkeit besteht, die einfach nur besagt, dass ein System genau dann steuerbar ist, wenn es in endlicher Zeit $T > 0$ von jedem beliebigen Anfangszustand $x(0)$ zu jedem beliebigen Endzustand $x(T)$ gesteuert werden kann. Man kann zeigen, dass diese Definition für LTI-Systeme zu der obigen Definition äquivalent ist.

***Beweisskizze:** Um den Beweis dieser Aussage zu skizzieren, können wir die allgemeine Formel

$$x(T) = e^{AT} \left(x(0) + \int_0^T e^{-At} B u(t) dt \right)$$

verwenden. Da e^{AT} invertierbar ist, sehen wir, dass man beliebige Endzustände genau dann erreichen kann, wenn man mit dem Integral $\int_0^T e^{-At} B u(t) dt$ durch Wahl der Trajektorien $u(t)$ den gesamten Raum \mathbb{R}^n erreichen kann. Jeder Term $e^{-At} B u(t)$ ist aufgrund der Matrixexponentialformel gegeben durch

$$e^{-At} B u(t) = B + AB(-t)u(t) + A^2 B \frac{(-t)^2 u(t)}{2!} + \dots + A^k B \frac{(-t)^k u(t)}{k!} \dots$$

Hierbei sind die Terme $\frac{(-t)^k u(t)}{k!}$ reelle Zahlen, und man kann natürlich nur die Richtungen erreichen, die im Spann von $B, AB, A^2 B, \dots$ sind, ganz egal, welchen Wert $-t$ und $u(t)$ haben. Das System ist also nur dann steuerbar, wenn

$$\mathbb{R}^n = \text{span}\{B, AB, A^2 B, \dots\}$$

und man kann zeigen, dass

$$\text{span}\{B, AB, A^2 B, \dots, A^{n-1} B, A^n B, \dots\} = \text{span}\{B, AB, A^2 B, \dots, A^{n-1} B\}$$

womit die Äquivalenz gezeigt ist: man kann genau dann das System von jedem Zustand zu jedem anderen steuern, wenn der Rang der Steuerbarkeitsmatrix \mathcal{C} gleich der Zustandsdimension n ist, die Spalten von \mathcal{C} also den gesamten Raum \mathbb{R}^n aufspannen.

Wir wollen zunächst an drei Beispielen die Steuerbarkeit von Matrizenpaaren (A, B) untersuchen.

Beispiel 1 (Der doppelte Integrator) Wir betrachten das System $\ddot{y} = u$, dessen Zustandsraumdarstellung durch die Matrizen $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $C = [1 \ 0]$, und $D = 0$ gegeben ist. Wir interessieren uns hier nur für das Paar (A, B) und wollen untersuchen, ob es steuerbar ist. Wir bilden deshalb $\mathcal{C} = [B|AB] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, und testen, ob diese Matrix den Rang $n = 2$ hat, bzw. ob sie invertierbar ist. Dies ist in der Tat der Fall, denn $\det(\mathcal{C}) = -1 \neq 0$.

Beispiel 2 Betrachten wir nun das Matrizenpaar $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ und $B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Wir bilden wieder $\mathcal{C} = [B|AB] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, und sehen, dass diese Matrix nur den Rang 1 hat, bzw. dass sie nicht invertierbar ist, denn $\det(\mathcal{C}) = 0$.

Beispiel 3 Als letztes untersuchen wir das Matrizenpaar $A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$ und $B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. Wir bilden wieder $\mathcal{C} = [B|AB] = \begin{bmatrix} b_1 & \lambda b_1 \\ b_2 & \lambda b_2 \end{bmatrix}$, und sehen, dass auch diese Matrix nur den Rang 1 hat, denn die zweite Spalte geht aus der ersten durch Multiplikation mit λ hervor. Man kann alternativ auch durch Rechnung sehen, dass $\det(\mathcal{C}) = 0$. Auch dieses System ist also nicht steuerbar.

12.2 Polvorgabe

Für steuerbare Systeme gilt eine sehr erstaunliche Eigenschaft, nämlich, dass wir mit Hilfe der Matrix K die Pole des geschlossenen Kreises *beliebig* vorgeben können. Wir betrachten nur den Fall eines SISO Systems mit $m = 1$.

Theorem 3 (Polvorgabe) *Wenn das Paar (A, B) steuerbar ist, dann kann man für beliebige n komplexe Zahlen $(\bar{s}_1, \dots, \bar{s}_n)$ die Matrix $K \in \mathbb{C}^{1 \times n}$ so wählen, dass die n Eigenwerte der Matrix $A - BK$ genau an den Stellen $(\bar{s}_1, \dots, \bar{s}_n)$ liegen.*

Der Beweis des Theorems nutzt eine Eigenschaft steuerbarer Systeme, die wir schon einmal in Abschnitt 3.3.1 verwendet haben, und in diesem Skript nicht beweisen wollen. Die Eigenschaft ist, dass jedes steuerbare System in die sogenannte *Regelungsnormalform* gebracht werden kann, also eine Zustandstransformation $x(t) = Tz(t)$ mit invertierbarer Matrix T existiert, so dass $A_z := T^{-1}AT$ und $B_z := T^{-1}B$ in Regelungsnormalform gebracht sind. Wir können diese Tatsache nutzen, um den Zustandsregler $u = -Kx = -K_z z$ mit $K_z := KT$ leichter zu analysieren. Ohne Einschränkung der Allgemeinheit nehmen wir an, dass das System bereits in Regelungsnormalform ist, also das Paar (A, B) durch folgende Ausdrücke gegeben ist:

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} \quad \text{und} \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \quad (12.1)$$

Die Systemmatrix $A_{\text{CL}} = A - BK$ des geschlossenen Kreises hat mit $K = [k_0, \dots, k_{n-1}]$ dann die Form

$$A_{\text{CL}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ (-a_0 - k_0) & (-a_1 - k_1) & \cdots & (-a_{n-2} - k_{n-2}) & (-a_{n-1} - k_{n-1}) \end{bmatrix}. \quad (12.2)$$

Jetzt wollen wir k_0, \dots, k_{n-1} so wählen, dass die Eigenwerte der Matrix A_{CL} genau die vorgegebenen Werte $(\bar{s}_1, \dots, \bar{s}_n)$ haben. Dies kann man durch einen Koeffizientenvergleich zweier Polynome erreichen. Zum ersten ist das charakteristische Polynom der Matrix A_{CL} gegeben durch

$$\det(sI - A_{\text{CL}}) = s^n + (a_{n-1} + k_{n-1})s^{n-1} \dots (a_1 + k_1)s + (a_0 + k_0) \quad (12.3)$$

Zum zweiten hätten wir gerne, dass das charakteristische Polynom seine Nullstellen genau bei den vorgegebenen Werten hat, dass also

$$\det(sI - A_{\text{CL}}) = \prod_{i=1}^n (s - \bar{s}_i).$$

Ausmultiplizieren des Ausdrucks auf der rechten Seite ergibt gewünschte Polynomkoeffizienten c_0, \dots, c_{n-1} in der folgenden Form

$$\prod_{i=1}^n (s - \bar{s}_i) = s^n + c_{n-1}s^{n-1} \dots c_1s + c_0. \quad (12.4)$$

Durch Vergleich von Gleichungen (12.4) und (12.3) erhalten wir eine explizite Formel für die Koeffizienten in der Matrix K , nämlich

$$k_i := c_i - a_i, \quad \text{für } i = 0, \dots, n-1. \quad (12.5)$$

Für ein System, das in Regelungsnormalform ist, haben wir damit einen einfachen Algorithmus zum Berechnen der Matrix K gefunden.

Da jedes steuerbare System in Regelungsnormalform gebracht werden kann, gilt die Konstruktion für alle steuerbaren Systeme: man bringt das System $\dot{x} = Ax + Bu$ durch eine Zustandstransformation $x = Tz$ erst in Regelungsnormalform $\dot{z} = A_z z + b_z u$, findet dann aus (A_z, B_z) wie beschrieben eine Matrix K_z , und muss dann am Ende nur noch die Zustandstransformation rückgängig machen, was man durch die Definition $K := K_z T^{-1}$ erreicht.

Das MATLAB Kommando “place”

In der Praxis kann man für die Polvorgabe den MATLAB Befehl `place` nutzen, der die folgende Syntax hat:

$$[K] = \text{place} (A, B, [s_1, s_2, \dots, s_n])$$

Hierbei gibt der Vektor $[s_1, s_2, \dots, s_n]$ die n Polstellen vor. Man beachte, dass man nur dann reelle Koeffizienten im charakteristischen Polynom und damit in der Matrix K bekommt, wenn man reelle Eigenwerte oder komplex konjugierte Paare von Eigenwerten vorgibt. Sonst bekommt die Matrix K komplexe Einträge, was in der regelungstechnischen Praxis ein Problem darstellen kann. Ausserdem gibt es eine kleine, in der Praxis leicht zu lösende Einschränkung beim Kommando `place`: man kann aus uns hier nicht interessierenden technischen Gründen nur paarweise verschiedene Polstellen vorgeben. Wenn man zwei gleiche Polstellen vorgeben möchte, muss man sie also leicht abändern, bevor man das Kommando `place` aufruft. Wir geben einen kleinen MATLAB-Beispiel-Code zur Verwendung des Kommandos.

```
>> A=[0 1 ; -1 2];
>> B=[0;1];
>> K=place(A,B,[-1.001,-0.999])
K =
   -0.0000    4.0000
```

Example 1 Wir wollen das Beispiel $A = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}$ und $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ auch noch einmal mit dem im Beweis des Theorems 3 beschriebenen Algorithmus lösen. Dies ist einfach möglich, denn das System hat bereits die Regelungsnormalform. Wir können also die

Koeffizienten $a_1 = -(2)$ und $a_0 = -(-1)$ einfach ablesen. Wenn wir als Pole zweimal (-1) vorgeben wollen, also $\bar{s}_1 = -1$ und $\bar{s}_2 = -1$, erhalten wir

$$\prod_{i=1}^2 (s - \bar{s}_i) = (s - (-1))(s - (-1)) = s^2 + 2s + 1,$$

also die die gewünschten Polynomkoeffizienten $c_0 = 1$ und $c_1 = 2$. Die Matrix $K = [k_0, k_1]$ muss dann genau die Koeffizienten

$$k_0 := c_0 - a_0 = 1 - 1 = 0 \quad \text{und} \quad k_1 := c_1 - a_1 = 2 - (-2) = 4$$

haben, also $K = [0 \quad 4]$. Wir haben also tatsächlich dasselbe Ergebnis wie in dem kleinen MATLAB Code herausbekommen.

12.3 Arbeitspunktwechsel

Bisher haben wir nur die Frage der Stabilisierung eines Systems auf den Ursprung betrachtet. Um einen Zustandsregler zu entwerfen, der einem Referenzsignal folgen kann, muss man zunächst den Zustand berechnen, der dem gewünschten Referenzsignal entspricht. Dafür beschränken wir uns auf Gleichgewichtszustände, das Ziel ist also, einem Referenzsprung möglichst gut zu folgen, bei dem sich der Referenzwert sehr lange nicht ändert, so dass das System Zeit hat, in den neuen Gleichgewichtszustand zu konvergieren. Betrachten wir also ein LTI-SISO System mit Referenzsignal $r(t) \in \mathbb{R}$. Der Gleichgewichtszustand (engl. *steady state*), der diesem Referenzsignal entspricht, besteht aus einem Zustand x_{ss} und einem Eingang u_{ss} , die folgende zwei Gleichungen erfüllen:

$$0 = Ax_{ss} + Bu_{ss} \quad \text{und} \quad r = Cx_{ss} + D_{ss}.$$

Durch Invertieren dieses linearen Gleichungssystems erhält man eine lineare Abhängigkeit des Gleichgewichtszustandes von $r(t)$, wie folgt:

$$\begin{aligned} \begin{bmatrix} x_{ss} \\ u_{ss} \end{bmatrix} &= \underbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{=: \begin{bmatrix} N_x \\ N_u \end{bmatrix}} r \end{aligned}$$

Mit Hilfe des Vektors $N_x \in \mathbb{R}^n$ und der Zahl $N_u \in \mathbb{R}$, die im voraus berechnet werden können, können wir den zu einer gewünschten Referenz $r(t)$ passenden Gleichgewichtszustand ausdrücken als

$$x_{ss}(t) = N_x r(t) \quad \text{und} \quad u_{ss}(t) = N_u r(t).$$

Das Steuerungssignal $u(t)$ des Zustandsreglers wird genauso wie zuvor bei der Stabilisierung auf den Ursprung berechnet, nur dass jetzt der Nullpunkt zu $x_{ss}(t)$ und $u_{ss}(t)$ verschoben ist:

$$u(t) = u_{ss}(t) - K(x(t) - x_{ss}(t)).$$

Es ist leicht zu zeigen, dass für ein nach einem Sprung konstant bleibendes Referenzsignal $r(t) \equiv r$ das System gegen den entsprechenden konstant bleibenden Gleichgewichtszustand x_{ss} und u_{ss} konvergiert. Wir definieren dafür den Zustandsregelfehler als $\Delta x(t) = x(t) - x_{ss}$ und berechnen seine Ableitung:

$$\frac{d\Delta x}{dt} = \dot{x} = Ax + Bu = Ax + \underbrace{Bu_{ss}}_{=-Ax_{ss}} - BK(x - x_{ss}) = (A - BK)\Delta x.$$

Wenn die Matrix $A_{CL} = A - BK$ so gewählt wurde, dass ihre Eigenwerte nur negative Realteile haben, dann konvergiert also $x(t)$ gegen x_{ss} , und damit auch $u(t)$ gegen u_{ss} .

Kapitel 13

Zustandsschätzung

Die im vorhergehenden Kapitel gemachte Annahme, dass der gesamte Zustand gemessen werden kann, ist in der Realität meist nicht erfüllt. Deshalb muss man bei Zustandsreglern den Systemzustand aus den verfügbaren Messinformationen rekonstruieren. Dies erfolgt mit Hilfe eines "Zustandsschätzers", den man oft auch "Beobachter" nennt. Die Zustandsschätzung ist nicht nur für die Regelungstechnik wichtig, sondern auch, wenn man mit Hilfe von Messinformationen aus der Vergangenheit die Zukunft voraussagen möchte, wie z.B. bei der Wettervorhersage. Man muss für eine gute Vorhersage neben einem guten Modell auch eine gute Schätzung des Anfangszustandes haben

13.1 Luenberger Beobachter und Beobachtbarkeit

Grundidee des Beobachters ist, einen geschätzten Zustand $\hat{x}(t)$ immer wieder mit den aktuellen Messungen zu vergleichen, und diesen bei Abweichungen so zu modifizieren, dass er besser zu den Messungen passt. Wir machen den folgenden Ansatz:

$$\frac{d\hat{x}}{dt}(t) = \underbrace{A\hat{x}(t) + Bu(t)}_{\text{open-loop simulation}} + L \underbrace{(y(t) - C\hat{x}(t) - Du(t))}_{\text{Abweichung/Vorhersagefehler}}$$

Hierbei nennt man die Matrix L auch das Luenberger-Gain. Um zu entscheiden, wie man L wählen sollte, betrachten wir den Schätzfehler

$$\Delta x(t) := \hat{x}(t) - x(t)$$

und seine Entwicklung über die Zeit

$$\begin{aligned} \frac{d\Delta x}{dt} &= \dot{\hat{x}} - \dot{x} \\ &= A(\hat{x} - x) + Bu - Bu + L(Cx + Du - C\hat{x} - Du) \\ &= (A - LC)\Delta x. \end{aligned}$$

Falls nun alle Eigenwerte der Matrix $(A - LC)$ stabil sind, wird der Fehler gegen Null konvergieren. Man kann tatsächlich die Eigenwerte dieser Matrix frei wählen, wenn das Matrix-Paar (A, C) die Eigenschaft der Beobachtbarkeit erfüllt, wie folgt.

Definition 4 (Beobachtbarkeit) Das Matrizen-Paar (A, C) mit $A \in \mathbb{R}^{n \times n}$ und $C \in \mathbb{R}^{p \times n}$ heißt "beobachtbar" genau dann, wenn die Matrix

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

den vollen Rang n hat.

Die Beobachtbarkeit ist dual zur Steuerbarkeit, denn durch Transposition geht die eine in die andere Bedingung über: (A, C) ist genau dann beobachtbar, wenn (A^\top, C^\top) steuerbar ist. Aufgrund dieser Dualität sehen wir auch sofort, warum man im Falle der Beobachtbarkeit die Pole der Matrix $(A - LC)$ frei wählen kann: da (A^\top, C^\top) steuerbar ist, kann man durch geschickte Wahl einer Matrix K_L die Pole der Matrix $(A^\top - C^\top K_L)$ frei wählen, und damit auch die Pole der Transponierten Matrix $(A - K_L^\top C)$. Man setzt dann einfach $L := K_L^\top$.

13.2 Vollständiger Zustandsregelkreis

In Abbildung 13.1 sehen wir, wie ein vollständiger Zustandsregelkreis mit Luenberger-Beobachter aussehen kann. In dieser Kombination wählt man die Eigenwerte des Beobachters typischerweise so, dass der Beobachterfehler schneller gegen Null konvergiert als der Regelfehler, man wählt also die Eigenwerte von $(A - LC)$ stärker negativ als die Eigenwerte von $(A - BK)$. Andererseits sollte man die Dynamik des Beobachters auch nicht zu schnell machen, denn dann reagiert er zu empfindlich auf Messrauschen.

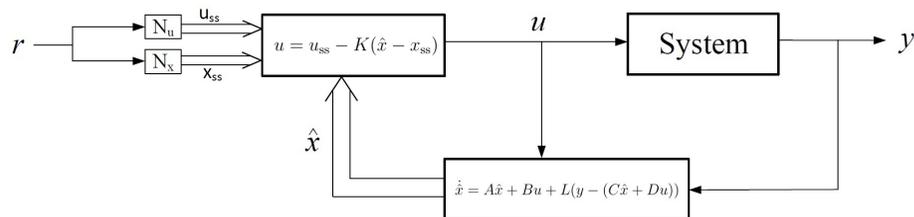


Abbildung 13.1: Vollständiger Zustandsregelkreis mit Beobachter.

13.3 Das Kalman Filter

Eine spezielle und in der Praxis weitverbreitete Wahl der Matrix L nutzt nicht, wie oben beschrieben, das Kommando `place`, sondern berechnet L mit Hilfe einer etwas komplexen Matrixgleichung. Diese Gleichung nennt sich die Algebraische Riccati Gleichung, und sie bestimmt aus zwei zu wählenden positiv definiten Matrizen Q und R eindeutig eine positiv definite Matrix P . Die Riccati Gleichung ist gegeben durch

$$0 = AP + PA^\top + Q - PC^\top R^{-1}CP.$$

Mit Hilfe der Matrix P bestimmt man die Matrix L des Kalman Filters jetzt durch

$$L_{\text{Kalman}} := PC^\top R^{-1}.$$

Die Motivation hinter dieser Wahl ist die folgende. Nimmt man an, dass das wirkliche System nicht deterministisch ist, sondern stochastisch, und sowohl durch Zustandsrauschen $w(t)$ als auch durch Messrauschen $v(t)$ beeinflusst wird, dann kann man zeigen, dass das Kalman Filter die Kovarianzmatrix des Schätzfehlers $\Delta x = \hat{x} - x$ minimiert. Die Annahmen dafür sind, dass das unterliegende Modell des stochastischen Systems gegeben ist durch

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

dass das Mess- und Zustandsrauschen unkorreliert sind, dass sie beide den Erwartungswert Null haben, und dass das Zustandsrauschen $w(t)$ die konstante Kovarianzmatrix Q und das Messrauschen $v(t)$ die konstante Kovarianzmatrix R hat. Die Matrix P gibt die – durch das Kalman-Filter minimierte – Kovarianzmatrix des Schätzfehlers an.

Anhang: Kleines Wörterbuch

Deutsch-Englisch

Deutsch	Englisch	Formelzeichen / MATLAB
Übertragungsfunktion	transfer function	$G(s)$
Eingangs-Ausgangsstabilität	input-output stability	
Regelung	feedback control	
Steuerung	(open-loop) control	
Verstärkung	gain	
statische Verstärkung	DC gain	$k_s = h(\infty)$
Proportionalglied	proportional gain	$G(s) = k_s$
Integrierglied	integrator	$G(s) = \frac{1}{Ts}$
Differenzierglied	derivative	$G(s) = Ts$
Verzögerung erster Ordn. (PT ₁)	first order lag	$G(s) = \frac{k_s}{Ts+1}$
Verzögerung zweiter Ordn. (PT ₂)	second order lag	$G(s) = \frac{k_s}{T^2s^2+2dT_s+1}$
Dämpfungsverhältnis	damping ratio	d
Eigenfrequenz	natural frequency	$\omega_0 = \frac{1}{T}$
Totzeitglied (T _t)	time delay	$G(s) = e^{-Ts}$
Bodediagramm	Bode plot	bode
Ortskurve	Nyquist plot	nyquist
Knickfrequenz	break point	
Polüberschuss, relativer Grad	relative degree	$n-q$ für $G(s) = \frac{b_q s^q + \dots + b_0}{a_n s^n + \dots + a_0}$
Sprungantwort (Übergangsfkt.)	step response	$h(t)$, step
Impulsantwort (Gewichtsfkt.)	impulse response	$g(t)$, impulse
Faltung	convolution	$(g*u)(t) = \int_0^t g(t-\tau)u(\tau) d\tau$
phasenanhebendes Korrekturglied	lead compensator	$G(s) = \frac{T_1s+1}{T_2s+1}$ mit $T_2 < T_1$
phasensenkendes Korrekturglied	lag compensator	$G(s) = \frac{T_1s+1}{T_2s+1}$ mit $T_2 > T_1$

Deutsch	Englisch	Formelzeichen / MATLAB
Regelstrecke	plant	$G(s)$
Regler	controller	$K(s)$
offene Kette	open loop transfer fun.	$G_0(s) = K(s)G(s)$
geschlossene Kette	closed loop transfer fun.	$T(s) = \frac{G_0(s)}{1+G_0(s)}$
Empfindlichkeitsfunktion	sensitivity function	$S(s) = \frac{1}{1+G_0(s)}$
komplementäre Empf.fkt.	complementary sens. fun.	$T(s) = \frac{G_0(s)}{1+G_0(s)}$
Bandbreite	bandwidth	ω_B mit $ T(j\omega_B) _{dB} = -3dB$
Durchtrittsfrequenz	crossover frequency	ω_s mit $ G_0(j\omega_s) = 1$
Phasenreserve/-rand	phase margin	$\Phi_R = \arg G_0(j\omega_s) - (-\pi)$
Amplitudenreserve/-rand	gain margin	$k_R = \frac{1}{ G_0(j\omega_{-180^\circ}) }$
Stabilitätsgrenzenmethode	ultimate sensitivity meth.	(nach Ziegler-Nichols)
Anstiegszeit	rise time	$T_{90\%}$
Überschwingzeit	peak time	
Überschwinghöhe	overshoot	Δh
Einschwingzeit	settling time	
innere Stabilität	internal stability	
gewöhnliche	ordinary	
Differentialgleichung	differential equation	$\dot{x} = f(x, u)$
ODE Integrator	ODE integrator	z.B. ode45
Spalte	column	
Zeile	row	
Einheitsmatrix	identity matrix	\mathbb{I} / eye
Eigenwerte	eigenvalues	eigs
Eigenvektor	eigenvector	
Matrixinverse	matrix inverse	inv(A)
transponierte Matrix	transposed matrix	A^T / A.'
konjugierte Matrix	conjugated matrix	A^H / A'
Mittelwert	mean	mean
Nenner	denominator	$N(s), D(s)$, den
Zähler	numerator	$Z(s), N(s)$, nom